The Journal
of the
Acoustical Society of America

LETTERS TO THE EDITOR

APPLIED ACOUSTICS PAPER: GENERAL LINEAR ACOUSTICS [20]

GENERAL LINEAR ACOUSTICS [20]

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA   VOL. 118, NO. 3, SEPTEMBER 2005

CONTENTS—*Continued from preceding page*

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Kenneth W. Grant—For contributions to auditory-visual speech processing.

James M. Kates—For contributions to signal processing for hearing aids.

Qing Liu—For contributions in computational elastodynamics and acoustics.

48th *Dallas Morning News*—Toyota Regional Science and Engineering Fair

More that 800 students from 86 North Texas schools competed in the 48th *Dallas Morning News*-Toyota Regional Science and Engineering Fair. This year one Senior and three Junior Division contestants won Outstanding Acoustics Project Awards.

Senior Division Winner Mr. Rick Yang, with the guidance of Ms. Rachel Carlson, Jasper HS, Plano ISD, reported sound waves at specifiable frequencies altered anaerobic respiration rates of yeast cultures. Although Mr. Yang offered no explanation, fermentation scientists expressed interest because the conditions under which rates increased did not arise from medium mixing or heating.

Mr. Martin Coffman, a student of Ms. Joyce Carson, Lowery Freshman Center, Allen ISD, received Senior Division Honorable Mention.

Junior Division Winners Ms. Alexa Hanks and Ms. Alianna Whiteaker-Chudecke with the guidance of Ms. Rachel Robb, St. Elizabeth of Hungary Catholic School, Dallas, used doves to affirm a central tenant of systematic biology: live and recorded dove calls elicited species-specific behavior; environmental sounds and calls of other song birds did not. Recorded hawk calls elicited behaviors common to doves and other song birds. Ms. Whiteaker-Chudecke and Ms. Hanks won Third Place in Junior Division Teams (see Fig.).

Junior Division Winner Ms. Jennifer Goebel, supervised by Ms. Deborah Johnston, Schimelpfenig Middle School, Plano ISD, demonstrated how drinking common beverages changed the surfaces lining mouth and throat, and altered the acoustic properties of a singer's voice.

Ms. Bethany Carson, a student of Ms. Doni Cash, Ford Middle School, Allen ISD, received Junior Division Honorable Mention, and the American Psychological Society "Outstanding Research Study" Award as well.

Dr. Peter Assmann, Dr. Greg Hughes, Dr. Ben Seep, and Dr. Michael Daly represented the Society.

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2004.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of 31 December 2004 and 2003 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

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

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of 31 December 2004 and 2003 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

O'CONNOR DAVIES MUNNS & DOBBINS, LLP
Accountants and Consultants
April, 1, 2005
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
AS OF 31 DECEMBER 2004 AND 2003

	2004	2003
Assets		
Cash and cash equivalents	$1,654,787	$2,349,616
Accounts receivable	337,150	222,195
Marketable securities, at market	6,458,334	5,053,884
Furniture, fixtures, and equipment—net	44,792	55,089
Other assets	451,527	483,789
Total assets	$8,946,590	$8,164,573
Liabilities		
Accounts payable and accrued expenses	$ 315,778	$ 327,250
Deferred revenue	1,604,183	1,196,196
Total liabilities	$1,919,961	$1,523,446
Net assets		
Unrestricted	$5,993,687	$5,640,080
Temporarily restricted	666,043	634,148
Permanently restricted	366,899	366,899
Total net assets	$7,026,629	$6,641,127
Total liabilities and net assets	$8,946,590	$8,164,573

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2004 AND 2003

	2004	2003
Changes in Unrestricted Net Assets		
Revenues:		
Dues..	$ 629,000	$ 664,508
Publishing—JASA..	2,156,994	2,132,642
Standards...	303,051	304,778
Spring Meeting...	414,755	186,353
Fall Meeting...	223,959	196,830
Other member services revenue..	12,715	17,488
Other..	101,708	102,190
Net assets released from restrictions..	69,970	68,348
	$3,912,152	$3,673,137
Expenses:		
Publishing..	$1,531,446	$1,566,927
Standards...	411,942	423,467
Spring Meeting...	467,159	236,943
Fall Meeting...	227,258	201,540
Member Services..	246,391	225,964
Other..	648,521	471,991
Administration..	484,939	430,978
	$4,017,656	$3,557,810
Net (loss) income from operations...............................	($ 105,504)	$ 115,327
Non-operating activities:		
Interest and dividends..	$ 124,903	$ 120,426
Realized gain (loss) on investments..	69,697	(569,494)
Unrealized gain on investments...	264,511	1,385,192
	$ 459,111	$ 936,124
Increase in unrestricted net assets.....................................	$ 353,607	$ 1,051,451
Changes in Temporarily Restricted Net Assets		
Contributions..	$ 21,855	$ 23,341
Interest and dividends..	20,708	22,768
Realized gain (loss)	12,368	(107,668)
Unrealized gain	46,934	261,885
Net assets released from restrictions..	(69,970)	(68,348)
Increase in temporarily restricted net assets.....................	$ 31,895	$ 131,978
Changes in Permanently Restricted Net Assets		
Contributions..	$ –	$ 1,880
Increase in permanently restricted net assets.......................	$ –	$ 1,880
Increase in net assets...	$ 385,502	$1,185,309
Net assets at beginning of year...	6,641,127	5,455,818
Net assets at end of year...	$7,026,629	$6,641,127

Annual Reports of the Technical Committees

Acoustical Oceanography

Fall 2004 Meeting (San Diego): Acoustical Oceanography (AO) sponsored a special session entitled "Acoustic Characterization of the Benthic Habitat," which was organized by Kenneth Foote (Woods Hole Oceanographic Institution). In addition AO co-sponsored a session with Underwater Acoustics (UW) entitled "Acoustic sensing of internal waves," which was organized by Jim Lynch (Woods Hole Oceanographic Institution) and featured a special lecture by Boris Katsnelson (Voronezh University, Russia). The Best Student Paper Awards in Acoustical Oceanography went to Eric Giddens (Scripps Institution of Oceanography), who won first place for his paper "Geoacoustic inversions in shallow water using Doppler shifted modes from a moving source," and Erica Summers-Morris (CSU Monterey Bay), who won second place for her paper "Geographical information systems (GIS) analysis of high-resolution multibeam bathymetry and remotely operated vehicle data to model rockfish habitat preference." Peter Worcester (Scripps Institution of Oceanography) and Peter Gerstoft (Marine Physical Laboratory) represented AO at the Technical Program Organizing Meeting.

Spring 2005 Meeting (Vancouver): The 2005 Medwin Prize in Acoustical Oceanography was awarded to Svein Vagle of the Institute of Ocean Sciences in Sidney, BC, "for development of acoustical techniques for probing the upper ocean boundary layer." He presented the Acoustical Oceanography Prize Lecture entitled "Acoustic explorations of the upper ocean boundary layer." AO sponsored two special sessions. Bruce Howe (Applied Physics Laboratory, University of Washington) organized a session jointly with UW and Animal Bioacoustics (AB) entitled "Acoustics and Ocean Observatories." Kenneth Foote (Woods Hole Oceanographic Institution) organized a session jointly with UW entitled "Riverine Acoustics." Srinivasan Jagannathan (Massachusetts Institute of Technology) won first place in the competition for the Best Student Paper Award in Acoustical Oceanography for his paper entitled "Could animals detect the approaching tsunami?" Ben Biffard (University of Victoria) was awarded second place for his paper entitled "Single beam echo sounding: considerations of depth and sea bed slope." Mike Wolfson (Applied Physics Laboratory, University of Washington) represented AO at the Technical Program Organizing Meeting.

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 (University of Rhode Island). 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.

This report is my first as Chair of the AO Technical Committee. I want to close by thanking all of the volunteers who have made the activities of the AO Technical Committee possible during this past year.

N. ROSS CHAPMAN
Chair

Animal Bioacoustics

Members of the Animal Bioacoustics Technical Committee (ABTC) continue to play an active role in the affairs of the Society and were well represented at both the San Diego (Fall 2004) and Vancouver (Spring 2005) meetings.

AB sponsored two named sessions at the San Diego meeting. One, on "Marine Mammal Acoustics," was held in honor of Ron Schusterman. It was chaired by Whit Au, Colleen Reichmuch Kastak, and Patrick Moore. A second session, on "Acoustic Monitoring of Large Whales—The Legacy of Paul Thompson," was held in memory of Paul Thompson. This session was chaired by John Hildebrand and Sue Moore. Andrea Simmons presented a lecture on "Hot topics in animal bioacoustics" at the interdisciplinary Hot Topics session. In addition, many AB members attended the tutorial lecture on "Ocean noise and marine mammals" presented by Gerald D'Spain and Douglas Wartzok. The ABTC meeting attracted over 50 attendees; discussion centered around public policy and standards issues.

We congratulate ABTC member D. Vance Holliday on his receipt of the Silver Medal in Acoustical Oceanography.

The recipients of the Best Student Paper Award were Mandy L. H. Cook, from the University of South Florida College of Marine Science, for her paper "Auditory brainstem response hearing measurements in free-ranging bottlenose dolphins (*Tursiops truncatus*)," and Jennifer Miksis-Olds, from the University of Rhode Island Graduate School of Oceanography, for her paper "Differential response by manatees to playbacks of sounds simulating approaching vessels."

ABTC sponsored three special sessions at the Vancouver meeting. Mardi Hastings organized and chaired a session on "Methodology for Measurements of Auditory Evoked Potentials in Aquatic Animals." Bill Burgess and Susanna Blackwell chaired sessions on "Tools for Animal Bioacoustics: New Designs and Directions." Ann Bowles and Ed Walsh organized a session and panel discussion on "Behavioral Audiometric Methods in Animal Bioacoustics: The Search for Standards," which was co-sponsored with the Committee on Standards. This is the first of a series of special sessions related to standards issues, which AB will continue to sponsor for future meetings. AB co-sponsored two other sessions, one with Psychological and Physiological Acoustics, and the other with Acoustical Oceanography and Underwater Acoustics. All of these sessions were heavily attended. Standards issues continued to dominate discussions at the ABTC meeting, which was attended by over 50 people.

Randy J. Hill, from the University of South Florida College of Marine Science, won the Best Student Paper Award for his paper, "Ground-truthing evoked potential measurements against behavioral conditioning in the goldfish, *Carassius auratus*."

Peter Narins and Andrea Simmons are currently organizing the Second International Workshop on Acoustic Communication, to be held at UCLA in the summer of 2007. This will follow up on the highly successful First Workshop, held at the University of Maryland in 2003.

New members of ABTC, whose terms began at the Vancouver meeting and extend to 2008, are Kelly Benoit-Bird, Tecumseh Fitch, Dorian Houser, Lee Miller, Larry Pater, Hiroshi Riquimaroux, Peter Scheifele, and Edward Walsh. Thanks to outgoing members Bill Burgess, Jim Finneran, Adam Frankel, Darlene Ketten, Tom Norris, and Peter Tyack for their service.

The ABTC could not function without the dedication and energy of our members. Mardi Hastings is a member of the Executive Council. Ann Bowles is the AB representative to the Committee on Standards. Charles Greene serves on the Medals and Awards Committee, and Darlene Ketten serves on the Membership Committee. Dave Mellinger is our webmaster. Marla Holt is the AB representative to the Student Council. Ann Bowles and Kate Stafford served as the TPOM representatives for the San Diego and Vancouver meetings, respectively. AB members, including Ann Bowles, Sheryl Coombs, Richard Fay, Mardi Hastings, Art Popper, Aaron Thode, and Ed Walsh, remain active in public policy and Standards issues. Activities include efforts to draft a policy statement on potential consequences of anthropogenic ocean noise on marine animals, and organizing a series of peer-reviewed papers on standards for acoustic measurements.

AB continues to grow in membership, with 665 ASA members now indicating AB as their primary, secondary, or tertiary interest. We encourage all members of ASA to visit our web site: http://cetus.pmel.noaa.gov/bioacoustics.html.

ANDREA SIMMONS
Chair

Architectural Acoustics

The Technical Committee on Architectural Acoustics has had a busy and exciting year. At the Fall 2004 meeting in San Diego, TCAA sponsored a special Topical Meeting, organized by Ning Xiang and Rendell Torres, on "Spatial and Binaural Evaluation of Acoustics in Performing Art Spaces," which spanned two half-day sessions. Other TCAA special sessions and their organizers included "Integration of Synthesis Techniques and 'Acoustical' Music" (K. Anthony Hoover), "Coupled Volume Systems: Design Processes and Implications" (David T. Bradley, Ralph Muehleisen), "Implementation of Classroom Acoustics" (Michael Stinson, Lily Wang), and "Speech in Architectural Spaces—Both Intelligibility and Privacy" (Ken Roy, Peter Mapp). Robin Glosemeyer served as TCAA representative to the Technical Program Organizing Meeting for this conference.

At the Spring 2005 meeting in Vancouver, TCAA sponsored a special Topical Meeting, organized by Murray Hodgson and Lily Wang, on "Classroom Acoustics: The Research Perspective," which included three half-day sessions and an evening public forum. Other TCAA special sessions and their organizers included "Architectural Acoustics and the Green Building Movement" (Ralph Muehleisen, Murray Hodgson), "Possibilities and Problems with Auralizations of Classroom Acoustics" (Bob Coffeen), "Preserv-

ing Acoustical Integrity in the Course of Renovation" (Dan Raichel), and "Soundscapes from an Architectural Viewpoint" (Brigitte Schulte-Fortkamp). Kerrie Standlee and Lily Wang served as TCAA representatives to the Technical Program Organizing Meeting for this conference. A special thanks to Murray Hodgson who served as General Chair for this excellent joint meeting with the Canadian Acoustical Association.

TCAA has several active subcommittees who made reports this past year. The Newman Student Award Fund continues to be active, providing awards to students who excel in their study of architectural acoustics and supporting educators who develop new educational materials through the Schultz Grant; more information may be found on their website at www.newmanfund.org. Another excellent Student Design Competition was organized for the Vancouver meeting by Bob Coffeen, Byron Harrison, and Norm Philipp, with first honors going to Lauren Ronsse and Andy Miller from the University of Kansas. Bill Dohn is leading the planning of the 3rd Concert Hall Research Group Summer School to be held in Aspen, Colorado in August 2006. The Acoustics Product Safety Working Group, chaired by Nancy Timmerman, is organizing a special session on product safety for the upcoming ASA conference in Minneapolis.

A new joint subcommittee with the Technical Committee on Noise concerning Speech Privacy was initiated at the San Diego meeting, chaired by Greg Tocci, and has been actively meeting throughout the year. Another new initiative approved in San Diego is to offer a TCAA Student Paper Award at each conference, along the same lines as many of the other ASA Technical Committees. The first awards will be given to students who present at the Fall 2005 meeting in Minneapolis. Interest in forming another new subcommittee focused on integrating Acoustics in Green Building Design was expressed at the Vancouver meeting; Ralph Muehleisen is leading those efforts.

Greg Miller and Robin Glosemeyer are finalizing the publication "Another Two Decades of Drama Theaters," based on posters submitted at the Spring 2004 meeting in New York City. Work has begun on the next such project: "Another Two Decades of Worship Spaces," organized by Adam Foxwell and Rosemary Su, with a poster session scheduled for the Spring 2006 meeting in Providence, Rhode Island.

Classroom Acoustics continues to be an important topic that TCAA members are strongly advocating. Special sessions on this topic were held at both San Diego and Vancouver, and a unique brainstorming session at the San Diego meeting focused on how to continue with Classroom Acoustics outreach. The TCAA is very grateful to David Lubman and Lou Sutherland for their ongoing efforts in this regard.

Congratulations to TCAA members who became ASA Fellows or were award recipients this year. New Fellows include Bennett Brooks, Klaus Genuit, Murray Hodgson, K. Anthony Hoover, and Volker Mellert. Lily Wang received the 2005 R. Bruce Lindsay Award at the Vancouver meeting.

Many thanks to the following individuals for their efforts this past year in ASA on behalf of TCAA. Red Wetherill is our representative on the Medals and Awards Committee. Greg Tocci serves on the Membership Committee, and George Winzer is TCAA representative to the ASA Committee on Standards. David T. Bradley from the University of Nebraska is our Student Council representative. The Associate Editors in Architectural Acoustics are Mendel Kleiner and Ning Xiang for *JASA* and J. D. Quirt for *ARLO*. Steven Brown is a member on the Executive Council. And last, a special thanks to Tony Hoover for his outstanding service as TCAA Chair for the previous three years.

LILY M. WANG
Chair

Biomedical Ultrasound/Bioresponse to Vibration

At the San Diego meeting the Biomedical Ultrasound/Bioresponse to Vibration Technical Committee (BU/BV) held its annual topical meeting. The one-day full meeting was organized by Kendall Waters and Brent Hoffmeister and the topic was Ultrasound Characterization of Cancellous and Cortical Bone. The topical meeting brought experts from around the world and the talks covered the gamut from historical perspective through the state-of-the-art and looking forward to the challenges in the future. It was concluded with an enthusiastic discussion on the barriers and potential of ultrasound in characterizing bone. Also at the San Diego meeting there was a special session organized by Ronald Roy and Todd Murray on the

combined use of Light and Sound for Imaging in the Body. We had two further sessions of contributed talks. Our TPOM representative was Franco Curra.

Two BU/BV members were honored at the San Diego meeting. James Miller was awarded the Silver Medal of the ASA and he was introduced by Matthew O'Donnell. David Blackstock was the first ever recipient of the Student Council Mentor Award. At the Technical Committee meeting on Thursday we held an election for a new chair of BU/BV and Michael Bailey from the University of Washington, Seattle was duly elected.

At the spring meeting in Vancouver BU/BV had a very strong showing with sessions all day, every day. Thanks go to the special sessions organizers: Mike Bailey and Reiner Schultheiss—Tissue response to shock waves (unfortunately Reiner could not attend due to a herniated disc); Tom Royston—Audible-frequency medical diagnostic methods, including multi-mode techniques; Pierre Mourad and Carr Everbach—Therapeutic ultrasound for the brain; and Pierre Mourad and Mike Averkiou—Gene therapy and molecular imaging. In addition BU/BV had a major impact on the Celebration session for Wesley Nyborg (Uncle Wes) that had been organized by Lawrence Crum and Junru Wu. Our TPOM representative was Michael Bailey. There were 53 people in attendance at the technical committee meeting on Thursday evening.

At the spring meeting BU/BV holds a student paper competition and in Vancouver there were 16 entries. The first prize went to Adam Maxwell from University of Washington and the second prize to Parag Chitnis from Boston University. I would also like to recognize two other students with honorable mentions: Wayne Kreider from University of Washington and Heather Argadine from the Mayo Clinic. Thanks to all the judges for their sterling service to the TC.

At the Minneapolis meeting BU/BV will be hosting a Topical Meeting on Detecting HIFU Lesions which is being organized by Emad Ebbini. We will have two special sessions: Medical Applications of Time Reversal Acoustics by Armen Sarvazyan and Acoustic Radiation Force Methods for Medical Imaging and Tissue Evaluation by Mostafa Fatemi.

My term as BU/BV chair comes to an end with this report and it is my pleasure to introduce the new chair Michael Bailey. I would like to take this opportunity to thank all the people who have helped me with the activities of BU/BV over the past three years. The continued vibrancy of the technical committee is due to the enthusiasm of members who volunteer their service to BU/BV. It has been my privilege to work with you all.

ROBIN CLEVELAND
Chair 2002–2005

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, San Diego in the Fall and Vancouver in the Spring.

In San Diego, the EATC sponsored two special sessions and one session of contributed papers. The special sessions and their organizers were MEMS Microphones: Fabrication, Calibration, and Application to High Density Arrays—Allan J. Zuckerwar and Qamar A. Shams and Hearing Aids—Gary W. Elko and James G. Ryan. In addition, EATC co-sponsored several other sessions with the various technical committees. The Committee offers many thanks to Steven R. Baker, who served as the representative to the San Diego TPOM. The EATC met in San Diego on the evening of Tuesday November 16.

In Vancouver, the EATC sponsored one special session and two sessions of contributed papers. The special session topic and organizers were Underwater Acoustic Sensor Technologies—Thomas R. Howarth and Dehua Huang. The Committee offers many thanks to Stephen C. Thompson who served as EATC representative to the Vancouver TPOM. The EATC met in Vancouver on the evening of Thursday May 17.

The recipients of the Best Student Paper awards for Engineering Acoustics were, for San Diego, Stefan A. L. Stijlen, "Ultrasonic nondestructive inspection of interface defects in anisotropic fiber-metal-laminates," (4aEA7), and, for Vancouver, Andrew P. Medley, "Radiated fields of thin and flexible high frequency loudspeakers," (1pEAa5).

The 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 attend EATC meetings to be held at future meetings

of the Society.
KIM C. BENJAMIN
Chair

Musical Acoustics

2004–2005 was another eventful year for the Technical Committee on Musical Acoustics. In addition to organizing special sessions, members of our technical committee were also responsible for some special events at the two ASA meetings and organized a workshop for instrument builders in conjunction with the Vancouver meeting.

There were some excellent special technical sessions in musical acoustics at the San Diego meeting in November 2004. These included a session on pipe organ acoustics organized by Tom Rossing, a session of invited talks and poster presentations on Computer Music organized by Julius Smith, "Musical Instruments of the Asia-Pacific Region," organized by Jim Cottingham, and "Synthesis Techniques with Acoustical Music," co-sponsored by Musical Acoustics and organized by Tony Hoover and Dick Campbell of Architectural Acoustics (AA). In addition to these, our technical committee was one of the many co-sponsors of the Robert Young memorial session. Following the session on Asian instruments, Christopher Adler of the University of San Diego presented a concert on the *khaen*, the free-reed mouth organ of Laos and Northeast Thailand. The performance also included demonstrations of two single-pipe free reed instruments and a duet in which Christopher Adler was joined by Supeena Insee Adler playing the three-string Thai zither *jakay*. About 150 people attended the concert. A tour and demonstration of two significant pipe organs were arranged in which a busload of participants visited the 73 rank Spreckels outdoor pipe organ in Balboa Park and the 107 rank pipe organ in the First Methodist Church. Our tour guide was Lyle Blackinton, curator of the Spreckels organ and builder of the organ in the Methodist church.

The Vancouver meeting included four special sessions in musical acoustics, two of which were co-sponsored by other technical committees. Courtney Burroughs and Tom Rossing organized "Vibration of and Radiation from Musical Instruments" (joint with Structural Acoustics). Jonas Braasch and William Martens co-chaired a joint session, "Low Frequency Content in Music" (joint with AA and Psychological and Physiological Acoustics). A session on virtual musical instruments was organized by Julius Smith and Tamara Smyth, and "String Instrument Design and Construction" was organized by Tom Rossing and Chris Waltham. Special technical initiative funding was used to support a workshop on acoustics for instrument builders, which was held following the Friday morning session on string instruments. The workshop, organized by Tom Rossing and Chris Waltham, had Friday afternoon and Saturday morning sessions. Many of the invited speakers from the Friday morning special session served as workshop leaders.

In addition to our ongoing efforts to promote student involvement in musical acoustics through the student paper competition, representatives from Musical Acoustics continued to play an active role in the ASA Student Council. Andrew Morrison of Northern Illinois University completed his term on the Student Council at the Vancouver meeting. Brian Monson of Brigham Young University has been appointed as our new Student Council representative. There were 11 entries in the student paper competition in San Diego and 13 in Vancouver. The winner of the First Prize award at San Diego was Rolf Bader of the Musikwissenschaftliches Institut, Hamburg, Germany, who presented "Multiphonic model of split pressure-phase modes using the finite-element method." Second Prize went to Yumiko Sakamoto of Kyushu University for "On the acoustical design of the ears of flue organ pipes." First Prize at Vancouver was awarded to Patricio de la Cuarda from CCRMA, Stanford University, for "A physical model and experimental test-bed for real-time simulation of flute-like instruments." The Second Prize winner was Niels Werner Larsen of Oersted DTU, Technical University of Denmark, who presented "A variable passive low-frequency absorber."

Jim Cottingham served as the representative to the Technical Program Organizing Meetings for San Diego and Vancouver, assisted by Chris Waltham at Vancouver. Representation from Musical Acoustics on ASA committees during the year included Uwe Hansen on the Membership Committee and Ian Lindevald on the Committee on Standards. Jim Beauchamp generously served an additional year on the Medals and Awards Committee after the expiration of his term in the spring of 2004. We expect his successor to be named soon. Diana Deutsch, who will work mainly with papers in the area of music perception, has been appointed as a second *JASA* Associate Editor for Musical Acoustics. She joins Neville Fletcher who will continue to serve as Associate Editor.

The Vancouver meeting was a time of transition, as Jim Cottingham ended his two terms of service as Technical Committee Chair. Jim Beauchamp has been elected as the new Technical Committee Chair, to serve until 2008. Members appointed or reappointed this year to serve on the Technical Committee for terms expiring in 2008 include James W. Beauchamp (Chair), George A. Bissinger, Annabel J. Cohen, James P. Cottingham, Diana Deutsch, Uwe J. Hansen, Roger J. Hanson, Peter L. Hoekje, James H. Irwin, Ian M. Lindevald, Stephen E. McAdams, Gary P. Scavone, and Chris Waltham.
JAMES P. COTTINGHAM
Chair (1999–2005)

Noise

At the Fall 2004 meeting of the ASA in San Diego, the Technical Committee on Noise organized or co-organized several interesting and diverse special sessions. These were "Special Session to Honor the Contributions of Robert W. Young to Acoustics" (organized by Alan Marsh), "Propulsion/Airframe Aeroacoustics" (Joe Posey), "Implementation of Classroom Acoustics" (John Erdreich, David Lubman and Louis Sutherland), "Active Noise Reduction for Hearing Protection: Technology Opportunities and Assessment of Effectiveness" (William Murphy), and "Soundscapes and Sound Quality" (Brigitte Schulte-Fortkamp and Patricia Davies). Our appreciation goes to Alan Marsh for serving as the Noise representative at the Technical Program Organizing Meeting.

The Spring 2005 Vancouver meeting was also host to an interesting array of special sessions organized or co-organized by Noise. These sessions were "Topical Meeting on Classroom Acoustics—The Research Perspective" (three half-days of technical presentations plus a public forum, organized by Murray Hodgson and Lily Wang), "Urban Noise Propagation" (Jian Kang, Ning Xiang, and Keith Wilson), "Progress in Managing Occupational Noise" (Hugh Davies and Murray Hodgson), "Preserving Acoustical Integrity in the Course of Renovation" (Dan Raichel), "Noise Control Feasibility: Technical, Legal and Economic Issues" (John Seiler and Angelo Campanella), "Characterization of Acoustical Materials" (Brandon Tinianov), and "Workshop on Methods for Community Noise and Annoyance Evaluation" (Brigitte Schulte-Fortkamp and Bennett Brooks). Jerry Lilly was the Noise representative at the Technical Program Organizing Meeting.

Classroom acoustics continues to be a major focus of the Noise Committee. The efforts of Lou Sutherland, David Lubman, Murray Hodgson, and others to promote good acoustics for classrooms are appreciated. A new subcommittee, joint between Noise and Architectural Acoustics and chaired by Greg Tocci, has been formed to address speech privacy in health care buildings.

The contributions of several Noise people were recognized at these two meetings. At the San Diego meeting, we had two winners of the Noise Young Presenter award, Beatriz Pinto for her talk "S-shaped glass also stands for soundless. Sound insulation measurements and implication for building practice" and Cesare Hall for his talk "Engine configurations for the Silent Aircraft." In Vancouver, the Noise Young Presenter award went to Connor Duke for his talk "Multi-channel active noise control on an axial fan using variable loads." New Fellows of the Society include Klaus Genuit, Tony Hoover, Bennett Brooks, and Murray Hodgson. Lily Wang was presented with the 2005 R. Bruce Lindsay Award. The winner of the 2005 Helmholtz-Rayleigh Interdisciplinary Silver Medal, in Noise and Physical Acoustics, is Gilles Daigle "for contributions to understanding the micrometeorology, topography, and ground properties on outdoor sound propagation."

Finally, the efforts of several volunteers should be recognized. Nancy Timmerman is the Noise representative on the Medals and Awards Committee, John Erdreich is our representative on the Membership Committee, and Richard Peppin is our representative on the ASA Committee on Standards. Our Student Council representative was Brian Anderson until the end of the San Diego meeting. Connor Duke is our new representative to Student Council and, following the practice initiated a few meetings ago, the secretary for the TC-Noise meetings. The Noise web page (http://www.nonoise.org/quietnet/tcn/) is maintained by Les Blomberg. Ken Cunefare completed his stint as coordinator for the Noise Young Presenter

Awards at the San Diego meeting. Ralph Muehleisen has taken over this role and he handled the coordination for the Vancouver meeting. Serving as an Associate Editor for *ARLO* is Mike Stinson and as *JASA* Associate Editors are Ken Cunefare, Keith Wilson, and Allan Zuckerwar.

MICHAEL R. STINSON
Chair

Physical Acoustics

It is my distinct pleasure to reflect on the past year's activities in Physical Acoustics. We have been redefining ourselves over the past few years. The "hot" fields of sonoluminescence and thermoacoustics have taken a back seat as funding levels have decreased (new discoveries may yet help stimulate more funding). New areas are also emerging (examples include monitoring the Earth, condensed matter, medical acoustics, and Homeland Security research). This is reflected in recent and upcoming special sessions: e.g., Infrasound, Sounds Speeds in Condensed Matter, Shock Wave Therapy, and Applications of Acoustic Radiation Pressure.

After the huge New York meeting, it would have been reasonable to have a small Physical Acoustics turnout at the Fall San Diego meeting, but that was not the case. We had sessions Monday through Thursday. Both special sessions had standing-room only crowds. The session on Sonoluminescence, Sonochemistry, and Sonofusion was exhilarating. Invited speakers came from all over the world, and from academia and industry, to share their most recent discoveries. It is worth noting that from the beginning, when single-bubble sonoluminescence was discovered (pre-1990), the leaders in this important field have had their roots in Physical Acoustics and the ASA. The special session on Applications of Acoustic Radiation Pressure was also standing room only. This session covered a wide range of topics, from basic physics to applications. We wish to thank the organizers for these sessions, Tom Matula and Greg Kaduchak. In addition to the special sessions, there was sufficient interest in the San Diego meeting (perhaps the weather had some role!) to generate several contributed sessions, including Sound Propagation in Ground, and in the Atmosphere, Materials Characterization, and Time Reversal.

The Spring meeting was held in Vancouver, BC. The meeting was larger than expected, and Physical Acoustics sessions ran all week, with some unavoidable concurrent sessions. A major special session was organized on Infrasound: A New Frontier in Monitoring the Earth, organized by Michael Hedlin. It is this author's impression that significant gaps remain between data gathering/reduction/analysis and predictive capabilities. This rapidly growing field is in dire need of Physical Acousticians (graduate students take note!). During the open TC meeting, Hank Bass even suggested that funding for this area may be increasing. This is good news for a TC that has taken some hits with decreased funding levels from the Office of Naval Research (ONR) over the past several years.

We also had special sessions in Scattering from Noncanonical Objects, organized by Steve Kargl; Stochastic Scattering of Elastic and Seismic Waves, organized by Richard Weaver, and another in the line of Physical Acoustics Living Legends sessions, organized by Larry Crum for Wes Nyborg. This session was co-run with Biomedical Ultrasound/Bioresponse to Vibration (B&B), as Wes's contributions significantly influenced the growth of B&B. There has been a lot of positive feedback on continuing these topical Living Legend sessions. Physical Acoustics also shared the stage during the "Hot Topics" session. Thanks to Albert Migliori for taking on the job of presenting our hot topics.

Physical Acoustics takes special pride in acknowledging our most recent award recipients: David Blackstock received the inaugural Student Council Mentoring Award in San Diego. David is continuously promoting and advancing student participation in science and the ASA. The Helmholtz-Rayleigh Interdisciplinary Silver Medal in Noise and Physical Acoustics went to Gilles Daigle. Gilles received the award at the Vancouver meeting. The Rossing Prize in Acoustics Education went to Allan Pierce at the San Diego meeting. Allan's recognition did not stop there; he was awarded the Gold Medal in Vancouver. Congratulations to all our award recipients. A job well done!

I would like to take this opportunity to advertise for the Physical Acoustics Summer School (PASS), held every two years. The next school will be held in June of 2006. This highly successful school takes students on a week-long intellectual journey through many areas touched by physical acoustics. Although funding for the school from ONR is gone, the National Center for Physical Acoustics and Penn State are trying to bridge the gap until new funding can be obtained.

I want to also take the time to acknowledge members of Physical Acoustics who volunteer their services on various committees: Robert "Bonzo" Keolian serves on the Medals and Awards committee; Steve Garrett serves on the Membership Committee; Phil Marston serves on the Books 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. It is also a pleasure to acknowledge Physical Acoustics members who are active in leadership roles in the ASA. Mark Hamilton is now past vice president of the ASA, the editor-in-chief is Allan Pierce, and Anthony Atchley was just elected President! Congratulations Anthony!

Finally, I wish to humbly thank the Physical Acoustics community for entrusting me to serve as Chair over the past three years. It has been a privilege and honor to serve you. I especially would like to thank the more experienced members for their advice and encouragement over the years. Now, I pass the crown to our new Chair, Murray Korman. Have a great time, Murray; I did!

THOMAS J. MATULA
Chair 2002–2005

Psychological and Physiological Acoustics

Reflecting Psychological and Physiological Acoustics' (P&P) strategy of emphasizing participation at ASA's spring meetings, the 148th Meeting of the Acoustical Society of America, the fall meeting in San Diego, was sparsely attended by the members of P&P. Nonetheless, P&P co-sponsored two special sessions: "Implementation of Classroom Acoustics I and II." In keeping with P&P policy, we will not be a primary sponsor of special sessions at the Minneapolis meeting in the fall, and P&P has asked that calls for papers provided by ASA include a brief explanation of our policy of deemphasizing the fall meetings.

In stark contrast, the spring meeting in Vancouver was buzzing with activity. There were eight 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 eight sessions, three were special sessions including one named session. We extend our thanks to the organizers of the special sessions: Bertram Scharf and Rhona P. Hellman (Psychoacoustics: In Memory of Søren Buus), Roy D. Patterson and Toshio Irino (Size Information in Speech and Animal Calls), and Lawrence L. Feth and Robert A. Fox (Current Issues in Auditory Spectral Integration). P&P also contributed as co-sponsor of six additional special sessions. Neal Viemeister accomplished the challenge of organizing the P&P sessions for the Vancouver meeting, a task for which we are most appreciative.

The P&P open meeting included an update of the Student Council report by the outgoing student representative, Joshua Bernstein. The P&P Associate Editors provided overview of their efforts, and urged the members of ASA to provide reviews well before the Peer X Press system generates an "overdue notice." The goal is to provide reviews within four weeks of receiving the submitted manuscript. We thank the P&P Associate Editors, Brenda L. Lonsbury-Martin, William P. Shofner, John H. Grose, Gerald D. Kidd, Armin Kohlrausch, Robert A. Lutfi, and Andrew J. Oxenham, for their efforts on our behalf. Finally, there was discussion concerning the need to continue to provide nominations of Fellows to the Acoustical Society of America. We express our thanks to John Culling for organizing the ballot for election to the Technical Committee, and thanks to members of the P&P community who so willingly added their names to the ballot.

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

VIRGINIA M. RICHARDS
Chair, 2002–2005

Speech Communication

The Speech Technical Committee (STC) supports the activities, meetings, publications, etc. for the largest technical area in the Society. The current members of the Committee are Jean Andruski, Patrice Beddor,

Lynne Bernstein, Catherine T. Best, Ocke-Schwen Bohn, Suzanne E. Boyce, Ann Bradlow, Dani Byrd, Roger Chan, Melissa Epstein, Robert Fox, Alexander Francis, Bruce Gerratt, Kenneth Grant, Helen Hanson, James Hillenbrand, Diane Kewley-Port, Jody E. Kreiman, Anders Lofqvist, Andrew Lotto, Benjamin Munson, Terrence Nearey, Peggy Nelson, Douglas O'Shaughnessy, Dwayne Paschall, Joseph Perkell, Astrid Schmidt-Nielsen, Brad Story, and Gary G. Weismer. Other *ex officio* members include Abeer Alwan, Shrikanth Narayanan, and Fredericka Bell-Berti, our representatives to the Membership Committee, ASACOS, and Medals and Awards, respectively; Jennell Vick is the Student Council representative. The continuing Associate Editors for speech are Alexander Francis, Kenneth Grant, Anders Lofqvist, and Douglas O'Shaughnessy; two newly appointed associate editors are Brad Story and Joan Sussman.

We are grateful for the help of our members this year: (1) Dani Byrd, Rob Fox, and Diane Kewley-Port, who revised the PACS numbers for the speech section. The revisions will take place in September 2005. (2) Our paper sorters, who arrange the technical programs: Jody Kreiman and Dani Byrd in San Diego and Terry Nearey and Richard Wright in Vancouver. Special thanks to Bryan Gick who served on the Vancouver local organizing committee. (3) Coordinators for student judging: Alex Francis, Amee Shah, and Ben Munson. We also are very happy to welcome three new fellows to the ASA from the speech communication TC: Catherine Best, Patricia Keating, and Robert Remez.

Student Activities

The Committee sponsored two student activities at each meeting, a competition with a cash award for best student presentation and an evening reception. The evening reception, which is joint with other technical committees, is intended to allow students to meet more senior ASA members informally. The reception was well attended. The student papers were judged by STC members and the winners were awarded $300 for first prize and $200 for second prize. In San Diego the first and second place winners were Jason Tourville of Boston University and Jerry Liu of the University of Southern California.

Special Sessions, Special Talks, and Workshops

Speech Communication has sponsored many interesting special sessions during the past year. In San Diego Sorin Dusin and Lawrence Rabiner organized a special session "Fifty years of progress in speech communication: Honoring the contributions of James L. Flanagan." In addition a session on "Foreign-accented speech: Production, perception and applications" was organized by Amee Shah. In Vancouver a considerable number of special sessions were held. Ruth Litovsky and Mario Svirsky coordinated the session "Communication abilities of congenitally deaf children" held jointly with Psychological and Physiological Acoustics (P&P), which considered cochlear implants, hair cell regeneration and related subjects. Sonya Bird organized "Indigenous languages of North America," which reduced the deficit of published research in this area. Linda Polka, Ocke-Schwen Bohn, and James Hillenbrand sponsored the session "Vowel systems and language learners." "Spectral integration of complex sounds and speech" was organized by Larry Feth, Eva Jacewicz, and Rob Fox, sponsored by P&P and co-sponsored by Speech. In addition, the "First ASA Workshop on Second Language Speech Learning" was held at Simon Fraser University immediately prior to the ASA meeting in Vancouver. The workshop, organized by Murray Munro and Ocke-Schwen Bohn, was attended by over 100 participants and was very highly praised. Finally, we appreciate the efforts of two members who gave excellent and informative presentations at the Vancouver meeting: the hot topics presentation "Imaging for understanding speech communication: Advances and challenges" by Shri Narayanan and the tutorial presentation "Tutorial lecture on automatic speech recognition" by Carole Espy-Wilson.
MAUREEN STONE
Chair

Structural Acoustics and Vibration

Many of the technical activities by members of the Structural Acoustics and Vibration Technical Committee (SAVTC) has been supported by funding by the US Navy on vibrations of, acoustic radiation from, and

self-noise in marine vessels. With Navy funding in these areas decreasing, SAVTC is facing new challenges, in addition to maintaining activities in past areas of interest, which is reflected in the SAVTC activities during this year. At the San Diego meeting in November 2004, there were three special technical sessions. Karl Grosh organized and chaired a session on "Structural Acoustics in MEMS." A session on "Surface Ship Vibration and Noise" was organized and chaired by Ray Fischer. Finally, Dan Russell organized and chaired a session on "Vibrations in Sports Equipments." In addition, there were two technical sessions on analysis and measurement methods. At the Vancouver meeting in May 2005, there were two special all-day technical sessions. David Feit organized and chaired a session on "Approximate Techniques for Fluid Loading Effects." Courtney Burroughs and Tom Rossing, from the Musical Acoustics Technical Committee, organized and chaired a session on "Vibration and Acoustic Radiation from Musical Instruments." In addition, there were SAVTC technical sessions on vibration and acoustic radiation.

This was the year for SAVTC to present a Hot Topics paper. At the San Diego meeting, Courtney Burroughs, Dean Capone, and Karl Grosh presented "Hot topics in structural acoustics; New directions and challenges." There was one new Fellow from the SAVTC. Two new Associate Editors for JASA were appointed. Andy Norris "retired" as Associate Editor after several years of service. Student paper award winners at the San Diego meeting were Matt Allen for paper 4pSA1 and Elizabeth Magluila for paper 2aSA11. In the Vancouver meeting, the two student paper winners were Jerome Pinsonnault for paper 5aSA2 and Dany Francoeur for paper 5aSA1. The student representative for SAVTC to the very active Student Council was Lance Locey (Penn State). He was replaced by Micah Shepard (Brigham Young) at the Vancouver meeting. The SAVTC representative to the San Diego TPOM was Courtney Burroughs and to the Vancouver TPOM, Claudio Bulfone. A special thanks to Claudio, who is not a member of SAVTC, for filling in when no one from SAVTC could attend the Vancouver TPOM.
COURTNEY B. BURROUGHS
Chair

Underwater Acoustics

The year began with preparation for the Fall 2004 meeting in San Diego, for which we thank Peter Gersoft and Peter Worcester for representing the Underwater Acoustics Technical Committee (UWTC) at the Technical Program Organizing Meeting (TPOM). Also, leading up to the San Diego meeting was the prominent cover-article of the October 2004 edition of *Physics Today*, entitled "Shallow-Water Acoustics" by William A. Kuperman and James F. Lynch.

In San Diego the UWTC sponsored special sessions entitled "Very High Frequency [O(100) kHz] Boundary Interaction" that was organized by Gary Heald and "Long Range Acoustic Propagation" that was organized by Peter Worcester. There were 16 papers submitted to the student paper competition for the San Diego meeting under the auspices of the UWTC. Elizabeth T. Kusel (Rensselaer Polytechnic Institute) was awarded first prize for her paper entitled "A single-scattering solution that handles large contrasts across interfaces," and Alexander O. MacGillivray (University of Victoria) received second prize for his presentation entitled "An airgun array source signature model for environmental impact assessments."

Gerald D'Spain presented a Tutorial Lecture on Ocean Noise and Marine Mammals at the San Diego meeting. Closely related to this topic, discussions at the San Diego venue commenced on an ASA Policy Statement on the "Potential Consequences of Anthropogenic Ocean Noise on Marine Animals," by the ASA Panel on Public Policy (PoPP), Working Group on Ocean Noise (James H. Miller and Darlene Ketten, Co-Chairs).

UWTC members Michael Ainslie, Robert Gragg, Phillippe Roux, and Manell Zakharia were recognized as new Fellows of the ASA, and the UWTC also recognizes the contributions of Walter Munk (ASA Honorary Fellowship), Van Holliday (Silver Medal in Acoustical Oceanography), and John Bouyoucos (Silver Medal in Engineering Acoustics), all of whom were recognized in San Diego. Finally, all the Technical Committee Chairs participated in the "Vision 2010 Conference" held at the Scripps Institution of Oceanography the Sunday before the San Diego meeting, with initial recommendations of this conference reported in the Winter 2005 edition of *Echoes*.

The Spring 2005 meeting in Vancouver, BC, held jointly with the Canadian Acoustical Association, was by all measures a fine success. Peter

Dahl represented the UWTC at the January TPOM, and the UWTC sponsored special sessions entitled "Acoustic Interaction with Ocean Boundaries: Single Bounce Measurements," which was organized by Charles Holland, and "Frequency Dependence of Sound Speed and Attenuation in Marine Sediments," which was organized by Gopu Potty and Peter Dahl. There were 21 papers submitted to the student paper competition for the Vancouver meeting under the auspices of the UWTC. Alexander O. MacGillivray (University of Victoria) received first prize for his presentation entitled "An acoustic modeling study of airgun noise from seismic surveys performed offshore British Columbia," and Jie Yang (Georgia Institute of Technology) received second prize for her work entitled "Influence of internal waves on vertical coherence of sound propagation in the East China Sea."

The ASA student paper awards in Underwater Acoustics receive contributions from other, related Technical Committees that do not offer this award; as a result both the paper count and competition is high. Coordinating the judging of these papers is a significant task (we seek at least two, ideally three independent evaluations for each paper), and the UWTC thanks Kyle Becker, who has carried out this coordination role on an *ad hoc* basis for the last three meetings. In Vancouver, the UWTC Chair proposed formalizing this role (*student judge coordinator* with position to last one year; Kathleen Wage has volunteered for this position that will commence with the Minneapolis meeting. We also note the graduation of UWTC Student Council representative Tom Weber (Applied Research Laboratory, Penn State University), who will be replaced by Andy Ganse (University of Washington). Finally the UWTC recognizes the contributions of Svein Vagle who received the Medwin Prize in Acoustical Oceanography in Vancouver.
PETER H. DAHL
Chair

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.

2005

17–21 Oct.   150th Meeting of the Acoustical Society of America joint with Noise-Con, Minneapolis, MN [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: http://asa.aip.org].

27–29 Oct.   5th International Symposium on Therapeutic Ultrasound, Boston, MA [www.istu2005.org; E-mail: info@istu2005.org].

2006

9–12 Feb.   46th Annual Convention, Illinois Speech-Language-Hearing Association, Rosemont, IL [ISHA, 230 E. Ohio St., Suite 400, Chicago, IL 60611-3265; Tel.: 312-644-0828; Fax: 315-644-8557; Web: www.aishil.org].

16–18 Feb.   31st Annual Conference, National Hearing Conservation Association, Tampa, FL [NHCA, 7995 E. Prentice Ave., Suite 100 East, Greenwood Village, CO 80111-2710; Tel: 303-224-9022; Fax: 303-770-1614; E-mail: nhca@gwami.com; WWW: www.hearingconservation.org].

6–9 June   151st Meeting of the Acoustical Society of America, Providence, RI Island [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: http://asa.aip.org].

17–21 Sept.   INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org ⟨http://www.interspeech2006.org/⟩]

28 Nov.–2 Dec.   152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, HI [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: http://asa.aip.org].

2008

28 July–1 Aug.   9th International Congress on Noise as a Public Health Problem (Quintennial Meeting of ICBEN, The International Commission on Biological Effects of Noise), Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, P. O. Box 1609, Groton, CT 06340-1609, Tel.: 860-572-0680; Web: www.icben.org; E-mail: icben2008@att.net].

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).
Volumes 105–114, 1999–2003: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 616. Price: ASA members $50; Nonmembers $90 (paperbound).

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 2006.

Technical Committees 2005–2006

Acoustical Oceanography
N. Ross Chapman, *Chair* to 2007

Term to 2008
Daniela Di Iorio
Gerald L. D'Spain
Gary J. Heald
Jean-Pierre Hermand
David P. Knobles
Timothy G. Leighton
James H. Miller
Daniel Rouseff
Emmanuel K. Skarsoulis
Jerome A. Smith
Dajun Tang

Term to 2007
Kyle M. Becker
Grant B. Deane
Christopher Feuillade
Peter Gerstoft
Oleg A. Godin
John K. Horne
Bruce M. Howe
Anthony P. Lyons
Ralph A. Stephen
Kathleen E. Wage
Peter F. Worcester

Term to 2006
Mohsen Badiey
Paul E. Chandler
Dezhang Chu
Kerry W. Commander
Christian de Moustier
Stan E. Dosso
Andone C. Lavery
Zoi-Heleni Michalopoulou
Jeffrey A. Nystuen
David R. Palmer
Simon D. Richards
Martin Siderius

Ex officio:
James F. Lynch, member of Medals and Awards Committee
Mohsen Badiey, member of Membership Committee
Anthony P. Lyons, member of ASACOS

Animal Bioacoustics
Andrea M. Simmons, *Chair* to 2006

Term to 2008
Whitlow W. L. Au
Kelly J. Benoit-Bird
W. Tecumseh Fitch
Dorian S. Houser
Lee A. Miller

Larry L. Pater
Hiroshi Riquimaroux
Peter M. Scheifele
Edward J. Walsh

Term to 2007
Sheryl L. Coombs
Richard R. Fay
Edmund R. Gerstein
Seth S. Horowitz
Cynthia F. Moss
James A. Simmons
Jeannette A. Thomas

Term to 2006
Ann E. Bowles
John R. Buck
Christopher W. Clark
William C. Cummings
Charles R. Greene
Mardi C. Hastings
D. Vance Holliday
David A. Mann
Sam H. Ridgway

Ex officio:
Charles R. Greene, member of Medals and Awards Committee
Darlene R. Ketten, member of Membership Committee
Ann E. Bowles, member of ASACOS

Architectural Acoustics
Lily M. Wang, *Chair* to 2007

Term to 2008
Wolfgang Ahnert
Christopher N. Blair
John S. Bradley
Christopher N. Brooks
Angelo J. Campanella
Quinsan Ciao
Robert C. Coffeen
Peter D'Antonio
Felicia M. Doggett
William Dohn
Timothy J. Foulkes
Richard D. Godfrey
Tyrone Hunter
Clare M. Hurtgen
J. Christopher Jaffe
Mendel Kleiner
Jeff P. Kwolkoski
Brad W. Lewis
Stephen J. Lind
David Lubman
Ralph T. Muehleisen
Michael T. Nixon
Boaz Rafaely
Daniel R. Raichel
Jack E. Randorff
Jonathan Rathsam
H. Stanley Roller
Steven R. Ryherd
B. Schulte-Fortkamp
Noral D. Stewart
Michael Vorländer
Ning Xiang

Term to 2007

Warren E. Blazier
Joseph F. Bridger
Norm Broner
Bennett M. Brooks
Steven M. Brown
Todd A. Busch
Richard H. Campbell
F. M. del Solar Dorrego
Erin L. Dugan
M. David Egan
Jesse J. Ehnert
Michael Ermann
Adam R. Foxwell
Ronald R. Freiheit
Klaus Genuit
Matthew V. Golden
Kenneth W. Good, Jr.
Brad N. Gover
Mark A. Holden
K. Anthony Hoover
Jerald R. Hyde
David W. Kahn
Martha M. Larson
Gary S. Madaras
Benjamin E. Markham
Charles T. Moritz
David L. Moyer
Edward T. Nykaza
Paul B. Ostergaard
Dennis A. Paoletti
Stephen W. Payne
Benjamin C. Seep
Neil A. Shaw
Ryan G. Sieler
Rose Mary Su
Jeff D. Szymanski
Richard H. Talaske
Michelle Vigeant
Ewart A. Wetherill
George E. Winzer
Michael R. Yantis

Term to 2006

Nils-Ake Andersson
Leo L. Beranek
Sergio Beristain
Leslie D. Blomberg
Erica Bowden
David T. Bradley
David Braslau
Todd L. Brooks
Courtney B. Burroughs
Paul T. Calamia
Alexander U. Case
William J. Cavanaugh
Dan Clayton
Elizabeth A. Cohen
Chris L. Davies
Damian Doria
John Erdreich
Robin S. Glosemeyer
Robert D. Hellweg
Murray R. Hodgson
Ian B. Hoffman

Jin Jeon
James A. Johnson
Bertram Y. Kinzey, Jr.
Iraklis E. Lampropoulos
Jerry G. Lilly
Peter A. Mapp
David E. Marsh
Gregory A. Miller
Hideo Miyazaki
Richard A. Moscoso-Bullon
Matthew A. Nobile
Bruce C. Olson
Richard J. Peppin
Scott D. Pfeiffer
James E. Phillips
Joseph Pope
Roy L. Richards
Carl J. Rosenberg
Kenneth P. Roy
Kevin P. Shepherd
Yasushi Shimizu
Gary W. Siebein
Abigail E. Stefaniw
Christopher A. Storch
Jason E. Summers
Louis C. Sutherland
Jiri Tichy
Nancy S. Timmerman
Brandon D. Tinianov
Gregory C. Tocci
Rendell R. Torres
Alfred C. C. Warnock
George P. Wilson

Ex officio:
Ewart A. Wetherill, member of Medals and Awards
Gregory C. Tocci, member of Membership Committee
George E. Winzer, member of ASACOS

Biomedical Ultrasound/Bioresponse to Vibration
Michael R. Bailey, *Chair* to 2008

Term to 2008

John S. Allen
Charles C. Church
Gregory Clement
Floyd Dunn
E. Carr Everbach
Christy K. Holland
R. Glynn Holt
Elisa E. Konofagou
Subha Maruvada
Wesley L. Nyborg

Term to 2007

Constantin C. Coussios
Diane Dalecki
J. Brian Fowlkes
Kullervo H. Hynynen
T. Douglas Mast
Thomas J. Matula
Robert J. McGough
Douglas L. Miller
Ronald A. Roy
Thomas J. Royston
Kendall R. Waters

Pei Zhong

Term to 2006
Shira L. Broschat
Ibrahim M. Hallaj
Joie P. Jones
Inder R. Makin
Michael V. Scanlon
Thomas L. Szabo
Janet M. Weisenberger
Junru Wu
Suk Wang Yoon
Ex officio:
Lawrence A. Crum, member of the Medals and Awards
Committee
E. Carr Everbach, member of the Membership Committee
and member of ASACOS

Engineering Acoustics
Kim C. Benjamin, *Chair* to 2006

Term to 2008
Steven R. Baker
David A. Brown
Stephen C. Butler
Robert D. Corsaro
Stephen E. Forsythe
Brian H. Houston
W. Jack Hughes
Robert M. Koch
L. Dwight Luker
Arnie L. Van Buren
Kenneth M. Walsh
Daniel M. Warren
Joseph F. Zalesak

Term to 2007
Mahlon D. Burkhard
James Christoff
Fernando Garcia-Osuna
Charles S. Hayden
Dennis F. Jones
Jan F. Lindberg
Yushieh Ma
Elizabeth A. McLaughlin
Alan Powell
Roger T. Richards
Kenneth D. Rolt
Neil A. Shaw
James F. Tressler

Term to 2006
Stanley L. Ehrlich
Gary W. Elko
Robert D. Finch
Guillermo C. Gaunaurd
Thomas R. Howarth
Dehua Huang
Sung Hwan Ko
Theodore J. Mapes
Victor Nedzelnitsky
James M. Powers
P. K. Raju
Yongrae Roh
Ahmet Selamet
Stephen C. Thompson

James E. West
George S. K. Wong

Ex officio:
Mahlon D. Burkhard, member of Medals and Awards
Committee and member of ASACOS
Thomas R. Howarth, member of Membership Committee

Musical Acoustics
James W. Beauchamp, *Chair* to 2008

Term to 2008
George A. Bissinger
Annabel J. Cohen
James P. Cottingham
Diana Deutsch
Neville H. Fletcher
Roger J. Hansen
Uwe J. Hansen
Peter L. Hoekje
James H. Irwin
Ian M. Lindevald
Stephen E. McAdams
Gary P. Scavone
Chris E. Waltham

Term to 2007
R. Dean Ayers
Judith C. Brown
Courtney B. Burroughs
John R. Buschert
Thomas M. Huber
Bozena Kostek
Barry Larkin
Daniel O. Ludwigsen
Thomas D. Rossing
David B. Sharp
Julius O. Smith
William J. Strong

Term to 2006
Anders G. Askenfelt
Xavier Boutillon
Jonas Braasch
Murray D. Campbell
Rene E. Causse
Antoine J. Chaigne
Nicholas J. Giordano
J. M. Harrison
William M. Hartmann
James M. Pyne
Daniel A. Russell
Punita G. Singh
Paul A. Wheeler
Shigeru Yoshikawa

Ex officio:
Uwe J. Hansen, member of Medals and Awards Committee
James P. Cottingham, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise
Michael R. Stinson, *Chair* to 2006

Term to 2008
Elliott H. Berger
Ann E. Bowles
Frank H. Brittain

Steven M. Brown
Mahlon D. Burkhard
Robert D. Collier
Lawrence S. Finegold
Samir N. Y. Gerges
Richard D. Godfrey
Matthew V. Golden
Murray R. Hodgson
Jerry G. Lilly
Stephen J. Lind
David Lubman
George A. Luz
Matthew A. Nobile
Richard J. Peppin
Robert A. Putnam
Jack E. Randorff
Stephen I. Roth
Paul D. Schomer
Michelle E. Swearingen
Nancy S. Timmerman
Brandon D. Tinianov
Gregory C. Tocci
Lily M. Wang

Term to 2007
Martin Alexander
Brian E. Anderson
Keith Attenborough
John P. Barry
Leo L. Beranek
Arno S. Bommer
James O. Buntin
John C. Burgess
Jim R. Cummins
Kenneth A. Cunefare
Paul R. Donavan
Ronald R. Freiheit
Klaus Genuit
Michael L. Gross
David C. Haser
Gerald C. Lauchle
George C. Maling
Thomas R. Norris
John P. Seiler
Noral D. Stewart
Louis C. Sutherland
Jiri Tichy
D. Keith Wilson
Ning Xiang

Term to 2006
Jorge P. Arenas
Sergio Beristain
Leslie D. Blomberg
Bennett M. Brooks
Robert D. Bruce
Angelo J. Campanella
William J. Cavanaugh
Gilles A. Daigle
Patricia Davies
Jesse J. Ehnert
Donna A. Ellis
Tony F. W. Embleton
John Erdreich
David J. Evans

William M. Hartmann
Robert D. Hellweg
David K. Holger
Seong-Woo Kang
William W. Lang
Richard H. Lyon
Alan H. Marsh
Ralph T. Muehleisen
William J. Murphy
Kenneth J. Plotkin
Joseph Pope
Daniel R. Raichel
Brigitte Schulte-Fortkamp
Ben H. Sharp
Kevin P. Shepherd
Scott D. Sommerfeldt
Kerrie G. Standlee
George S. K. Wong
James C. Yu
Ex officio:

Nancy S. Timmerman, member of Medals and Awards
Committee
John Erdreich, member of Membership Committee
Richard J. Peppin, member of ASACOS

Physical Acoustics
Murray S. Korman, *Chair* to 2008

Term to 2008
Robert T. Beyer
Robin O. Cleveland
Lawrence A. Crum
Kenneth E. Gilbert
Robert A. Hiller
R. Glynn Holt
Bart Lipkens
Thomas J. Matula
Ralph T. Muehleisen
Harry Simpson
John S. Stroud
Preston S. Wilson

Term to 2007
David T. Blackstock
David A. Brown
John A. Burkhardt
Kerry W. Commander
Bruce C. Denardo
Logan E. Hargrove
D. Kent Lewis
Julian D. Maynard
George Mozurkewich
Lev A. Ostrovsky
Andrea Prosperetti
Neil A. Shaw
Victor W. Sparrow
Richard Stern
Roger M. Waxler

Term to 2006
Henry E. Bass
Yves H. Berthelot
James P. Chambers
David I. Havelock
Philip L. Marston

Philip S. Spoor
Larry A. Wilen
D. Keith Wilson
Evgenia A. Zabolotskaya

Ex officio:
Robert M. Keolian, member of Medals and Awards
Committee
Steven L. Garrett, member of Membership Committee
Sameer I. Madanshetty, member of ASACOS

Psychological and Physiological Acoustics
Elizabeth A. Strickland, *Chair* to 2008

Term to 2008
Michael A. Akeroyd
Alain de Cheveigne
Brent W. Edwards
John H. Grose
Gerald D. Kidd, Jr.
Armin Kohlrausch
Brenda L. Lonsbury-Martin
William P. Shofner
Lynne A. Werner

Term to 2007
David A. Eddins
Lawrence L. Feth
Hedwig E. Gockel
Jennifer Lentz
Mario A. Ruggero

Term to 2006
Michelle L. Hicks
Lynne Marshall
Christopher J. Plack
Donal G. Sinex
Magdalena Wojtczak

Ex officio:
Joseph W. Hall, member of the Medals and Awards
Committee
Leslie R. Bernstein, member of Membership Committee
Peggy B. Nelson, member of ASACOS

Signal Processing in Acoustics
Charles F. Gaumond, *Chair* to 2006

Term to 2008
Frank A. Boyle
David H. Chambers
Joe A. Clark
R. Lee Culver
David J. Evans
David M. Fromm
Howard A. Gaberson
David I. Havelock
Jean-Pierre Hermand
George E. Ioup
Juliette Ioup
Matti A. Karjalainen
Sean K. Lehman
Lance L. Locey
Zoi-Heleni Michalopoulou
Brian B. Monson
Joseph Pope
Leon H. Sibul
Randall W. Smith

Krykidos Tsiappoutas
James E. West
Gary R. Wilson
George S. K. Wong
Ning Xiang

Term to 2007
Max Deffenbaugh
Alireza A. Dibazar
Gary W. Elko
Alan W. Meyer
Daniel J. Sinder
David C. Swanson
Robert C. Waag
Preston S. Wilson
Lixue Wu

Term to 2006
James V. Candy
William M. Carey
Geoffrey S. Edelson
Stanley L. Ehrlich
Brian Ferguson
Paul J. Gendron
Peter Gerstoft
William M. Hartmann
Kevin D. Heaney
William S. Hodgkiss
Paul D. Hursky
John M. Impagliazzo
Hua Lee
Patrick J. Loughlin
Jens M. Meyer
Hassan Namarvar
Joseph W. Posey
James C. Preisig
Brian D. Rapids
Edmund J. Sullivan

Ex officio:
Leon H. Sibul, member of Medals and Awards Committee
David I. Havelock, member of Membership Committee
David J. Evans, member of ASACOS

Speech Communication
Maureen L. Stone, *Chair* to 2007

Term to 2008
Norma S. Barroso
Fredericka Bell-Berti
Ann R. Bradlow
Dani M. Byrd
Roger W. Chan
Alexander Francis
Kenneth W. Grant
Anders Lofqvist
Terrance M. Nearey
Douglas D. O'Shaughnessy
Joseph S. Perkell
Astrid Schmidt-Nielsen
Joan E. Sussman
Gary G. Weismer

Term to 2007
Patrice S. Beddor
Melissa A. Epstein
Robert Allen Fox

Helen M. Hanson
Diane Kewley-Port
Jody E. Kreiman
Andrew J. Lotto
Shrikanth S. Narayanan
Dwayne Paschall

Term to 2006
Jean E. Andruski
Lynne E. Bernstein
Ocke-Schwen Bohn
Suzanne E. Boyce
Bruce R. Gerratt
Benjamin R. Munson
Peggy B. Nelson
Brad H. Story

Ex officio:
Fredericka Bell-Berti, member of Medals and Awards
Committee
Abeer Alwan, member of Membership Committee
Shrikanth S. Narayanan, member of ASACOS

Structural Acoustics and Vibration
Courtney B. Burroughs, *Chair* to 2006

Term to 2008
Dean E. Capone
Joel Garrelick
Peter C. Herdic
Teik C. Lim
Thomas J. Royston
Angie Sarkissian
Jeffrey S. Vipperman

Term to 2007
Jeffrey E. Boisvert
Stephen C. Conlon
Linda P. Franzoni
Robert C. Haberman
Rudolph Martinez
Koorosh Naghshineh
Michael F. Shaw

Term to 2006
Benjamin A. Bard
Alain C. Berry
Kenneth A. Cunefare
Joseph W. Dickey
David Feit
Allison B. Flatau
Guillermo C. Gaunaurd
Karl Grosh
Sabih I. Hayek
Louis A. Herstein
Francis Kirschner
Jean R. Nicolas
Andrew N. Norris
Sally J. Pardue
Allan D. Pierce
Mauro Pierucci
Scott D. Sommerfeldt
Earl G. Williams
Sean F. Wu

Ex officio:
Courtney B. Burroughs, member of Membership Committee

Mauro Pierucci, member of Medals and Awards Committee
Sabih I. Hayek, member of ASACOS

Underwater Acoustics
Peter H. Dahl, *Chair* to 2006

Term to 2008
Juan I. Arvelo, Jr.
Pierre-Philippe Beaujean
Shira L. Broschat
Geoffrey F. Edelmann
Peter Gerstoft
John H. Glattetre
Brian T. Hefner
Jean-Pierre Hermand
Charles W. Holland
John C. Osler
Kevin B. Smith
Brian J. Sperry
Christopher T. Tindle
Alexandra I. Tolstoy

Term to 2007
David C. Calvo
Jee Woong Choi
Christian P. de Moustier
Stan E. Dosso
Nicholas C. Makris
Zoi-Heleni Michalopoulou
Tracianne B. Neilsen
Robert I. Odom
Marshall H. Orr
Gregory J. Orris
James Preisig
T. Martin Siderius
Kathleen E. Wage
Jixun Zhou

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

Ex officio:
Henrik Schmidt, member of Membership Committee

Eric I. Thorsos, member of Medals and Awards Committee
Joseph M. Zalezak, member of ASACOS

Administrative Committees 2005–2006

Archives and History
Julian D. Maynard, *Chair* to 2007

Term to 2008
Anthony A. Atchley
Leo L. Beranek
William J. Cavanaugh
Steven L. Garrett
Logan E. Hargrove
Allan D. Pierce
Victor W. Sparrow

Term to 2007
Henry E. Bass
David T. Blackstock
E. Carr Everbach
William W. Lang
David L. Moyer
Richard Stern
Rosalie M. Uchanski

Term to 2006
Jont B. Allen
Ralph R. Goodman
David I. Havelock
Wesley L. Nyborg
Richard J. Peppin
William J. Strong

Audit
George V. Frisk, *Chair* to 2006

Term to 2008
Diane Kewley-Port

Term to 2007
Judy R. Dubno

Books +
David L. Bradley, *Chair* to 2008

Term to 2008
James P. Cottingham
Nancy S. McGarr
Jeffrey A. Nystuen
Neal A. Shaw
Emily A. Tobey

Term to 2007
Stanley L. Chin-Bing
Robert C. Spindel

Term to 2006
Jerry H. Ginsberg
Philip L. Marston
Joseph Pope
Robert A. Walkling

Allan D. Pierce, Editor-in-Chief, *ex officio*

College of Fellows Steering
Janet M. Weisenberger, *Chair* to 2007

Term to 2008
Stanley L. Ehrlich
E. Carr Everbach

Term to 2007
Peter G. Cable
M. David Egan
Uwe J. Hansen
Diane Kewley-Port
Thomas D. Rossing

Term to 2006
Joseph W. Dickey
Judy R. Dubno
Jiri Tichy

William J. Cavanaugh, *ex officio* as past Chair
Richard H. Lyon, *ex officio* as past Chair

Education in Acoustics
Uwe J. Hansen, *Chair* to 2006

Term to 2008
David T. Blackstock
Courtney B. Burroughs
Robin O. Cleveland
Kenneth A. Cunefare
D. Michael Daly
Mary Florentine
Logan E. Hargrove
Mardi C. Hastings
Peter L. Hoekje
Darrell R. Jackson
Michel T. T. Jackson
Murray S. Korman
Luc Mongeau
James M. Sabatier
Neil A. Shaw
Kevin B. Smith
Ralph A. Stephen
James E. West
Wayne M. Wright

Term to 2007
George A. Bissinger
David A. Brown
Robert D. Collier
Corinne M. Darvennes
Margaritis S. Fourakis
Carole E. Gelfer
Daniel O. Ludwigsen
Sharon Y. Manuel
Philip L. Marston
Ralph T. Muehleisen
Andrew A. Piacsek
Daniel R. Raichel
Thomas D. Rossing
Ronald A. Roy
Dawn R. Schuette
Scott D. Sommerfeldt
William Thompson, Jr.
Robert A. Walkling
George S. K. Wong

Term to 2006
William A. Ahroon
Takayuki Arai
Anthony A. Atchley
Fredericka Bell-Berti
Suzanne E. Boyce
Robert D. Celmer

Annabel J. Cohen
E. Carr Everbach
Thomas B. Gabrielson
Katherine S. Harris
Elizabeth S. Ivey
Joie P. Jones
Maria B. Mody
Amy T. Neel
P. K. Raju
Deborah M. Rekart
Daniel A. Russell
M. Roman Serbyn
Victor W. Sparrow
Emily A. Tobey

Ethics and Grievances
Janet M. Weisenberger, *Chair* to 2008

Term to 2008
William J. Cavanaugh
Gerald L. D'Spain
Barbara G. Shinn-Cunningham

Investments
Ilene Busch-Vishniac, *Chair* to 2007

Term to 2008
Lawrence A. Crum
Richard H. Lyon

Term to 2006
Patricia K. Kuhl

David Feit, Treasurer, *ex officio*

Medals and Awards
David L. Bradley, *Chair* to 2006
Term to 2008
Charles R. Greene	Animal Bioacoustics
Mauro Pierucci	Structural Acoustics and Vibration
Eric I. Thorsos	Underwater Acoustics
Ewart A. Wetherill	Architectural Acoustics

Term to 2007
Uwe J. Hansen	Musical Acoustics
Fredericka Bell-Berti	Speech Communication
Mahlon D. Burkhard	Engineering Acoustics
James F. Lynch	Acoustical Oceanography
Leon H. Sibul	Signal Processing in Acoustics

Term to 2006
Joseph W. Hall	Psychological and Physiological Acoustics
Lawrence A. Crum	Biomedical Ultrasound/Bioresponse to Vibration
Robert M. Keolian	Physical Acoustics
Nancy S. Timmerman	Noise

Meetings: May 2005–October 2005
Clark S. Penrod, *Chair* to 2008
Whitlow W. L. Au, Fall 2006, Honolulu and ASA Vice President-Elect
Damian Doria, Spring 2004, New York
Donna L. Neff, Vice President
Murray S. Hodgson, Spring 2005, Vancouver
Russell Johnson, Spring 2004, New York
William A. Kuperman, Fall 2004, San Diego
James H. Miller, Spring 2006, Providence
Elaine Moran, ASA Office Manager, *ex officio*

Peggy B. Nelson, Fall 2005, Minneapolis
Charles E. Schmid, Executive Director, *ex officio*

Meetings: October 2005–June 2006
Clark S. Penrod, *Chair* to 2008
Whitlow W. L. Au, Fall 2006, Honolulu and ASA Vice President-Elect
Donna L. Neff, Vice President
Murray D. Hodgson, Spring 2005, Vancouver
William A. Kuperman, Fall 2004, San Diego
James H. Miller, Spring 2006, Providence
Elaine Moran, ASA Office Manager, *ex officio*
Peggy B. Nelson, Fall 2005, Minneapolis
Charles E. Schmid, Executive Director, *ex officio*
Scott D. Sommerfeldt, Spring 2007, Salt Lake City

Membership
Peter H. Rogers, *Chair* to 2006

Term to 2008
E. Carr Everbach	Biomedical Ultrasound/Bioresponse to Vibration
Courtney B. Burroughs	Structural Acoustics and Vibration
Burton G. Hurdle	International Members
John Erdreich	Noise

Term to 2007
Mohsen Badiey	Acoustical Oceanography
Steven L. Garrett	Physical Acoustics
David I. Havelock	Signal Processing in Acoustics
Henrik Schmidt	Underwater Acoustics
Gregory C. Tocci	Architectural Acoustics

Term to 2006
Abeer Alwan	Speech Communication
Leslie R. Bernstein	Psychological and Physiological Acoustics
James P. Cottingham	Musical Acoustics
Thomas R. Howarth	Engineering Acoustics
Darleen R. Ketten	Animal Bioacoustics

Public Relations
Barbara J. Sotirin, *Chair* to 2006

Term to 2007
Kelly J. Benoit-Bird
E. Carr Everbach
Christy K. Holland
Joe W. Posey
Stephen C. Thompson

Term to 2006
Paul A. Baxley
Paul Hursky
Elizabeth A. McLaughlin
Jack E. Randorff
Kevin B. Smith
James C. Yu

Term to 2005
Geoffrey F. Edelmann
Katherine Kim
Ellen S. Livingston
Andrew Piacsek
Brigitte Schulte-Fortkamp

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
Floyd Dunn, *Chair* to 2006

Term to 2008
Diane Dalecki
James F. Lynch

Term to 2007
Charles C. Church
Mark F. Hamilton
Mardi C. Hastings

Term to 2006
Jont B. Allen
David I. Havelock

Anthony A. Atchley, President-Elect, *ex officio*
Allan D. Pierce, Editor-in-Chief, *ex officio*

Regional Chapters
Juan I. Arvelo, *Cochair* to 2008
Elizabeth A. McLaughlin, *Cochair* to 2008

Angelo J. Campanella	Central Ohio
Robert M. Keolian	Central Pennsylvania
Ernest M. Weiler	Cincinnati
Rebecca Mercuri	Delaware Valley
Timothy J. Foulkes	Greater Boston
Michael J. Anderson	Inland Northwest
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Sergio Beristain	Mexico City
Roger T. Richards	Narragansett
George Bissinger	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Juan Arvelo	Washington, DC
Thomas M. Disch	Wisconsin

Uwe J. Hansen, Chair, Education in Acoustics, *ex officio*
David Feit, Treasurer, *ex officio*
Conner Duke, Student Council representative, *ex officio*

Rules and Governance
William M. Hartmann, *Chair* to 2008

Term to 2008
Ilene J. Busch-Vishniac
Tony F. W. Embleton
Richard H. Lyon

Term to 2007
William J. Cavanaugh
Floyd Dunn

Term to 2006
Elaine Moran
Charles E. Schmid

Prizes and Special Fellowships
Wayne M. Wright, *Chair* to 2007

Term to 2008
Uwe J. Hansen

Term to 2007
Anthony A. Atchley
Constantine Trahiotis

Term to 2006
Fredericka Bell-Berti
James E. West

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
Richard 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
Donna L. Neff, Chair of ASA Technical Council, *ex officio*
Anthony P. Lyons, 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
Sabih I. Hayek, 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

Student Council

Brian Monson	Chair and Signal Processing in Acoustics
Tessa C. Bent	Speech Communication
Joshua G. Bernstein	Psychological and Physiological Acoustics
Erica Bowden	Regional Chapter Committee Liaison
David T. Bradley	Architectural Acoustics
Conor Duke	Noise
Andrew Ganse	Underwater Acoustics
Marla Holt	Animal Bioacoustics
Wheeler B. Howard	Physical Acoustics
Nicolas Le Dantec	Acoustical Oceanography
Alexandra Loubeau	Engineering Acoustics
Andrew C. H. Morrison	Musical Acoustics
Neil Owen	Biomedical Ultrasound/Bioresponse to Vibration
Micha Shepard	Structural Acoustics and Vibration

Tutorials

to be appointed, *Chair* to 2008

Term to 2008
Kenneth A. Cunefare
David R. Dowling
Barbara G. Shinn-Cunningham

Term to 2007
Ann R. Bradlow
James V. Candy
James P. Chambers

Term to 2006
Gerald D. Kidd
Lily M. Wang
Charles E. Schmid, Executive Director, *ex officio*

Women in Acoustics

Peggy B. Nelson, *Chair* to 2006

Term to 2008
Mardi C. Hastings
Brigitte Schulte-Fortkamp
Sophie Van Parijs
Lily M. Wang

Term to 2007
Sarah Hargus Ferguson
Benjamin R. Munson
Donna L. Neff
Nancy S. Timmerman

Term to 2006
David T. Bradley
Chi-Feng Chen
Melissa A. Epstein
Subha Maruvada
Katherine Sawicki
Whitlow W. L. Au, *ex officio* as Vice President-Elect

Ad Hoc Committees 2005–2006

ARLO Editorial Advisory

D. Keith Wilson, *Chair*

E. Carr Everbach
Anders Lofqvist

Philip L. Marston
Andrew N. Norris
Allan D. Pierce
Michael R. Stinson

International Research and Education

Gilles A. Daigle, *Chair* to 2005

Sergio Beristain
Lawrence A. Crum
Malcolm J. Crocker
Samir N. Y. Gerges
Vera A. Khokhlova
William M. Hartmann
Konstantin A. Naugolnykh
Oleg Sapozhnikov
Brigitte Schulte-Fortkamp
Michael Vorlander
Suk Wang Yoon

Future of ASA Publications

William A. Yost, *Chair*
Anthony A. Atchley
Ilene J. Busch-Vishniac
Lawrence A. Crum
Jody Kreiman
Allan D. Pierce
Edward J. Walsh

JASA Editorial Board

Term to June 2008
S. B. Blaeser, Acoustical News—Standards
W. M. Carey, Signal Processing in Acoustics
K. A. Cunefare, Noise, Its Effects and Control
D. Deutsch, Music and Musical Instruments
D. Feit, Structural Acoustics and Vibration
N. H. Fletcher, Music and Musical Instruments
K. G. Foote, Underwater Sound
K. W. Grant, Speech Perception
G. C. Lauchle, Atmospheric Acoustics and Aeroacoustics
R. A. Lutfi, Psychological Acoustics
P. L. Marston, Acoustical Reviews—Books
W. G. Mayer, Acoustical News—International
L. D. Rice, Acoustical Reviews—Patents
W. P. Shofner, Physiological Acoustics
R. A. Stephen, Underwater Sound
B. H. Story, Speech Production
J. E. Sussman, Speech Perception
A. J. Szeri, Ultrasonics; Physical Effects of Sound
R. M. Waxler, General Linear Acoustics
D. K. Wilson, Noise: Effects and Controls
G. S. K. Wong, Acoustical Standards News
A. J. Zuckerwar, Applied Acoustics; Transduction; Acoustical Measurements

Term to June 2007
W. W. L. Au, Bioacoustics—Animal
P. E. Barbone, Physical Acoustics
Y. H. Berthelot, Ultrasonics; Physical Effects of Sound
A. R. Bradlow, Speech Perception
C. C. Church, Bioacoustics
A. J. M. Davis, General Linear Acoustics
F. Dunn, Bioacoustics—Biomedical
A. L. Francis, Speech Perception
L. P. Franzoni, Structural Acoustics and Vibration
J. H. Ginsberg, Structural Acoustics and Vibration
M. F. Hamilton, Nonlinear Acoustics

ACOUSTICAL STANDARDS NEWS

Susan B. Blaeser, Standards Manager

ASA Standards Secretariat, Acoustical Society of America, 35 Pinelawn Rd., Suite 114E, Melville, NY 11747 [Tel.: (631) 390-0215; fax: (631) 390-0217; e-mail: asastds@aip.org]

George S. K. Wong

Acoustical Standards, Institute for National Measurement Standard, National Research Council, Ottawa, Ontario K1A 0R6, Canada [Tel.: (613) 993-6159; fax: (613) 990-8765; e-mail: george.wong@nrc.ca]

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please, contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar National

At the 150th ASA Meeting joint with Noise-Con, Minneapolis, MN at the Hilton Hotel and Towers, 17–21 October 2005, the ASA Committee on Standards (ASACOS) and ASACOS STEERING Committees will meet as below:

- *Monday, 17 October 2005*
  ASACOS Steering Committee
- *Thursday, 20 October 2005*
  ASA Committee on Standards (ASACOS). Meeting of the Committee that directs the Standards Program of the Acoustical Society.

International Standard Committee Meetings

12–16 December 2005

ISO/TC 108/SC 5 Condition monitoring and diagnostics of machines

This committee and its working groups will meet at the SeaTech Campus of the Florida Atlantic University Graduate School of Ocean Engineering, in Dania Beach, Florida.

American National Standards Call for Comment on Proposals Listed

This section solicits comments on proposed new American National Standards and on proposals to revise, reaffirm, or withdraw approval of existing standards. The dates listed in parentheses are for information only.

AMCA (Air Movement and Control Association)

Revisions

BSR/AMCA 204-200x, Balance Quality and Vibration Levels for Fans (revision of ANSI/AMCA 204-1996)

This standard defines appropriate fan balance quality and operating vibration levels to individuals who specify, manufacture, use, and maintain fans. (20 June 2005)

Final actions on American National Standards

The standards actions listed below have been approved by the ANSI Board of Standards Review (BSR) or by an ANSI-Audited Designator, as applicable.

ASA (ASC S1) (Acoustical Society of America)

Revisions

ANSI S1.13-2005, Measurement of Sound Pressure Levels in Air [revision of ANSI S1.13-1995 (R1999)]

Specifies requirements and describes procedures for the measurement of sound pressure levels in air at a single point in space. These apply primarily to measurements performed indoors but may be utilized in outdoor measurements under specified conditions. This is a fundamental standard applicable to a wide range of measurements and to sounds that may differ widely in temporal and spectral characteristics; more specific ANS complement its requirements. A classification is given of the types of sound generally encountered, and the preferred descriptor for each type is identified. (25 July 2005)

ASA (ASC S1) (Acoustical Society of America)

Reaffirmations

ANSI S1.16-2000 (R2005), Method for Measuring the Performance of Noise Discriminating and Noise Canceling Microphones (reaffirmation of ANSI S1.16-2000): (27 April 2005)

ASA (ASC S1) (Acoustical Society of America)

Withdrawals

ANSI S2.38-1982, Field Balancing Equipment—Description and Evaluation [withdrawal of ANSI S2.38-1982 (R2001)]: (27 April 2005)

ANSI S2.40-1984, Mechanical Vibration of Rotating and Reciprocating Machinery—Requirements for Instruments for Measuring Vibration Severity [withdrawal of ANSI S2.40-1984 (R2001)]: (27 April 2005)

ANSI S2.45-1983 (R2001), Electrodynamics Test Equipment for Generating Vibration—Methods of Describing Equipment Characteristics [withdrawal of ANSI S2.45-1983 (R2001)]: (30 June 2005)

ANSI S2.58-1983 (R2001), Auxiliary Tables for Vibration Generators—Methods of Describing Equipment Characteristics [withdrawal of ANSI S2.58-1983 (R2001)]: (30 June 2005)

ASA (ASC S3) (Acoustical Society of America)

Revisions

ANSI S3.4-2005, Procedure for the Computation of Loudness of Steady Sounds [revision of ANSI S3.4-1980 (R2003)]: (1 June 2005)

BSR/ASHRAE 87.3P-2001 (R200x), Method of Testing Propeller Fan Vibration (reaffirmation of ANSI/ASHRAE 87.3P-2001)

This standard establishes laboratory and on-site diagnostic test methods for identifying causes of vibration problems involving direct-driven propeller fans for condenser cooling in air-conditioning units, heat pumps, and others.

ASTM (ASTM International)

Reaffirmations

ANSI/ASTM E1662-1995A (R2005), Classification for Serviceability of an Office Facility for Sound and Visual Environment [reaffirmation of ANSI/ASTM E1662-1995A (R99)]: (1 May 2005)

IEEE (ASC C63) (Institute of Electrical and Electronics Engineers)

Revisions

BSR C63.19-200x, Methods of Measurement of Compatibility between Wireless Communications Devices and Hearing Aids (revision of ANSI C63.19-2001)

Sets forth uniform methods of measurement and parametric requirements for the electromagnetic and operational compatibility and accessibility of hearing aids used with wireless communications devices operating in the range of 800 MHz to 3 GHz. However, this version is focused on existing services, which are in common use. (20 June 2005)

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers of the initiation and the scope of activities expected to result in new or revised American National Standards. This information is a key element in planning and coordinating American National Standards.

The following is a list of proposed new American National Standards or revisions to existing American National Standards that have been received from ANSI-Accredited standards developers that utilize the periodic maintenance option in connection with their standards. Directly and materially affected interests wishing to receive more information should contact the standards developer directly.

ASAE (American Society of Agricultural Engineers)

BSR/ASAE/ISO 5007-200x, Agricultural wheeled tractors—Operator's seat—Laboratory measurement of transmitted vibration (identical national adoption)

Specifies a laboratory method for measuring and evaluating the effectiveness of the suspension of operator seats on agricultural wheeled tractors.

BSR/AISI COFS/HEADER-200x, Cold-Formed Steel Framing—Header BSR/ASAE/ISO 5008—200x, Agricultural wheeled tractors and field machinery—Measurement of whole body vibration of the operator (identical national adoption)

Specifies methods for measuring and reporting the whole body vibration to which the operator of an agricultural wheeled tractor or other field machine is exposed when operating on a standard test track.

TIA (Telecommunications Industry Association)

BSR/TIA 1083-200x, Measurement procedures and performance requirements for handset generated in-band magnetic noise (new standard)

Documents the measurement techniques required for measuring in-band (audio) magnetic noise radiated from a telephone handset. The handset may be corded or cordless. Establish performance requirements for magnetic noise signals radiated from a telephone handset.

Project Need: Provide a method and criteria for evaluating the impact to a hearing aid of unwanted magnetic signals (noise) radiating from a telephone handset.

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization

ISO Standards

ACOUSTIC (TC 43)

ISO 140-11:2005, Acoustics—Measurement of sound insulation in buildings and of building elements—Part 11: Laboratory measurements of the reduction of transmitted impact sound by floor coverings on lightweight reference floors

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 7919-5:2005, Mechanical vibration—Evaluation of machine vibration by measurements on rotating shafts—Part 5: Machine sets in hydraulic power generating and pumping plants

ISO 18437-2:2005, Mechanical vibration and shock—Characterization of the dynamic mechanical properties of visco-elastic materials—Part 2: Resonance method

ISO 18437-3:2005, Mechanical vibration and shock—Characterization of the dynamic mechanical properties of visco-elastic materials—Part 3: Cantilever shear beam method

ISO 16063-22:2005, Methods for the calibration of vibration and shock transducers—Part 22: Shock calibration by comparison to a reference transducer

BUILDING CONSTRUCTION (TC 59)

ISO 12680-1:2005, Methods of test for refractory products—Part 1: Determination of dynamic Youngs modulus (MOE) by impulse excitation of vibration

IEC Standards

PIEZOELECTRIC AND DIELECTRIC DEVICES FOR FREQUENCY CONTROL AND SELECTION (TC 49)

IEC 61019-2 Ed. 2.0 en: 2005, Surface acoustic wave (SAW) resonators—Part 2: Guide to the use

ENVIRONMENTAL CONDITIONS, CLASSIFICATION AND METHODS OF TEST (TC 104)

IEC 60068-2-80 Ed. 1.0 b: 2005, Environmental testing—Part 2-80: Tests—Test Fi: Vibration—Mixed mode

IEC Technical Specifications

PIEZOELECTRIC AND DIELECTRIC DEVICES FOR FREQUENCY CONTROL AND SELECTION (TC 49)

IEC/TS 61994-4-4 Ed. 1.0 en: 2005, Piezoelectric and dielectric devices for frequency control and selection—Glossary—Part 4-4: Materials—Materials for Surface Acoustic Wave (SAW) devices

IEC 62276 Ed. 1.0 en: 2005, Single crystal wafers for surface acoustic wave (SAW) device applications—Specifications and measuring methods

ISO Draft Standard

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 13374-2, Condition monitoring and diagnostics of machines—Data processing, communication and presentation—Part 2: Data-processing—(21 August 2005)

ISO 10326-1/ DAmd1, Mechanical vibration—Laboratory method for evaluating vehicle seat vibration—Part 1: Basic requirements—Amendment 1—(20 August 2005)

HYDROMETRIC DETERMINATIONS (TC 113)

ISO/DIS 4366, Hydrometry—Echo sounders for water depth measurements—(8 September 2005)

IEC Draft Standard

29/584/FDIS, IEC 60118-8 Ed.2: Electroacoustics—Hearing aids—Part 8: Methods of measurement of performance characteristics of hearing aids under simulated *in situ* working conditions (2 September 2005)

46A/718/FDIS, IEC 61196-1-107: Coaxial communication cables—Part 1-107: Electrical test methods—Test for cable microphony charge level (mechanically induced noise) (1 July 2005)

65B/558/FDIS, IEC 60534-8-1: Industrial-process control valves—Part 8-1: Noise considerations—Laboratory measurement of noise generated

by aerodynamic flow through control valves (12 August 2005)

International documents submitted to the U.S. for vote and/or comment

Some of the documents processed recently by the ASA Standards Secretariat. Dates in parentheses are deadlines for submission of comments and recommendation for vote, and they are for information only.

U.S. TAG	ISO documents
S2	**Systematic Review of ISO 4867:1984** "Code for the measurement and reporting of shipboard data" (2 June 2005)

Systematic Review of ISO 4868:1984 "Code for the measurement and reporting of local vibration data of ship structures and equipment" (2 June 2005)

Systematic Review of ISO 6070:1981 "Auxiliary tables for vibration generators—Method of describing equipment characteristics" (2 June 2005)

Systematic Review of ISO 6954:2000 "Mechanical vibration—Guidelines for the measurement, reporting and evaluation of vibration with regard to habitability on passenger and merchant ships" (2 June 2005)

Systematic Review of ISO 10816-1:1995 "Mechanical vibration—Evaluation of machine vibration by measurements on nonrotating parts—Part 1: General guidelines" (2 June 2005)

ISO/FDIS 16063-22 "Methods for calibration of vibration and shocktransducers—Part 22: Shock calibration by comparison with a reference transducer"

Systematic Review of ISO 10816-5:2000 "Mechanical vibration—Evaluation of machine vibration by measurements on nonrotating parts—Part 5: Machine sets in hydraulic power generating and pumping plants" (2 June 2005)

Systematic Review of ISO 10816-6:1995 "Mechanical vibration—Evaluation of machine vibration by measurements on nonrotating parts—Part 6: Reciprocating machines with power ratings above 100 kW" (2 June 2005)

Systematic Review of ISO 14964:2000 "Mechanical vibration and shock vibration of stationary structures-Specific requirements for quality management in measurement and evaluation of vibration" (2 June 2005)

S3   **ISO/DIS 389-5** "Acoustics—Reference zero for the calibration of audiometric equipment—Part 5: Reference equivalent threshold sound pressure levels for pure tones in the frequency range 8 kHz to 16 kHz"

S12   **Fourth ISO/CD 17201-3** "Acoustics—Noise from shooting ranges—Part 3: Guidelines for sound propagation calculations" (2 May 2005)

ISO/DIS 362-1 "Acoustics—Engineering method for the measurement of noise emitted by accelerating road vehicles—Part 1: Vehicles of categories M and N" (2 May 2005)

ISO/DIS 362-2 "Acoustics—Engineering method for the measurement of noise emitted by accelerating road vehicles—Part 2: Vehicles of category L" (2 May 2005)

ISO/DIS 1996-2.2 "Acoustics—Description, assessment and measurement of environmental noise—Part 2: Determiniation of environmental noise levels" (15 February 2005)

Systematic Review of ISO 3743-2: 1994 "Acoustics—Determination of sound power levels of noise sources using sound pressure—Engineering methods for small, movable sources in reverberent fields—Part 2: Methods for special reverberation test rooms" (25 March 2005)

Systematic Review of ISO 3740: 2000 "Acoustics determination of sound power levels of noise sources—Guidelines for the use of basic standards" (25 March 2005)

Systematic Review of ISO 4869-2: 1994 "Acoustics—Hearing protectors —Part 2: Estimation of effective A-weighted sound pressure levels when hearing protectors are worn" (25 March 2005)

Systematic Review of ISO 7196: 1995 "Acoustics—Frequency-weighting characteristic for infrasound measurements" (25 March 2005)

OBITUARY

Stanley J. Bolanowski ● 1950–2005

Stanley (Sandy) J. Bolanowski, a Fellow of this Society, died suddenly and unexpectedly on January 13, 2005 while on sabbatical leave at Vanderbilt University in Nashville, Tennessee.

Dr. Bolanowski was a professor in the Department of Biomedical and Chemical Engineering in the L.C. Smith College of Engineering and Computer Science and was associate director of the Institute for Sensory Research (ISR) at Syracuse University.

Dr. Bolanowski was born on February 22, 1950 in Utica, New York, where he attended both elementary and high school. He went on to study electrical technology at the State University of New York at Morrisville, where he earned an associate's degree in electrical technology in 1970. At Syracuse University he earned a bachelor's degree in psychology (1973) and was one of the first students to receive a bachelor's degree in electrical engineering—biosystems (1974). It was during his undergraduate studies at Syracuse that he met and worked with Dr. Ronald Verrillo, founder of the somatosensory laboratories at ISR. He received a Ph.D. in sensory sciences at Syracuse University in 1981. He built the vibrotactile physiology laboratory at ISR and utilized it for his Ph.D. thesis titled "Intensity and Frequency Characteristics of Pacinian Corpuscles." His graduate advisers were Dr. Ronald Verrillo and Dr. Jozef Zwislocki.

After receiving his doctorate, he was awarded a postdoctoral fellowship (1980–1982) and then was appointed assistant professor at the Center for Brain Research (1983–1987) at the University of Rochester School of Medicine and Dentistry, and assistant professor at the Center for Visual Science at the University of Rochester (1983–1987). While at Rochester he studied the cortical activity within the visual systems of monkeys and the psychophysical responses of humans to *ganzfeld* stimulation. In this work he collaborated with the noted neurophysiologist Dr. Robert Doty.

In 1988, Dr. Bolanowski returned to Syracuse University and continued his collaboration with Dr. Verrillo, and with Dr. George Gescheider of the nearby Hamilton College Department of Psychology. His 1988 JASA article with coauthors G.A. Gescheider, R.T. Verrillo, and C.M. Checkosky has been described as the "now-classic paper that provided evidence of four different mechanoreceptors…" with "basic findings" that to this day "remain the best theory of mechanoreceptive function" [J.E. Weisenberger, "History of Bioresponse to Vibration in the Acoustical Society of America" in *ASA at 75*, (2004) p. 81]. He was appointed associate professor of bioengineering and neuroscience in 1988 and was promoted to full professor in 1994, a position that he held until the time of his death. From 1995 to 1999 he was director of the Computational Neuroscience Program and in 2002 he was appointed associate director of the Institute for Sensory Research under its current director Robert L. Smith. After his return to Syracuse he continued his association with the University of Rochester where he held an adjunct professorship in the Department of Surgery (Neurosurgery/Neuroendocrine Unit) from 1988–1992. In 1992–1993 and again in 1994–1995 he was a visiting professor in the Department of Otolaryngology at the Keck Center for Integrative Neurobiology and the Coleman Laboratory, University of California at San Francisco. In this position he did collaborative research with Dr. Michael Merzenich on the structure and functioning of the pain, temperature, and touch sensory systems. His collaboration activity with scientists outside of Syracuse was extensive, including M. Holmes, Department of Mathematical Sciences at Rensselaer Polytechnic Institute and J. Bell, Department of Mathematics, SUNY Buffalo, with whom he modeled the mechanics and transduction mechanisms of the Pacinian corpuscle. With D. Dishman of the New York Chiropractic College, Seneca Falls, NY he studied the double-crush syndrome and other spine/nervous system disorders. In this collaboration he also developed a device for controlled spinal manipulations. With V. Apkarian at the SUNY Health Sciences Center at Syracuse he did research on functional MRI and stochastic resonance. His collaboration with U.S. Navy scientists included the design and execution of psychophysical and clinical tests for deep-sea divers at the Naval Submarine Medical Research Laboratory in Groton, Connecticut. With the Naval Air Medical Research Laboratory, Pensacola, Florida, he participated in the design and execution of experiments to evaluate devices to aid in the situational awareness of pilots.

The recipient of numerous research grants from government, military, and industrial institutions, Dr. Bolanowski's expertise spanned many areas of sensory science, including neurophysiology, anatomy, psychophysics, and the modeling of sensory systems. His contributions to sensory science include: The precise determination of the intensity and frequency characteristics of the Pacinian corpuscle at the level of the receptor and spike potentials; the effects of response criteria and skin temperature on the responses to vibration in humans and in excised Pacinian corpuscles; the effects of monocular and binocular contourless stimuli on the visual responses of humans and on the cortical activity of monkeys; the modeling of physiological and psychophysical responses of the tactile sensory channels in human glabrous and hairy skin; the determination of the columnar compression patterns in the lamellae of Pacinian corpuscles by direct video observation; and the distribution patterns of rapidly and slowly adapting mechanoreceptors in cutaneous tissues. His engineering skills manifested themselves in the numerous laboratory instruments that he developed and perfected for use in the somesthetic laboratories at the Institute for Sensory Research. Some of these systems were built and shipped to other laboratories in other parts of the world.

Dr. Bolanowski became a member of ASA in 1984 and was elected a Fellow of the Society in 1993. He became a member of the Technical Committee on Biomedical Ultrasound/Bioresponse to Vibration (1987) and served on the Standards Committee, American Physics Institute since 1992. He was a member of the Planning Committee for the 117th meeting of the Acoustical Society in Syracuse, NY in May 1989. At this meeting he served as registrar and treasurer and organized a special session on "Structural and Functional Aspects of Mechanotransduction" for that meeting. From 1977 to the present he participated in no fewer than 29 oral presentations at meetings of the Society. A frequent contributor to somatosensory and neuroscience journals, he addressed meetings of the Society for Neuroscience more than 50 times in his career.

A manuscript in preparation at the time of his death describes new and revolutionary views of vibrotactile transduction, and NIH funded him in late 2004 to continue these studies, stating in the summary statement that he was "uniquely skilled in anatomical, psychophysical, and cellular approaches to mechanotransduction and well recognized for his contributions in all these areas." The team of researchers that he gathered together and mentored plan to continue his work in the laboratory he built, and add to our knowledge of how we experience the world through our somatosensory system, an area in which his work will be long remembered.

ROBERT L. SMITH
RONALD T. VERRILLO

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*

SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*

JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*

MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*

DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*

DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*

CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*

WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*

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

ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,873,570

43.30.Pc HIGH RESOLUTION BATHYMETRIC SONAR SYSTEM AND MEASURING METHOD FOR MEASURING THE PHYSIOGNOMY OF THE SEABED

Weiqing Zhu *et al.*, assignors to Institute of Acoustics Chinese Academy of Sciences
29 March 2005 (Class 367/88); filed in China 13 September 2001

A high-resolution sidescan sonar system for ocean bottom profiling is described. It involves an underwater vehicle, either an autonomous vehicle, a remotely operated vehicle, or a towed body, with two sonar arrays mounted symmetrically on the two sides of the lower part of the vehicle. Each of these two sonar arrays has a transmitting line array and three or more parallel, longitudinal, receiving arrays, equally spaced between one-half and a full wavelength at the operational frequency. The system is said to be a high-resolution system having increased precision of measurement, range, and operational efficiency.—WT

6,873,572

43.30.Yj LOW-FREQUENCY SONAR COUNTERMEASURE

Thomas A. Frank, assignor to The United States of America as represented by the Secretary of the Navy
29 March 2005 (Class 367/174); filed 3 May 2004

The design of a low-frequency, expendable, cylindrical-shaped, ship-deployable, sonar projector for countermeasure usage is described. The figure illustrates one subsection of the device which might consist of a number of these subsections connected in coaxial fashion. The actual transducer unit **11** can be either magnetostrictive or piezoceramic and is excited into longitudinal vibrations. Its annular ends **12** are sealed by thin flexible membranes or discs **16**. The transducer is sealed at its outer cylindrical face by a protective shell **13** while a portion **17** of the inner volume is filled with an inert liquid **18** such as a high-purity isoparaffinic solvent. Module **20**, containing battery plus requisite electronics to drive the transducer, is resiliently
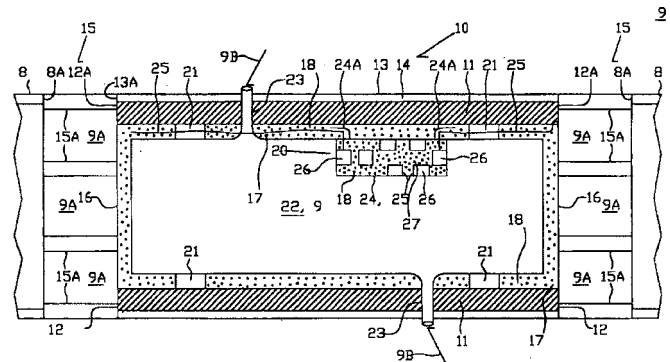
mounted from transducer **11** via supports **21**. Since the battery is likely to be a seawater battery, the remainder of the interior is free-flooded via ducts **23**. Adjacent subsections are connected to one another by an open truss **15** consisting of a number of longitudinally oriented rigid members **15A** that create another free-flooded cavity whose portions are identified as **9A**. As the cylindrical-shaped transducers undergo longitudinal vibrations, the ensemble of these free-flooded cavities **9A** radiate to the surrounding medium **9**. It is alleged that the device will operate at a much lower frequency than a conventional cylindrical transducer of the same size.—WT

6,874,212

43.35.Ns METHOD OF MAKING AN ACOUSTIC WAVE RESONATOR

John Dwight Larson III, assignor to Agilent Technologies, Incorporated
5 April 2005 (Class 29/25.35); filed 8 November 2001

This patent describes a method of temperature compensating piezo-electric resonators using INVAR alloys in a layered-composite design. IN-VAR alloys have a low thermal expansion coefficient, so that the composite structure can be engineered to have zero temperature-dependent frequency shift. The use of INVAR in MEMS structures is unusual, and sputtering is used to deposit the material, which is then chemically etched.—JAH

6,874,211

43.35.Ns METHOD FOR PRODUCING THIN FILM BULK ACOUSTIC RESONATORS (FBARS) WITH DIFFERENT FREQUENCIES ON THE SAME SUBSTRATE BY SUBTRACTING METHOD AND APPARATUS EMBODYING THE METHOD

Paul D. Bradley *et al.*, assignors to Agilent Technologies, Incorporated
5 April 2005 (Class 29/25.35); filed 5 March 2001

This patent discloses a method for stagger tuning mass-produced thin film bulk acoustic resonators (FBARs) on silicon wafers. The method used is to add material to the top or bottom of the resonator area, and then selectively etch those that one wants higher in frequency. The examples given are for cell phones at 1900 MHz frequencies using AlN as the piezo material. There are a few model designs given but very little on the performance (bandwidth, electromagnetic coupling) of the resulting filters. What is described here is not novel and is routinely used by practitioners of the MEMS art.—JAH

6,878,604

43.35.Ns SEMICONDUCTOR ELEMENT COMPRISING A SEQUENCE OF LAYERS FOR CONVERTING ACOUSTIC OR THERMAL SIGNAL AND ELECTRICAL VOLTAGE CHANGES INTO EACH OTHER AND METHOD FOR PRODUCING THE SAME

Robert Aigner *et al.*, assignors to Infineon Technologies AG
12 April 2005 (Class 438/396); filed in Germany 20 July 2000

It is not clear what is being claimed as novel here. The patent is very brief and mentions bulk acoustic wave resonators as the application. The claims for layering a Bragg reflector are not new, and the claims for the use of tungsten, molybdenum, and platinum as electrode materials are not new. What is left is the method of fabrication, which involves nothing novel except the step of chemical/mechanical planarization of the metal electrodes prior to deposition of the piezoelectric material. This is emphasized by the authors as improving the coupling coefficient of the resulting layer. That's all.—JAH

6,876,484

43.35.Ty DEFORMABLE SEGMENTED MEMS MIRROR

Dennis S. Greywall, assignor to Lucent Technologies Incorporated
5 April 2005 (Class 359/291); filed 24 March 2003

This patent describes an interesting approach to deformable mirror design. The mirrors are tiled to cover an array of electrodes and are suspended by ingenious hinges. The hinges have several degrees of freedom to allow the mirrors to move independently. It is proposed that this reduces the complexity of the control logic. It is an interesting patent to read.—JAH

6,879,085

43.35.Ty RESONANCE SHIFTING

Lior Shiv *et al.*, assignors to Nanomotion Limited
12 April 2005 (Class 310/312); filed 12 December 2002

The authors of this patent claim to have discovered the principle of ''resonance shifting'' by means of drilling holes or adding protuberances on a piezoelectric plate. Unfortunately, Chladni discovered this and Rayleigh formulated a mathematical description that quantifies the effects of this procedure well in advance of this patent. What remains are some details on how to suspend a plate so that it doesn't break when put into a cell phone.—JAH

6,866,630

43.35.Wa VIVO BIOCOMPATIBLE ACOUSTIC COUPLING MEDIA

Margaret J. Larson *et al.*, assignors to Sonotech, Incorporated
15 March 2005 (Class 600/437); filed 12 March 2001

The patent describes a medical ultrasound coupling medium and lubricant, in gel or liquid form, consisting of polyethylene oxide (PEO) and aqueous solutions. It is asserted that this coupling medium provides long-term biocompatibility with *vivo* human tissue and body fluids. This medium is formulated and manufactured in such a manner and form that render the acoustic medium sterile, non-cross-linked, pseudoplastic, and containing acceptably low levels of pyrogens.—DRR

6,869,407

43.35.Yb ACOUSTIC WAVE DEVICE

Moshe Ein-Gal, Ramat Hasharon, Israel
22 March 2005 (Class 601/2); filed 12 September 2001

Claim is made for a device for generation and focusing of acoustic waves through the use of electromagnetic energy. The device includes a truncated conical acoustic wave transducer. A parabolic reflector may be arranged with respect to the transducer so as to focus acoustic waves towards a focal point located at the apex of the transducer. An electrical
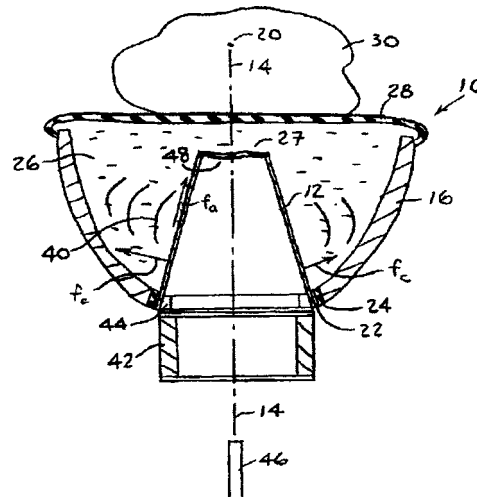
element on the noncylindrical, nonflat, outer contour of a support of the acoustic wave transducer radiates acoustic waves outwardly from that contour's surface. It is envisioned that this apparatus would be used for medical treatment, such as stone fragmentation or orthopedic procedures.—DRR

6,877,975

43.35.Zc ROLLING PIN HORN

David G. Wuchinich, Yonkers, New York
12 April 2005 (Class 425/174.2); filed 16 October 2001

Resonant vibrations of metal horns are used in industry for ultrasonic welding, sealing, laminating, etc. Cylindrical horns capable of rotation permit the continuous processing of materials, but generally tend not to provide uniform radial vibrations. The horns described in the present patent are intended to eliminate this shortcoming. A cylindrical horn described in this patent is tubular, with an interior space that is shaped somewhat like a rolling pin—that is, thin in the middle and becoming thicker towards the ends. A driver that generates axial vibrations is placed in the interior space and connected to the horn via an arrangement of lugs and ridges, so as to induce radial vibrations in the horn.—EEU

6,865,281

43.38.Dv LIQUID COOLED SPEAKER

Jeff B. Jordan, Baton Rouge, Louisiana
8 March 2005 (Class 381/397); filed 28 September 2000

Heat is a byproduct of electrodynamic transduction. This heat needs to be removed to prevent the transducer from burning up. Various schemes such as vented gaps, thermally conductive voice coil formers, forced air circulation, and the like have been patented, and some even implemented, to help remove heat from the motor structure, especially the voice coil. The present invention describes a system that uses a liquid cooled yoke **25** with internal cavity **33** for loudspeaker **1**. The yoke is part of a cooling system
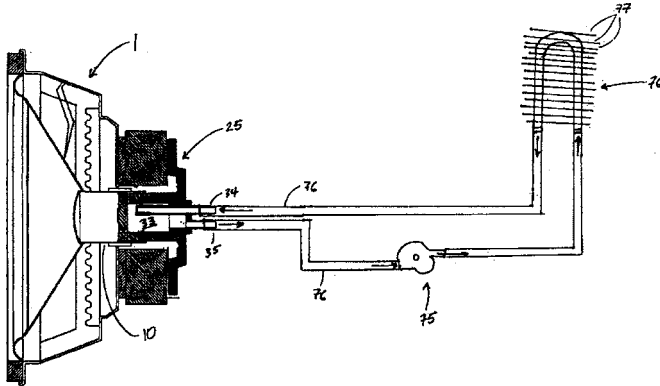
that includes pump **75**, radiator fins **77**, piping **76**, and inlet and outlet connections **34** and **35**. Heat from the voice coil **10** is transferred to yoke **25**, where the fluid in void **33** conducts it away. For special applications this may be a practical method of getting more power handling and less power compression. When used with some present methods that lower the thermal resistance of the air gap, such as ferrofluids, the liquid cooling could be even more effective.—NAS

6,868,166

43.38.Dv CHORD-LIKE VIBRATION LOUDSPEAKER

Xiafu Zhang, Kunming, Yunnan, China *et al.*
15 March 2005 (Class 381/399); filed in Canada 10 June 1998

A ribbon loudspeaker is described. The patent maintains that the magnetic body can be a hollow square column (although in the figure, it appears to be rectangular), a tube, as well as other shapes. This poorly written, short
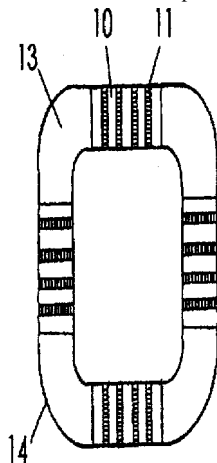
patent (a page and a half long) includes seven figures (all on one page), none of which are adequately described in the patent text. The utility of this invention is unclear as the radiating surface may interact with itself and have limited power-handling capabilities.—NAS

6,877,385

43.38.Gy CONTACT TYPE MICRO PIEZORESISTIVE SHEAR-STRESS SENSOR

Yean-Kuen Fang *et al.*, assignors to National Science Council
12 April 2005 (Class 73/777); filed in Taiwan, Province of China
16 November 2000

This patent describes a shear-stress sensor based on two piezoresistive sensors embedded in a thin membrane of silicon. Shear stresses are coupled to the membrane by a stub, similar in principle of operation to a joystick. The diagrams and wording in the patent are not very clear, so do not seek this one out for fabrication information.—JAH

6,862,360

43.38.Ja SPEAKER SYSTEM

Jen-Hui Tsai, Wen Shen Suburb, Taipei, Taiwan, Province
of China
1 March 2005 (Class 381/351); filed 19 April 2001

A novel understanding and interpretation of the laws of physics, how electrodynamic loudspeakers and their constituent parts operate, the why and how of enclosure and port design, among other things, lead to an inter-

esting patent that is an excellent example of wishful thinking. The patent basically describes a plurality of volumes **14** and ports **20**, each tuned to a different frequency band, that solve a multiplicity of problems, none of which are adequately described.—NAS
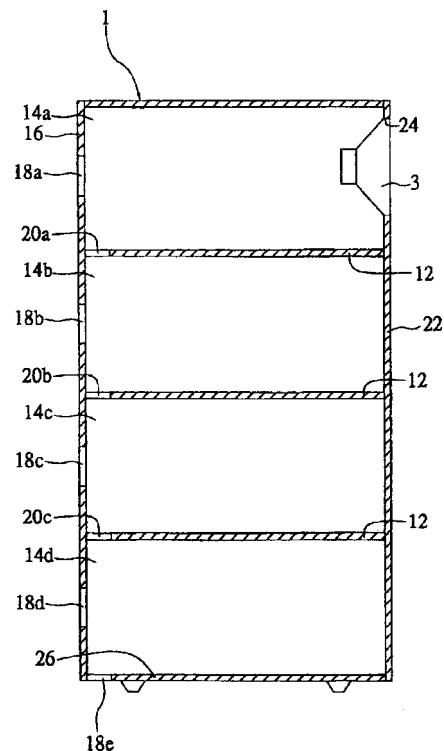
6,845,165

43.38.Ja ACOUSTIC EFFECT APPARATUS AND METHOD AND PROGRAM RECORDED MEDIUM THEREFOR

Haruhiko Motohashi *et al.*, **assignors to Korg Incorporated**
18 January 2005 (Class 381/98); filed in Japan 24 December 1998

To enhance the bass response of a bass drum or string bass, the inventors describe using a nonlinear mechanism to generate overtones and then filtering the result. As shown, a bandpass filter **31** between 200 and 400 Hz
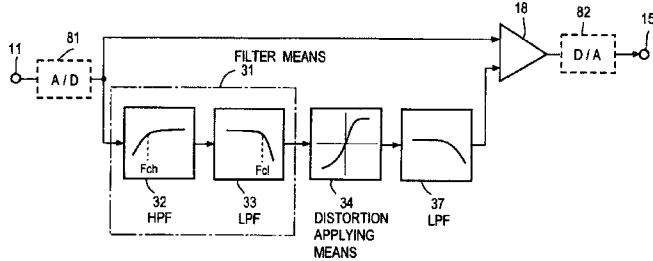
filters the input. This is distorted, filtered, and summed with the input. This is not the first time that nonlinear blocks have been used in this fashion (see United States Patent 6,792,115 [reviewed in J. Acoust. Soc. Am. **117**(3), 990 (2005)]).—MK

6,873,705

43.38.Lc COMBINE AUDIO AND RINGING MODE

Klaus Graef and Edgar Sexauer, assignors to Dialog Semiconductor GmbH
29 March 2005 (Class 379/418); filed 27 November 2001

In this mobile phone audio amplifier, two identical analog circuits are bridged to drive loudspeaker **9**. Each circuit includes a ring-signal generator **7** and **8** and pairs of switches **5**, **15** and **6**, **16** to select audio or ringing operation. Level adjustment in either mode is controlled by ladder attenuators **3**, **13** and **4**, **14**. The use of analog circuitry is said to avoid the high power dissipation and slew rate problems of pulse width modulation. It
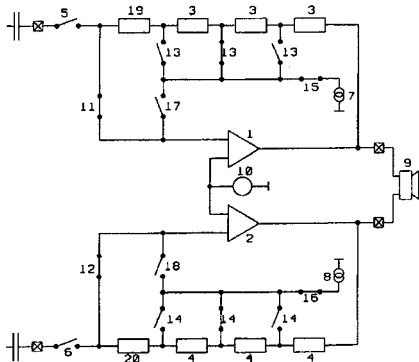
seems apparent that at least ten components could be eliminated by adding a voltage amplifier and then restricting bridged operation to the output stage alone, but with integrated circuit fabrication the trade-offs may favor the approach taken.—GLA

6,826,282

43.38.Md MUSIC SPATIALISATION SYSTEM AND METHOD

Francois Pachet *et al.*, **assignors to Sony France S.A.**
30 November 2004 (Class 381/61); filed in the European Patent Office 27 May 1998

In conjunction with a spatialization unit, a constraint solver **3** can be used to control the position parameters of the sound sources. Typical con-

straints include minimal distances from sound sources to the listener, angular limits with respect to the listener, and maintaining constant distance to sound sources.—MK
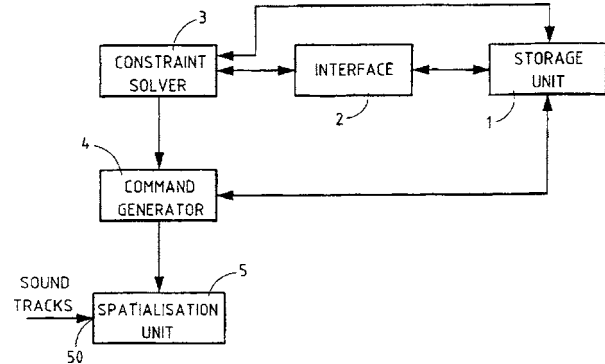
6,868,224

43.38.Md METHOD AND APPARATUS FOR PROVIDING MULTIMEDIA PLAYBACK

Blair B. A. Birmingham and Cheryl B. Giblon, assignors to ATI International SRL
15 March 2005 (Class 386/68); filed 5 May 1999

To assist searching an audio/video recording, the inventors propose (1) using pitch shifting while in fast forward to bring the spectrum down to an understandable range, (2) removing "blank" portions, thereby shortening the time, and (3) reversing the audio samples during rewind.—MK

6,876,847

43.38.Si CONTROL OF SYNCHRONOUS DISPLAY OF MELODY INFORMATION AND DIFFERENT INFORMATION ON MOBILE COMMUNICATION TERMINAL

Kazuhiro Kudoh, assignor to NEC Corporation
5 April 2005 (Class 455/403); filed in Japan 4 October 2000

A typical fold-up cellular phone houses a number of output devices in addition to a loudspeaker—a graphic display, a vibrator, and a signal light, for example. It may also include a programmable tone generator that can play a melody as an incoming call signal. In some instances, according to this patent, melody generation may also be linked to a computer game in which input and output information must be synchronized with the melody in response to certain actions from the user. In the control method described,
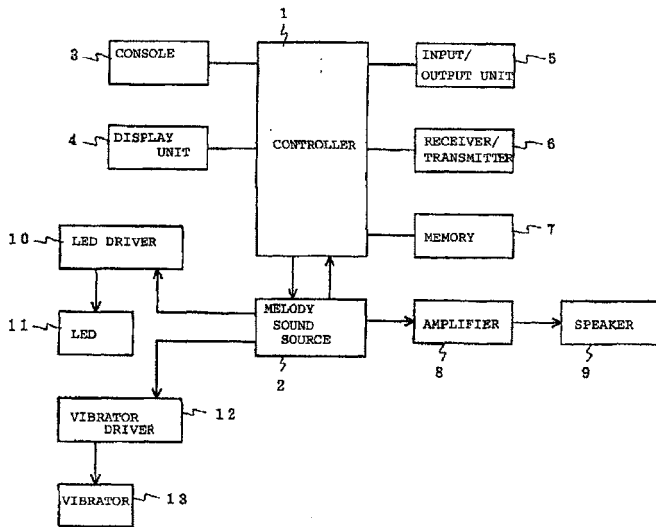
each note of a given melody is coded as a series of attributes, such as frequency, intensity, and duration. This makes it easy to synchronize various outputs with a melody in various modes of operation.—GLA

6,879,698

43.38.Si CELLULAR TELEPHONE, PERSONAL DIGITAL ASSISTANT WITH VOICE COMMUNICATION UNIT

Peter V. Boesen, Des Moines, Iowa
12 April 2005 (Class 381/328); filed 31 May 2001

In his earlier United States Patent 6,094,492, the inventor proposed a method of combining air-conduction and bone-conduction transducers for improved voice pickup. Here, he describes how this method can be advantageously used for mobile, hands-free communications. In the illustration,
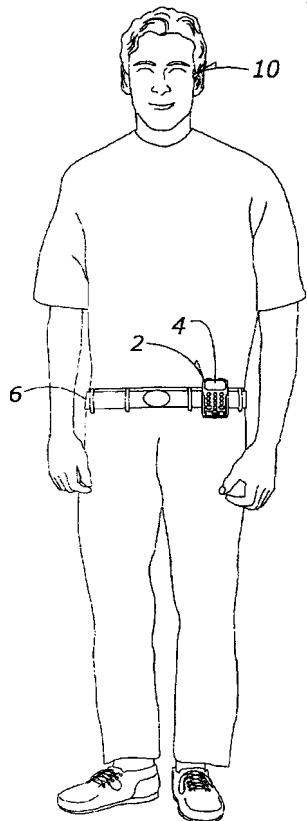
earpiece **10** is connected to cellular transceiver **4** via a short-range wireless link.—GLA

6,879,848

43.38.Si METHOD AND APPARATUS FOR ALERTING AN OPERATOR OF AN INCOMING CELLULAR OR PORTABLE PHONE CALL

Edward A. Lygas, assignor to Sony Corporation; Sony Electronics Incorporated
12 April 2005 (Class 455/567); filed 14 December 2001

During the past few years there has been considerable interest in producing audible signals from modulated ultrasonic beams. In a seemingly unrelated field, developments in medical ultrasonic therapy are regularly patented. The invention disclosed in this Sony patent is a kind of hybrid. Instead of a vibrator, a cellular phone employs two concentric, ring-shaped ultrasonic transducers driven at two different frequencies. In this case, the difference frequency lies below the audible band, perhaps only 10 or 12 Hz. The two ultrasonic signals somehow pass through heavy clothing with negligible attenuation and combine in the user's tissue to generate a tactile sensation. There is nothing in the patent to suggest that a working model has yet been tested.—GLA

6,870,942

43.38.Tj LOUDSPEAKER FOR LINE ARRAY SOUND SYSTEM

Curtis H. Graber, Woodburn, Indiana
22 March 2005 (Class 381/349); filed 3 September 2003

A two-way line array system is described. A ported enclosure **24** that can be trapezoidal in section contains a conventional low-frequency unit **18** and a planar transducer **16**. Many, including the manufacturers of line arraysystems that are similar, if not identical, to the described invention, may
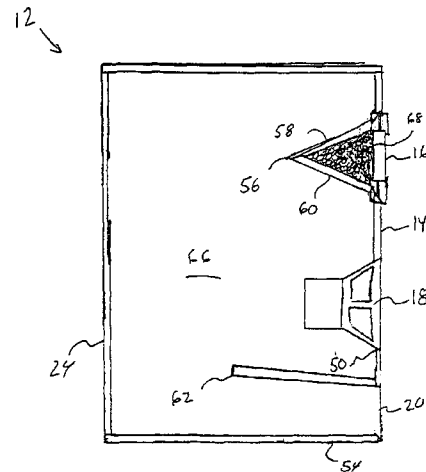
be interested in how this invention may not only fail to be novel, but how it may also fail to be a nonobvious, or even obvious, improvement over the prior art.—NAS

6,865,482

43.38.Vk METHOD AND ARRANGEMENT FOR GUIDING A USER ALONG A TARGET PATH

Richard Hull, assignor to Hewlett−Packard Development Company, L.P.
8 March 2005 (Class 701/211); filed in the United Kingdom 6 August 2002

The inventor uses the "Oboe" audio glidescope system of World War II as an auditory display exemplar. Using a positioning system, stereo cues can be given in the headphones to direct the user inside a space such as a museum.—MK

6,829,582

43.38.WI CONTROLLED ACCESS TO AUDIO SIGNALS BASED ON OBJECTIONABLE AUDIO CONTENT DETECTED VIA SOUND RECOGNITION

Eric Lawrence Barsness, assignor to International Business Machines Corporation
7 December 2004 (Class 704/275); filed 10 October 2000

For those who can't bear modern media, this patent proposes machine-

assisted censorship. Using audio and sound recognition, the offensive passages can be blanked. Hear no evil, see no evil, speak no evil.—MK
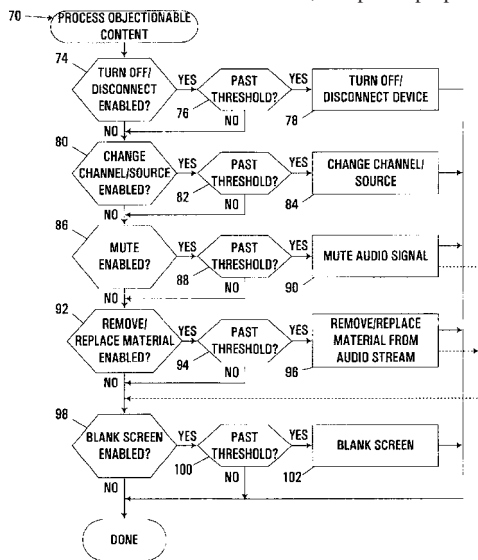
6,878,237

43.40.Le MACHINE AND PROCESS FOR OPERATING A MACHINE TO MONITOR VIBRATIONS

Eva Scheideler, assignor to Voith Paper Patent GmbH
12 April 2005 (Class 162/198); filed in Germany 27 March 2002

In paper-making machines and the like, material that is passed between rollers may exhibit small periodic thickness variations that are undesirable. In order to detect these variations, an acoustic sensor is placed near the material web or near a roller and the resulting signal is subjected to frequency analysis. Observation of excessive amplitudes informs the machine operators to take remedial action.—EEU

6,879,224

43.40.Sk INTEGRATED FILTER AND IMPEDANCE MATCHING NETWORK

Michael L. Frank, assignor to Agilent Technologies, Incorporated
12 April 2005 (Class 333/189); filed 12 September 2002

This patent concerns the creation of high-pass filters having impedance transforming characteristics along with a built-in notch coming from bulk acoustic wave resonators inserted into the filter network. The novelty stems from the tuning of these notches to coincide with certain cell phone frequency bands, and little else. The patent describes the band rejection performance of some prototype filters built this way.—JAH

6,871,565

43.40.Tm VIBRATION CONTROL BY CONFINEMENT OF VIBRATION ENERGY

Daryoush Allaei, assignor to Quality Research, Development & Consulting, Incorporated
29 March 2005 (Class 74/574); filed 17 October 2003

"Confinement devices" are used to keep vibrations that are injected in a structure at given locations out of locations that are to be protected. These devices, which are intended to block the transmission of vibrational energy, may consist of combinations of translational and torsional springs and masses and/or of active vibration control systems. The embodiments illustrated in the patent apply to one-dimensional structures, such as beams and shafts, but the utility of the approach to two-dimensional structures is also described.—EEU

6,877,711

43.40.Tm VIBRATION ISOLATION TABLE

Masakuni Kainuma *et al.*, assignors to Fujikura Rubber Limited
12 April 2005 (Class 248/638); filed in Japan 28 November 2002

This isolation system, intended to protect sensitive instruments from floor vibrations, is configured to provide very low natural frequencies in the horizontal direction. Vertical isolation is provided primarily via an air spring arrangement consisting of a piston **A** in an air chamber **S1**, with a rolling seal **7** enabling relative motion. Horizontal isolation is obtained via an arrangement in which a rod **19** that rests on a curved surface at the base of a
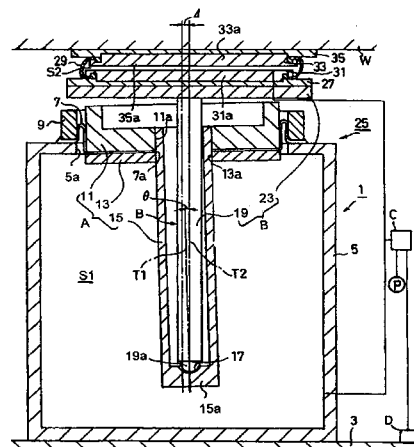
tube **15a** (which in effect is part of the vertical piston) can deflect relatively freely through a small angle θ from the vertical. A second air space **S2** between a plate **31** that is affixed to the top of the rod and a plate **35** that supports the payload, with the air space sealed by a diaphragm **29**, provides additional horizontal and vertical isolation. The two air spaces are pressurized from a source **C** that is controlled on the basis of floor vibrations detected by a sensor **D**.—EEU

1254   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Reviews of Acoustical Patents   1254

6,878,873

43.40.Tm DAMPING SLEEVE WITH RESILIENT SUPPORT MEMBERS

Samuel B. Fryberger, Jr. and Brian M. Snyder, assignors to Federal−Mogul World Wide, Incorporated
12 April 2005 (Class 174/48); filed 8 August 2002

The sleeves described in this patent are intended for reducing the rattling noise associated with wiring harnesses and the like in motor vehicles, and also for reducing the vibrations transmitted to these harnesses. A typical sleeve consists of a layer of a material, such as a felt or plastic, that is sewn to a backing material via parallel seams that are spaced apart, creating pockets between the two layers. Elongated flexible elements, which may be visualized as wires, are positioned in these pockets in order to hold the sleeves in place around the items on which they are placed.—EEU

6,876,715

43.40.Yq MEASUREMENT SYSTEM OF TORSION VIBRATION FOR REACTOR INTERNAL PUMP

Naoto Hikida and Seiichi Matsumura, assignors to Hitachi, Limited
5 April 2005 (Class 376/372); filed in Japan 19 July 2002

Reactor internal pumps are sealed and immersed in reactors, so that generally only a limited portion near one of the ends of a pump is available for vibration monitoring. Torsional vibrations are of concern, because matching of a motor frequency to a torsional frequency of the impeller system may lead to structural fatigue. An electrodynamic sensor (which produces a signal proportional to velocity) is affixed to the pump's cylindrical casing at a selected location so as to measure the axial vibrations. An accelerometer is placed at 90° from that location, oriented to measure vibrations in the tangential direction. The accelerometer's signal is integrated and low-pass filtered, then the axial vibration signal (which is unaffected by torsional vibrations) is subtracted from it, resulting in a measure of the torsional vibrations—EEU

6,866,474

43.50.Gf NOISE REDUCTION BY VORTEX SUPPRESSION IN AIR FLOW SYSTEMS

Robert B. Uselton, assignor to Lennox Industries, Incorporated
15 March 2005 (Class 415/121.2); filed 27 January 2003

This device was developed to reduce acoustic vibrations generated by von Kármán vortex streets by shaping members interposed in moving air flowstreams, such as fan guards and grilles, used in forced-flow air conditioning systems. Airflow systems are provided with structural modifications that reduce the formation of, and the shedding of, vortices on the downstream side of such structural elements. Typical configurations are relatively flat rectangular cross-section members with interrupted or nonlinear leading or trailing edges formed by approximately sawtooth or sinusoidal wave forms or connected to spaced-apart structural members, or cylindrical members formed in the shape of a sawtooth or sinusoidal form, or presented with separated rings or grooves interrupting the cylindrical cross section of the member.—DRR

6,877,585

43.55.Ev ACOUSTICAL CEILING TILES

Brandon Dillan Tinianov, assignor to Johns Manville International, Incorporated
12 April 2005 (Class 181/290); filed 30 April 2001

This sound-absorbing ceiling tile uses a first substrate of porous insulation material (like glass fiber, mineral wool, or foam) and a separate facing material (like plastic, semiporous paper, fabric, or perforated film). The two layers have different flow resistances to achieve broadband acoustical absorption.—CJR

6,870,934

43.58.Vb AUDIO LOUDSPEAKER DETECTION USING BACK-EMF SENSING

Andrew C. Krochmal et al., assignors to Visteon Global Technologies, Incorporated
22 March 2005 (Class 381/59); filed 15 July 2002

A method is described to verify that the audio speakers 11 are correctly connected in an automotive audio system during assembly in an environment which can be noisy. The invention uses a scheme that measures the back emf generated by the speakers from a repeatable, computer-controlled
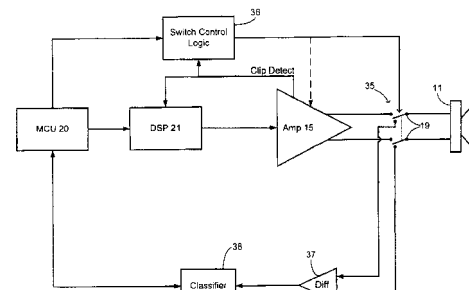
signal to automate the signal and test. This is an improvement over the use of a human test device (the repeatability and attitude is variable) and the door-slam method of generating a back emf signal.—NAS

6,838,994

43.58.Wc ADAPTIVE ALARM SYSTEM

Srinivas Gutta et al., assignors to Koninklijke Philips Electronics N.V.
4 January 2005 (Class 340/575); filed 26 October 2001

Some sleepers truly need disruptive alarm clocks. This invention monitors movement—if the sleeper has not stirred from the bed, then the
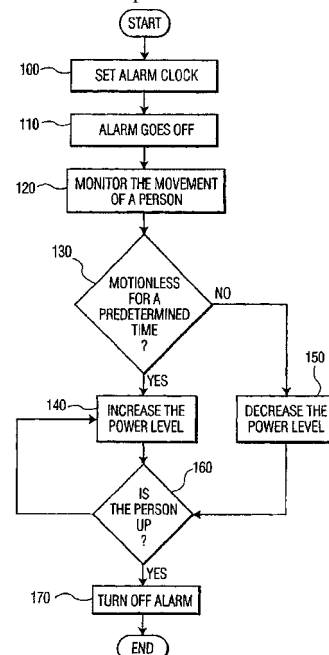
power level is increased as shown in the flowchart. Eventually, coupled with a suitable stimulus, it should work.—MK

6,847,587

43.60.Jn SYSTEM AND METHOD FOR IDENTIFYING AND LOCATING AN ACOUSTIC EVENT

Frank K. Patterson, Woodward, Oklahoma *et al.*
25 January 2005 (Class 367/127); filed 24 January 2003

A system is described by which a number of acoustic sensors may be deployed over a large area for the purpose of detecting and identifying gunshots within the monitored area and, more specifically, to determine the location of the shooter. This system differs from prior patented systems in that the individual sensors include a transmission capability, and so do not depend on wired connections. They also include more processing capability to perform some of the analysis before transmission to a central location. The individual sensors may also be equipped with GPS receiving devices, providing each sensor with accurate time and position information.—DLR

6,853,880

43.60.Jn AUTONOMOUS ACTION ROBOT

Yoshiaki Sakagami *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
8 February 2005 (Class 700/259); filed in Japan 22 August 2001

This robotic device would use both audio and visual inputs to determine the direction of an audio source, turn toward the direction of the sound, and identify objects, specifically one or more faces, at the target location,
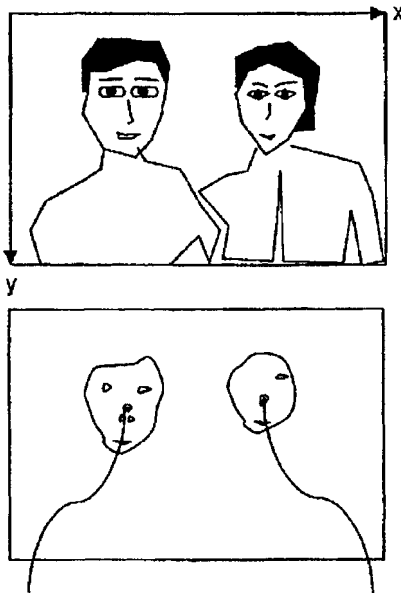
and then perform various analyses of the detected sources, such as facial expression analysis or speech recognition to extract commands or other information from a speech signal.—DLR

6,879,952

43.60.Jn SOUND SOURCE SEPARATION USING CONVOLUTIONAL MIXING AND A PRIORI SOUND SOURCE KNOWLEDGE

Alejandro Acero *et al.*, assignors to Microsoft Corporation
12 April 2005 (Class 704/222); filed 25 April 2001

One current approach to sound source separation that takes reverberation into account is convolutional independent component analysis. This frequency-domain approach can separate the various frequency components of different sound sources, but has a crucial permutation ambiguity: it is impossible to tell which separated component came from which sound

source. This leads to the possibility of reconstructing a source sound as a meaningless mixture of frequency components from several sources. The patented technique attempts to resolve this ambiguity using prior knowledge about the target source, in particular that it is human speech. A typical cepstral vector speech recognition system can be used by the convolutional scheme (specific methods are presented) to reconstruct a source which is most likely a proper speech signal instead of a garble of permuted components. Since this is computationally expensive, a more practical version is described using a vector quantization codebook of linear prediction vectors.—SAF

6,879,543

43.60.Pt ACOUSTIC PROCESSING FOR ESTIMATING SIZE OF SMALL TARGETS

Anthony A. Ruffa, assignor to The United States of America as represented by the Secretary of the Navy
12 April 2005 (Class 367/131); filed 7 August 2003

This is a signal processing technique for estimating the size of an object that is small compared to the wavelength of the sound it either radiates or reflects. A line array containing a number of receivers of that sound generates that number of independent signals from which a different number of time series are produced via the introduction of sets of time delays. These time series are further processed to produce information indicative of the size of the object in question.—WT

6,879,695

43.66.Ts PERSONAL SOUND LINK MODULE

Albert A. Maltan and Janusz A. Kuzma, assignors to Advanced Bionics Corporation
12 April 2005 (Class 381/315); filed 3 October 2002

In one embodiment, an easily removable module, containing acoustic input and output transducers, signal processing circuitry, wireless connection capability, and a rechargeable battery, is shaped for insertion into the retro-auricular space connected to the ear canal. An antenna and telemetry
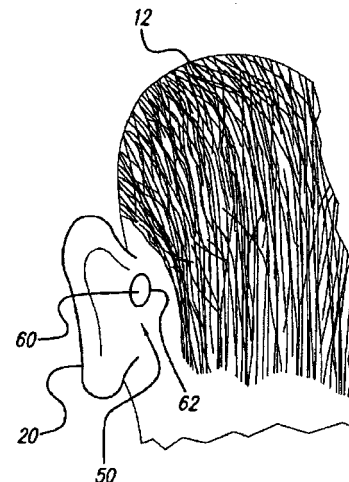
circuitry are included for communicating with a remote source such as a radio, computer, CD player, or television receiver.—DAP

6,876,750

43.66.Ts METHOD AND APPARATUS FOR TUNING DIGITAL HEARING AIDS

Rustin W. Allred et al., assignors to Texas Instruments Incorporated
5 April 2005 (Class 381/314); filed 28 September 2001

Methodology is provided to automatically tune the frequency response of a hearing aid within a specified tolerance of a predetermined target fre-

quency response. The desired target is calculated from an audiogram, which-may be input, or derived in real time, or from a direct input of the desired response.—DAP
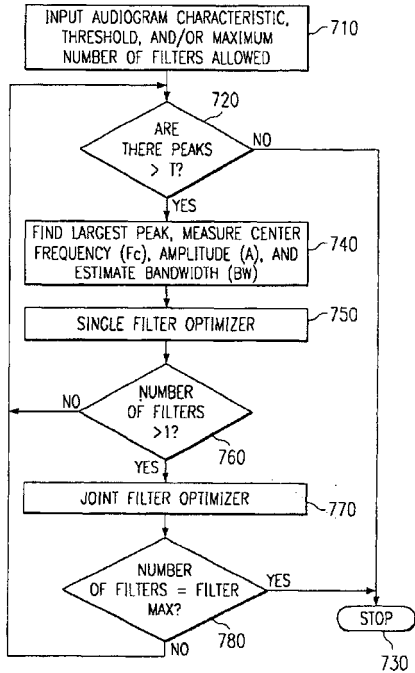
6,876,751

43.66.Ts BAND-LIMITED ADAPTIVE FEEDBACK CANCELLER FOR HEARING AIDS

Shawn X. Gao et al., assignors to House Ear Institute
5 April 2005 (Class 381/318); filed 30 September 1999

This patent adds a restricted bandwidth feature to existing feedback canceller configurations. The system is said to adapt more quickly and pro-

duce less audible artifacts by restricting the bandwidth of the feedback cancellation algorithm to frequency regions having unstable feedback.—DAP
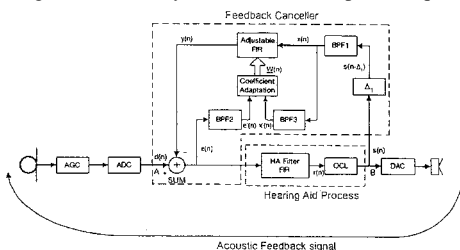
6,879,692

43.66.Ts HEARING AID WITH A SELF-TEST CAPABILITY

Kim Hjortgaard Nielsen and Lars Baekgaard Jensen, assignors to Widex A/S
12 April 2005 (Class 381/60); filed 9 July 2001

A test controller disconnects selected portions of a hearing aid and injects test signals at appropriate points in the signal path to diagnose whether the hearing aid is operating properly. For some tests, the hearing aid
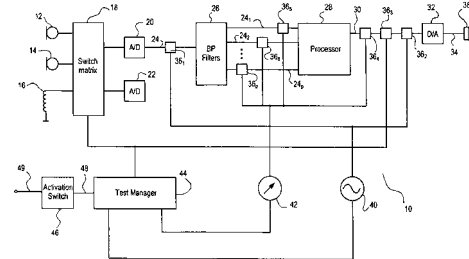
is placed in a compartment with hard walls to generate a reflected acoustic signal for use as an input. Level detectors in the test controller determine whether a defect exists.—DAP

6,876,749

43.66.Ts MICROPHONE FOR HEARING AID AND COMMUNICATIONS APPLICATIONS HAVING SWITCHABLE POLAR AND FREQUENCY RESPONSE CHARACTERISTICS

Mead C. Killion et al., assignors to Etymotic Research, Incorporated
5 April 2005 (Class 381/122); filed 11 July 2000

Building on an old idea, this patent proposes that circuitry sense the position of a mechanical shutter that closes and opens the rear tube of a switchable, directional/omnidirectional system. This feature allows the hear-

ing aid to optionally equalize frequency response or do other processing in either the directional or omnidirectional mode, something not included in the 1970s patents.—DAP
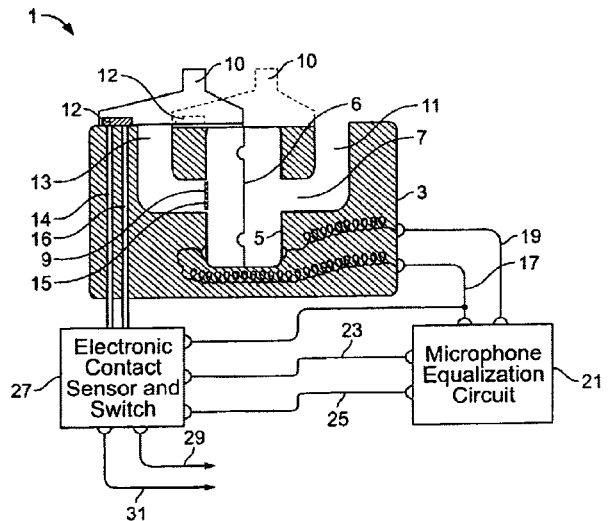
6,879,696

43.66.Ts IN-EAR HEARING AID AND METHOD FOR ITS MANUFACTURE

Andi Vonlanthen, assignor to Phonak AG
12 April 2005 (Class 381/322); filed 6 June 2000

A freely vibrating, acoustically transparent membrane protects the out-

put of a hearing aid from debris, such as ear wax, and reduces the need to clean the hearing aid.—DAP
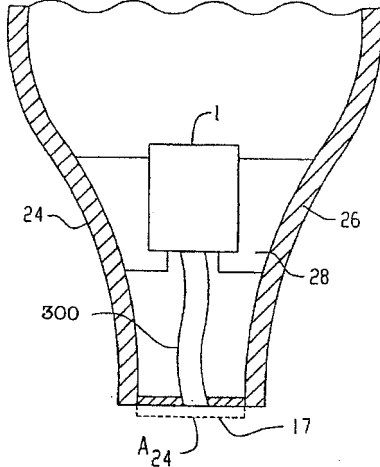
6,879,697

43.66.Ts HEARING AID WITH A FACE PLATE THAT IS AUTOMATICALLY MANUFACTURED TO FIT THE HEARING AID SHELL

Jan Tøpholm, assignor to Widex A/S
12 April 2005 (Class 381/322); filed in the European Patent Office 26 March 2001

To manufacture a custom hearing aid, digital data regarding the shape of the auditory canal are acquired and converted to a three-dimensional
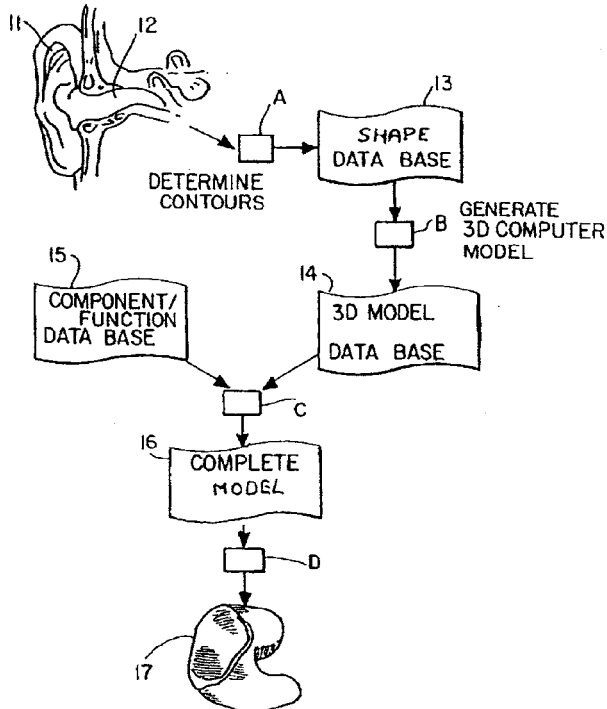
computer model of a hearing aid shell. Computer-controlled manufacture of a complete custom hearing aid housing is then performed, including a model of the faceplate.—DAP

6,785,645

43.72.Ar REAL-TIME SPEECH AND MUSIC CLASSIFIER

Hosam Adel Khalil *et al.*, assignors to Microsoft Corporation
31 August 2004 (Class 704/216); filed 29 November 2001

Both time domain and frequency domain features are combined to form a classifier. These features include spectral flux **VSF**, energy contrast **EC**, spectral centroid **VSC**, line spectrum pair correlation **VLSP**, and long-term prediction gain **LTP Gain**. As shown, these features can be used in a decision tree to distinguish music from speech. Short-time features (from the
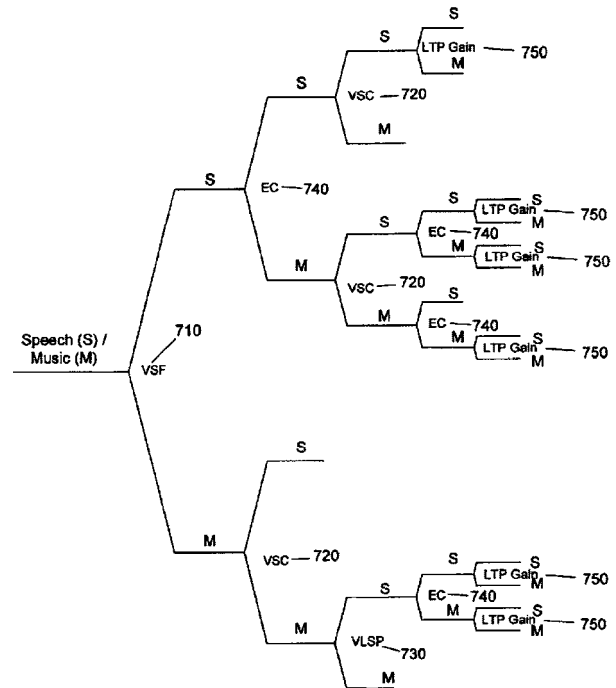
time domain) provide faster localization, whereas long-term features (like the frequency domain methods that use frame-based spectral estimation) offer more accurate classification. The patent text is reasonably clear and explores the tradeoffs between the two types of features.—MK

6,839,672

43.72.Ja INTEGRATION OF TALKING HEADS AND TEXT-TO-SPEECH SYNTHESIZERS FOR VISUAL TTS

Mark Charles Beutnagel *et al.*, assignors to AT&T Corporation
4 January 2005 (Class 704/260); filed 31 December 1998

This patent describes the integration of text-to-speech systems into the MPEG-4 video coding structure and also an apparently minor modification to the standard MPEG implementation consisting of a delineation of which elements of the synthesis process are to be driven by the transmitted data and which may be defined locally at the site of the video/audio re-creation. An additional enhancement consists of a method for smoothing the transitions from one synthesis element to the next. The patent includes detailed descriptions of the animation components, but, not being familiar with the MPEG standard, this reviewer cannot comment upon how much of that material is original in the patent.—DLR

6,876,969

43.72.Ja DOCUMENT READ-OUT APPARATUS AND METHOD AND STORAGE MEDIUM

Makiko Nakao, assignor to Fujitsu Limited
5 April 2005 (Class 704/260); filed in Japan 25 August 2000

A simple idea is proposed, though with limited detail, of using a standard speech synthesizer to fast-forward through a read-out of a document. The idea is to speed up and slow back down to a normal speech rate seamlessly at the user's indication, or perhaps by preset keywords in the document. The notion of rewinding during a document read-out session is also mentioned, though one wonders what rapidly reading backwards would sound like.—SAF

6,832,196

43.72.Ne SPEECH DRIVEN DATA SELECTION IN A VOICE-ENABLED PROGRAM

David E. Reich, assignor to International Business Machines
Corporation
14 December 2004 (Class 704/270.1); filed 30 March 2001

As the patent abstract says more clearly than the title, this is about using a speech recognizer to dynamically format the machine dialog for a voice-enabled database access system. In a network environment, the data-

base will typically reside on a server. A speech processor, also usually on the server, is connected using a voice-enabled markup language, such that the recognizer text output can be incorporated into the interaction dialog.—DLR
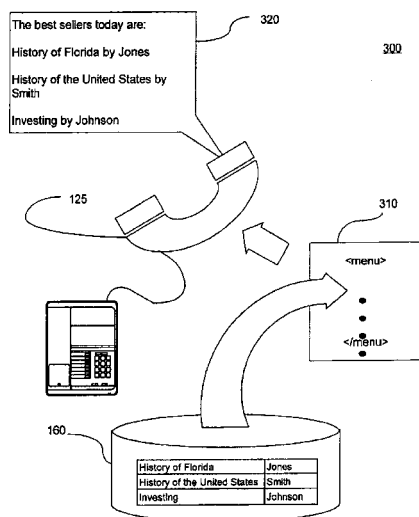
6,834,264

43.72.Ne METHOD AND APPARATUS FOR VOICE DICTATION AND DOCUMENT PRODUCTION

Mark Lucas et al., assignors to Provox Technologies Corporation
21 December 2004 (Class 704/235); filed 11 July 2001

This patent says that if you have a computer and a speech recognizer and you need to generate documents in a database, then you can use the recognizer to locate the database, and the record of interest, and produce text to be saved. An application is described in which a doctor would be able to produce records immediately instead of having to send out his notes to a transcriber. Not only is it obvious, but it's been done. This reviewer worked for a company in the 1990s which sold recognizers for medical transcription.—DLR

6,839,669

43.72.Ne PERFORMING ACTIONS IDENTIFIED IN RECOGNIZED SPEECH

Joel M. Gould et al., assignors to ScanSoft, Incorporated
4 January 2005 (Class 704/246); filed 10 June 1999

This very long patent covers two topics related to the use of speech recognition, each the subject of about half of the text and half of the claims.
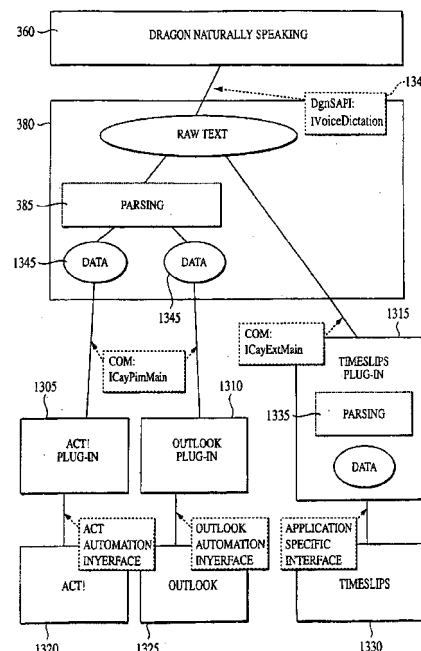
The first part deals with the mechanics of getting your voice into the machine. For example, you can use a hand-held recorder while you're away from the computer. These ideas all seem pretty obvious. The second part consists of extensive coverage of a parser to manage the continuous-speech input. The essence seems to be that a large set of finite-state "patterns" are matched against fragments of the input. These patterns are then tied into the semantic structures which govern how a phrase gets interpreted.—DLR

6,839,670

43.72.Ne PROCESS FOR AUTOMATIC CONTROL OF ONE OR MORE DEVICES BY VOICE COMMANDS OR BY REAL-TIME VOICE DIALOG AND APPARATUS FOR CARRYING OUT THIS PROCESS

Walter Stammler et al., assignors to Harman Becker Automotive
Systems GmbH
4 January 2005 (Class 704/251); filed in Germany 11 September 1995

This speech recognition patent describes the simultaneous use of speaker-dependent (SD) and -independent (SI) recognizers, which cooperate by working on the specific and the general aspects, respectively, of the speech input in order to accurately interpret the utterances. The primary application intended for such a system is in the operation of a vehicle. The SI recognizer uses a word template approach to construct a representation of the general form of an input command, while the SD recognizer uses feature analysis to identify specific words or phrases which have been trained by the particular user.—DLR

6,823,308

43.72.Ne SPEECH RECOGNITION ACCURACY IN A MULTIMODAL INPUT SYSTEM

Robert Alexander Keiller and Nicolas David Fortescue, assignors to Canon Kabushiki Kaisha
23 November 2004 (Class 704/256); filed in the United Kingdom 18 February 2000

This patent introduces the concept of multimodal input in a very general way, describing and claiming feature analyzers for various modalities as if they might be parallel to those needed for speech. The reality is a bit more mundane. The only input considered in detail, other than speech, is a mouse. Clicks would be detected and the mouse location used as the referent for demonstratives, such as "this" and "that." A finite-state syntactic structure is assumed to be up to the task of identifying such words and the verb structures which refer to them.—DLR

6,845,356

43.72.Ne PROCESSING DUAL TONE MULTI-FREQUENCY SIGNALS FOR USE WITH A NATURAL LANGUAGE UNDERSTANDING SYSTEM

Mark E. Epstein, assignor to International Business Machines Corporation
18 January 2005 (Class 704/235); filed 31 January 2001

A method is described for integrating a DTMF (Touch Tone®) signal decoder into a speech recognition facility. A step prior to the actual recognition would make a decision on whether or not the input included a DTMF tone. The actual tone decoding is well-known prior art. Once recognized, the tone code could be converted into any of several text forms for direct use by the text analysis portion of the recognition system. For example, a series of digits may be individual digits, a multidigit number, a telephone number, or a date. Each would be presented differently in the text form.—DLR

6,873,686

43.72.Ne NETWORK AND METHOD FOR PROVIDING A CALLING NAME TELECOMMUNICATIONS SERVICE WITH AUTOMATIC SPEECH RECOGNITION CAPABILITY

Zeeman Zhang and Joseph Knoerle, assignors to BellSouth Intellectual Property Corporation
29 March 2005 (Class 379/88.03); filed 9 August 2000

Enhanced features, such as calling name, call forwarding, and automatic recall, are enabled via predetermined voice commands over either landline or wireless telephones rather than requiring users to manually enter dial feature codes on a telephone or computer.—DAP

6,876,970

43.72.Ne VOICE-ACTIVATED TUNING OF BROADCAST CHANNELS

Edward Michael Silver and Linda Ann Roberts, assignors to BellSouth Intellectual Property Corporation
5 April 2005 (Class 704/275); filed 13 June 2001

The patent proposes an automatic method of channel selection via voice commands, either from all persons or only from authorized persons. Specific tuner broadcast channels are associated with appropriate voice commands, such as call letters or program categories, and stored. A DSP chip determines which channel to select, or group of channels to browse, from a voice command.—DAP

6,861,581

43.75.Gh ACOUSTIC GUITAR RESONATOR

John F. Geiger, Atlanta, Georgia
1 March 2005 (Class 84/294); filed 12 June 2003

The use of resonators in acoustic guitars dates back to at least the 1930s (see Schireson's United States Patent 1,887,861). The inventor
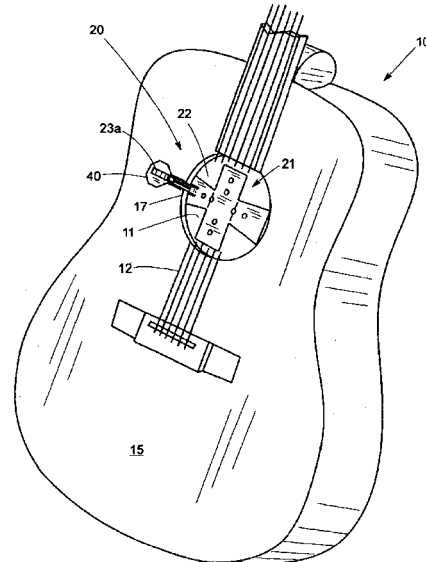
proposes that placing an iron-cross-shaped "sound collector" in the sound hole of a guitar will enable the player to obtain "quality sound and high volume."—MK

6,787,689

43.75.St FAST BEAT COUNTER WITH STABILITY ENHANCEMENT

Fang-Chu Chen, assignor to Industrial Technology Research Institute Computer & Communication Research Laboratories
7 September 2004 (Class 84/600); filed 1 April 1999

Beat tracking is an important task in machine listening. Scheirer used a filter bank approach while others, including Brown, used autocorrelation. The inventor wants a computationally simpler and faster algorithm.
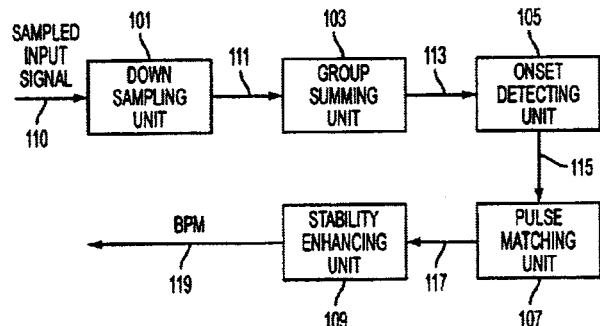
Accordingly, the input is downsampled, summed, and then smoothed and differentiated in block **105**. The beat tracking is done by looking at the values of other peaks in a 2-s frame (block **107**). The final stage tries to avoid tracking short-time variations in timing.—MK

1260   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Reviews of Acoustical Patents   1260

6,867,356

43.75.Wx MUSICAL TONE GENERATING APPARATUS, MUSICAL TONE GENERATING METHOD, AND PROGRAM FOR IMPLEMENTING THE METHOD

Tetsuo Nishimoto, assignor to Yamaha Corporation
15 March 2005 (Class 84/603); filed in Japan 13 February 2002

At issue is the loss of the upper spectrum if you undersample a signal. The inventor proposes modulating the lower spectrum up and then filtering it, scaling it, and adding it to the original (deficient) waveform. It's a simple enough method, but assumes that the upper spectrum is a copy of the lower.—MK

6,867,359

43.75.Wx MUSICAL INSTRUMENT CAPABLE OF CHANGING STYLE OF PERFORMANCE THROUGH IDLE KEYS, METHOD EMPLOYED THEREIN AND COMPUTER PROGRAM FOR THE METHOD

Shinya Koseki and Haruki Uehara, assignors to Yamaha Corporation
15 March 2005 (Class 84/615); filed in Japan 28 February 2003

Consider adding an electronic synthesizer to an acoustic piano, with the additional ability to silence the acoustic mechanism. If the range (written
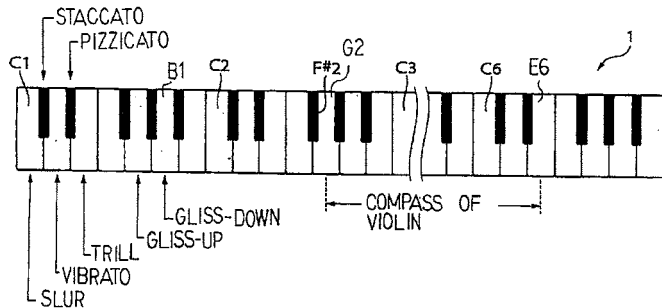
''compass'') of the instrument is smaller than the piano range, then you can use the other keys for various other functions like adding dynamics to the synthetic sounds.—MK

6,879,544

43.80.Ka MANATEE VOCALIZATION DETECTION METHOD AND SYSTEM

Christopher A. Sermarini, assignor to The United States of America as represented by the Secretary of the Navy
12 April 2005 (Class 367/136); filed 31 October 2003

This is a detection system geared to detecting a specific species of sea mammals, namely the manatee. Their vocalizations are discerned by executing a variety of frequency-domain processing operations on sound signals
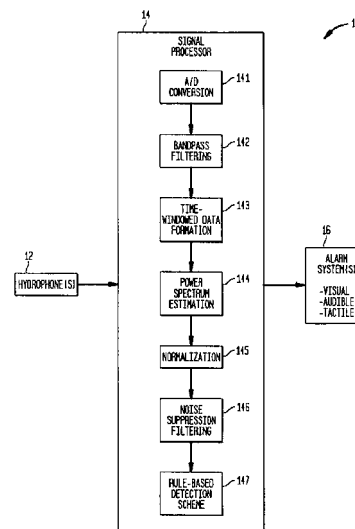
detected underwater. The overlapping, time-windowed segments of the digitized sound data are converted into the frequency domain. The spectra are then normalized and scored in a series of steps to indicate a manatee vocalization when a threshold score is achieved across a portion of the overlapping time-windowed segments.—DRR

6,869,404

43.80.Qf APPARATUS AND METHOD FOR CHRONICALLY MONITORING HEART SOUNDS FOR DERIVING ESTIMATED BLOOD PRESSURE

Randal C. Schulhauser *et al.*, assignors to Medtronic, Incorporated
22 March 2005 (Class 600/528); filed 26 February 2003

An implantable device, said to be minimally invasive, monitors a patient's blood pressure by recording heart sounds and deriving an estimated blood pressure on the basis of custom heart sound spectral analysis and processing. The device detects pressure waves with an acoustical sensor, preferably piezoelectric. The resulting analog signals are digitized and processed to determine blood pressure based on custom spectral analyses of the first and second heart sounds. The device is also equipped with subcutaneous ECG electrodes and a memory for storing ECG and heart sound data. A detected R-wave on a sensed ECG signal may be used to trigger sampling of the heart sound sensor signal for a specified period of time for detecting the first heart sound. A detected T-wave may be used to trigger sampling of the second heart signal. Heart sound and ECG are continuously sampled and stored in a looping memory based on a first-in, first-out basis. Calibration includes measuring a patient's blood pressure using a standard clinical method and performing regression analysis on multiple variables to identify a set of best fit weighted equations for predicting blood pressure. Concurrent ECG and estimated blood pressure may be displayed for review by a physician.—DRR

6,878,118

43.80.Qf ESOPHAGEAL STETHOSCOPE WITH SEPARATE COVER MEMBER

John L. Atlee III, Hartland, Wisconsin
12 April 2005 (Class 600/528); filed 6 February 2003

This is a really different sort of stethoscope, one that is inserted into the esophagus instead of being placed on a chest in the usual manner as with the conventional stethoscope. It is multifunctional in that it is capable of performing recording, monitoring, or stimulation functions from or within the esophagus, among these being transesophageal monitoring of cardiac activity, transesophageal cardiac stimulation (e.g., cardiac pacing), or
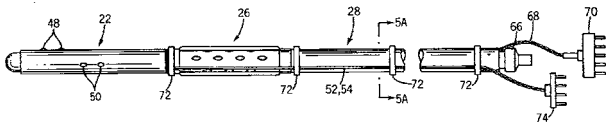
esophageal oximetry. The device consists of a flexible, tubular body incorporating an acoustic input region. A carrier member, separate from the tubular body, engages with the tubular body either proximally or distally of the acoustic input region. On the carrier member may be mounted one or more devices for enabling esophageal recording, monitoring, or stimulation functions. In order to effect cardiac pacing, electrodes may be mounted onto the carrier member.—DRR

6,866,670

43.80.Sh APPARATUS FOR REMOVING PLAQUE FROM BLOOD VESSELS USING ULTRASONIC ENERGY

Robert A. Rabiner and Brad A. Hare, assignors to OmniSonics Medical Technologies, Incorporated
15 March 2005 (Class 606/128); filed 6 August 2003

The patent covers a method and apparatus for removing plaque, fatty deposits, and other occlusions from blood vessels through the use of ultrasonic energy. A special application includes the removal of plaque from the cardoid artery in a nonthermal manner. The device is designed to present a small cross-sectional profile, thus allowing the apparatus to be used in a minimally noninvasive manner. The authors maintain that the device can be
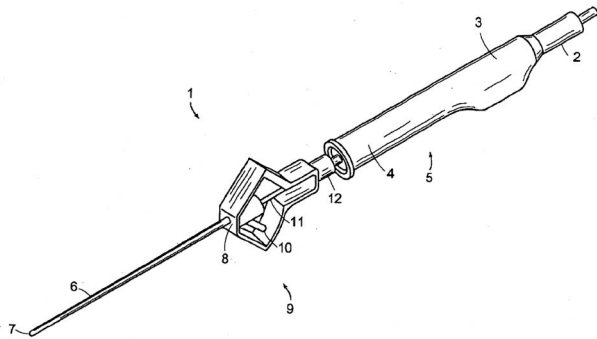
used in both surgical and outpatient treatment with minimal postoperative complications and minimal damage to areas other than the region of treatment. The ultrasonic probe may feature aspiration channels on its outer surface. An aspiration sheath may surround the ultrasonic probe in such a manner that the location of an aspiration port may be varied axially relative to the ultrasonic tip.—DRR

6,869,439

43.80.Sh ULTRASONIC DISSECTOR

Jeffrey S. White et al., assignors to United States Surgical Corporation; Misonix, Incorporated
22 March 2005 (Class 606/169); filed 19 August 2002

The patent disclosure relates to an ultrasonic instrument having an angled blade member and a clamp member that is said to be particularly suited for use in performance dissection and coagulation of tissue. The device includes a vibration coupler supported within a housing and operably connected to an ultrasonic generator. The angled blade member connects to
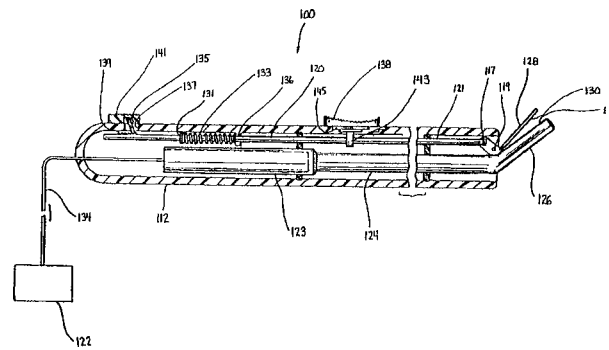
the distal end of the vibration coupler to conduct high-frequency vibration from the ultrasonic generator to the blade member. A clamp member placed adjacent to the blade member is movable to various positions. The clamp member and angled blade member act together to promote contact between the tissue and the blade member in the course of the operation of the instrument.—DRR

6,875,178

43.80.Vj RECEIVE CIRCUIT FOR ULTRASOUND IMAGING

Robert N. Phelps et al., assignors to Siemens Medical Solutions USA, Incorporated
5 April 2005 (Class 600/447); filed 27 June 2002

This circuit accommodates different signal formats from ultrasound transducers. The circuit adapts to time-division multiplexed signals by using an analog-to-digital converter that is synchronized with delays selected from
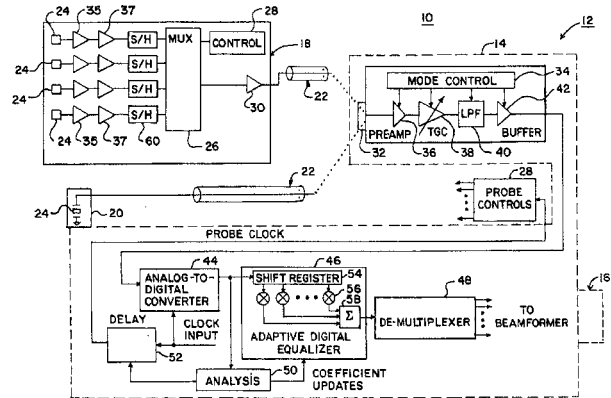
an analysis of test signals from the transducer. Filter characteristics of a digital equalizer are also selected from the analysis.—RCW

6,872,180

43.80.Vj DEVICE AND PROCESS FOR QUANTIFYING BODIES BY MEANS OF ULTRASOUND

Michael Reinhardt et al., assignors to Schering AG
29 March 2005 (Class 600/443); filed in Germany 28 March 2002

Ultrasonic images are obtained from overlapping sections of the object being examined to obtain sets of images from which the concentration of markers such as bubble contrast agents can be determined.—RCW

6,875,420

43.80.Vj METHOD OF ULTRASOUND IMAGING

Steven C. Quay, assignor to Amersham Health AS
5 April 2005 (Class 424/9.52); filed 18 July 1997

Gases are specially selected to form microbubbles for use as ultrasound contrast agents. The microbubbles have long life spans and are sufficiently small to traverse the lungs. The selection of gases is based on calculations that use the physical properties of gases. A way is described to determine microbubble lifetimes, given the properties of a gas.—RCW

6,876,879

43.80.Vj MAMMOGRAPHY METHOD AND APPARATUS

Kris A. Dines *et al.*, assignors to XDATA Corporation
5 April 2005 (Class 600/427); filed 21 March 2003

A three-dimensional image of a compressed breast of a subject is formed using an x-ray mammography unit and an ultrasound b-scan imaging unit. Fixtures permit spatial registration of the ultrasound and x-ray images. The apparatus includes a compression paddle through which ultrasound images are obtained.—RCW

6,878,115

43.80.Vj THREE-DIMENSIONAL ULTRASOUND COMPUTED TOMOGRAPHY IMAGING SYSTEM

Donald P. Dione *et al.*, assignors to Ultrasound Detection Systems, LLC
12 April 2005 (Class 600/459); filed 28 March 2003

This system uses an imaging chamber in which ultrasound transducer elements are arranged in a number of cylindrical rings. Cone-shaped beams are formed for transmission and reception. An algebraic reconstruction technique is used to produce images of sound speed and slope of attenuation.—RCW

LETTERS TO THE EDITOR

Aircraft noise annoyance in recreational areas after changes in noise exposure: Comments on Krog and Engdahl (2004) (L)

Ronny Klæboe[a)]
Institute of Transport Economics, Grensesvingen 7, Oslo, Norway

When Gardermoen replaced Fornebu as the main airport for Oslo, aircraft noise levels increased in recreational areas near Gardermoen and decreased in areas near Fornebu. Krog and Engdahl [J. Acoust. Soc. Am. **116**, 323–333 (2004)] estimate that recreationists' annoyance from aircraft noise in these areas changed more than would be anticipated from the actual noise changes. However, the sizes of their estimated "situation" effects are not credible. One possible reason for the anomalous results is that standard regression assumptions become violated when motivational factors are inserted into the regression model. Standardized regression coefficients (beta values) should also not be utilized for comparisons across equations. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1984787]

I. INTRODUCTION

There are not many socioacoustic studies published on the effect of aircraft noise on outdoor recreationists, and papers bringing in results in this research area are especially welcome. They provide information on the importance of silent areas having recuperative effects, and establish results that are important input for more precise research hypotheses, improved data collection, and better statistical analyses. However, models of recreationists' annoyance reactions when undertaking voluntary trips in an area exposed to aircraft noise are conceptually more complex than the simpler dose response relationships for residential noise annoyance. The structure of such models should therefore be considered carefully.

In the paper: "Annoyance with aircraft noise in local recreational areas, contingent on changes in exposure and other context variables," Krog and Engdahl (2004) link a set of explanatory variables to aircraft noise annoyance in recreational areas. The data for the analyses come from two sets of before-after surveys; each set consisting of separate samples (not panel data). The first dataset stems from a study area close to Fornebu airport before and after it was closed down. The second set is from a study area near Gardermoen airport before and after it replaced Fornebu as the main airport for Oslo and Southeast Norway. In all surveys the equivalent aircraft noise exposure values $L_{Aeq,air}$ for the duration of each recreationist's trip as well as $L_{Asel,air}$ values

were calculated on the basis of measurements at fixed locations at each interview site deemed representative for the respective walking areas. An additional aircraft noise exposure indicator, the percentage of time aircraft noise was audible during each trip, was also calculated from these measurements.

The socio-acoustic surveys were designed to capture possible changes in the aircraft noise annoyance experienced by recreationists in two "situations:"

(1) After a marked decrease in $L_{Aeq,air}$ values (average reduction *ca.* 22 dB A)
(2) After a slight increase in $L_{Aeq,air}$ values (average increase *ca.* 1 dB A)

II. REGRESSION ANALYSES YIELD VERY HIGH "SITUATION" EFFECTS

Krog and Engdahl (2004) analyzed each of the studies (before and after) separately by means of multiple regression analyses. The degree of aircraft noise annoyance experienced during a recreational trip was the dependent variable, while the independent variables were as follows:

(1) $L_{Aeq,air}$.
(2) Duration of the visit.
(3) Whether the area is the primary recreational area or not.
(4) Reason for participating: The acoustic nature.
(5) Reason for participating: The visual nature experience.
(6) An indicator (a dummy variable) of whether the data stem from the after survey or not and that may capture a "situation" effect.

[a)]Electronic mail: rk@toi.no

The indices for participation in recreational trips in general (not the specific trip) were chosen as "mediator" variables.

Based on the results from the multiple regression analyses, Krog and Engdahl (2004) conclude that there is evidence of "situation" effects. People were less annoyed after the removal of air traffic from Fornebu, and people became more annoyed following the modest increase in average $L_{Aeq,air}$ in Romeriksåsen than predicted from the primary explanatory variables alone (1–5). However, the sizes of these "situation" effects are not credible. Simple calculations based on the published results, Krog and Engdahl (2004: Tables II and III), show that the "situation" accounts for a reduction in aircraft noise annoyance comparable to that of a 134 dB A reduction in $L_{Aeq,air}$ near Fornebu Airport. [The figure of −134 dB is obtained by dividing the estimated effect size of the "situation" (−1.476), with the estimated effect size of the aircraft noise exposure indicator (0.011)]. Vice versa, Gardermoen becoming the main airport for Oslo results in a "situation" effect comparable to a 21 dB A increase in $L_{Aeq,air}$ (0.398/0.019). Both effect sizes are in excess of what one could expect from the relatively well-behaved dose-response relationships for the combined datasets that are illustrated in Krog and Engdahl (2004: Figs. 4 and 5). In the Fornebu before situation the recreationists were exposed to a median $L_{Aeq,air}$ value of 69 dB. According to Krog and Engdahl (2004: Fig. 4), these recreationists, on average, became *rather annoyed*. In the complex multivariate regression model, aircraft noise annoyance in the Fornebu before the situation becomes mainly "explained" by the regression *constant—see* Krog and Engdahl (2004: Table II). The regression model estimated by Krog and Engdahl thus attempts to predict why recreationists become *not annoyed* by aircraft noise annoyance, but fails to explain why they *were* annoyed.

III. THE CONCEPTUAL MODEL IS COMPLEX, AND STANDARD REGRESSION ASSUMPTIONS MAY BE VIOLATED

Taking a trip in a recreational area exposed to aircraft noise is a voluntary decision. For analyzing this decision, it is clearly necessary to have information from recreationists that decide against taking a trip, or choose alternative recreational areas for taking the trip. The effect of aircraft noise on noise annoyance that is estimated is thus contingent on the fact that the recreationists have decided to take a trip in spite of the aircraft noise exposure in the area. The reason for recreational trips may be to experience silence, or the recreationist may have other reasons. It is therefore easy to sympathize with the efforts of Krog and Engdahl (2004) in bringing motivational factors into the conceptual model. No doubt, the reason for taking recreational trips may influence the experience of the trip.

However, incorporating motivational factors among the independent variables for explaining aircraft noise annoyance during the trip yields a highly complex explanatory model. All questionnaire items were filled in "on their way out of the area" and thus in the context of the specific trip. The recreationists' attitudes, noise sensitivity and personality traits may affect both motivation and annoyance. Such non-

modeled factors affecting both independent and dependent variables violate the standard regression assumption of independence between the error term and the dependent variable. The estimated impacts of the motivational factors on aircraft noise annoyance may thus be spurious and the integrity of the regression models compromised. An additional challenge is that when the variation in aircraft noise exposure is small or actual noise exposure has little effect on aircraft noise annoyance, the variation in the aircraft noise annoyance responses mainly reflects individual or other differences between the recreationists. The causal directions of the analytic regression model motivated by the presupposed dose-response relationship then need reexamining. Using the trip duration purely as an independent variable could also be considered problematic.

One possible approach for disentangling the impacts of the different variables is to estimate and report the results from simpler explanatory models before estimating the more complex models. In particular, it would be helpful to know the explanatory power and full report of a model where the exposure variable $L_{Aeq,air}$ is the sole predictor for aircraft noise annoyance. Thereafter models statistically controlling for situation and, possibly, whether the area is the primary recreational area could be established. This would improve transparency and make it easier to follow the implications of the changes in the estimated coefficients and possible quality improvement from introducing additional explanatory variables.

IV. THE FALLACY OF COMPARING STANDARDIZED COEFFICIENTS, BETA VALUES

Krog and Engdahl (2004) are also interested in comparing the effect sizes of the different aircraft noise exposure indicators. For such comparisons, the relative lack of variation in the exposure values in the studies near Gardermoen poses a problem. All recreationists are exposed to fairly low noise levels ($L_{Aeq,air}$) during the trip. To compare aircraft noise exposure indicators, an adequate exposure range for all proposed exposure indicators is required. This is not the case near Gardermoen, and measurement errors and factors other than exposure may play a large role in determining a model fit. As the comparisons of aircraft noise exposure indices are not fully reported, it is difficult to assess the quality of the results.

In an attempt to reduce the problems of comparing different aircraft noise exposure measures with different metrics, Krog and Engdahl (2004) establish separate (complex) regression models, standardize the coefficients (obtain beta values), and proceed to compare the estimated beta values for the different aircraft noise exposure indicators. The exposure indicator having the larger estimated beta value is deemed "better" than an indicator having a lower estimated beta value.

When comparing the explanatory variables that are part of the *same* multiple regression model, beta values may indicate which variable has the largest relative impact on the dependent variable. However, the use of beta values for comparing effect sizes between *different* regression equations is
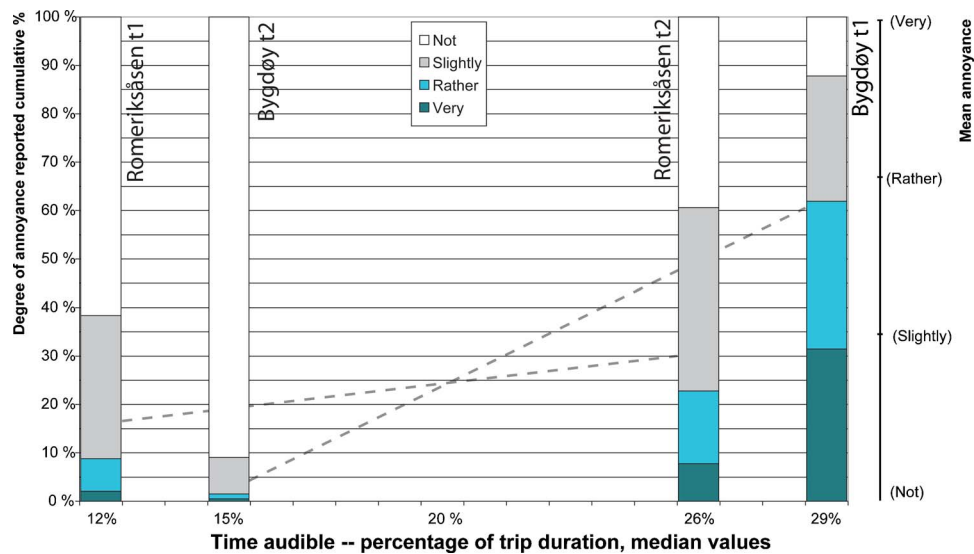
FIG. 1. (Color online) The proportion of recreationists in two before and after surveys reporting different degrees of aircraft noise annoyance as a function of the percentage of the trip duration that aircraft noise was audible. Mean annoyance response changes indicated by broken lines—annoyance scale to the right.

misguided. The results of such comparisons not only depend on effect size, but also the variation in exposure values. The warning from Greenland *et al.* (1986) against utilizing standardized coefficients to compare results from different analyses is particularly pertinent when the variation in some of the exposure variables is small.

For a dummy variable such as the "situation" effect indicator, the standardized regression coefficient becomes a direct function of the proportion of interviews made in the before and after surveys. This affects both the within-equation and between-equation comparisons, and the beta effect sizes become directly dependent on the relative number of respondents in the before and after survey.

V. TIME AUDIBLE AS A PREDICTOR OF AIRCRAFT NOISE ANNOYANCE

As the methodology for analyzing the impacts of "time audible" on aircraft noise annoyance is the same as for $L_{Aeq,air}$, these results also need to be reexamined. The aggregate results reported by Krog and Engdahl (2004: Figs. 2 and 3) are summarized in Fig. 1. As simple comparisons violate the situational approach advocated by Krog and Engdahl, the relevant changes in mean annoyance responses are indicated using dashed lines. The disparate results from both simple and context sensitive comparisons of the dose-response relationships, suggest one should exercise caution in relying too much on estimation results for partial relationships and that researchers need consider all four datasets when attempting to draw conclusions from the data.

VI. THE DATASETS COULD BE MERGED FOR IMPROVED ANALYSES

The authors analyze the Fornebu and Gardermoen studies separately. The rationale provided is that the situation

effect has different signs in each of the two studies. If we disregard the problems with the conceptual model that have already been discussed in this paper, it would have been possible to incorporate two dummy variables in the multiple regression model instead of one. This would capture all three alternatives: "steady state," "noise increase," or "noise decrease" and data from both before and after studies could have been combined. The whole dataset could then have been utilized to estimate the parameter values for the variables common to both analyses.

VII. CONCLUSION

The sizes of the "situation" effects (removal of air traffic from Fornebu and added traffic at Gardermoen airport) are unbelievably large and the analyses should be reexamined. They authors should provide more detailed information about their conceptual models and provide a rationale for the multiple regression models utilized for the analyses. If possible, a sequential approach reporting fully the results from simpler explanatory models that serve as baselines before proceeding to more complex models could be adopted. This would improve transparency and facilitate the reader's understanding of the contribution of the various explanatory factors in the model. The conclusions reached by comparing standardized regression coefficients should also be reexamined.

Greenland, S., Schlesselman, J. J., and Criqui, M. H. (**1986**). "The fallacy of employing standardized regression coefficients and correlations as measures of effect," J. Epidemiol. **123**, 203–208.

Krog, N. H., and Engdahl, B. (**2004**). "Annoyance with aircraft noise in local recreational areas, contingent on changes in exposure and other context variables," J. Acoust. Soc. Am. **116**, 323–333.

A Wigner approximation method for wave propagation (L)

Patrick J. Loughlin

University of Pittsburgh, Department of Electrical and Computer Engineering, 348 Benedum Hall, Pittsburgh, Pennsylvania 15261

Leon Cohen

City University of New York, Department of Physics, 695 Park Avenue, New York, New York 10021

An approximation method for pulse propagation based on the Wigner position-wavenumber representation is presented. The method is very easy to apply and moreover is physically illuminating. One obtains the evolved approximate Wigner distribution from the initial Wigner distribution by a simple linear translation in phase space. Each phase space point propagates at constant velocity given by the group velocity at the phase space point. Dissipative propagation (damping) is also taken into account. From the approximate Wigner distribution, one can obtain the approximate magnitude of the evolved pulse and the approximate local wavenumber, that is, the spatial derivative of the phase of the pulse. Examples are given to illustrate the method. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2001488]

I. INTRODUCTION

The fundamental nature of dispersive wave propagation is that different frequencies travel at different velocities, and hence one would suspect that a transformation of the pulse into a joint position-wavenumber (x,k) representation would be well suited to study such dispersive propagation. We present an approximation method for dispersive propagation based on the position-wavenumber Wigner representation of the pulse. The approximation is remarkably simple to apply and gives considerable insight into the nature of pulse propagation. For simplicity of presentation, we focus on the case of one spatial dimension and single mode propagation, but the approach extends to multiple dimensions and to multiple modes, as we discuss subsequently.

Our method is as follows. Starting with the initial pulse $u(x,0)$, one calculates the Wigner distribution at time zero, $W(x,k,0)$. Then, the approximate evolved Wigner distribution is obtained easily according to

$$W(x,k,t) \approx e^{2\omega_I(k)t}W(x-v_g(k)t,k,0), \qquad (1)$$

where $v_g(k)$ is the group velocity, as usually defined, and $\omega_I(k)$ is the imaginary part of the dispersion relation. Moreover, the approximate magnitude $|u(x,t)|$ and local wavenumber, $\langle k \rangle_{x,t}$, of the pulse $u(x,t)$, are obtained by

$$|u(x,t)|^2 \approx \int W(x,k,t)dk, \qquad (2)$$

$$\langle k \rangle_{x,t} \approx \frac{1}{\int W(x,k,t)dk} \int kW(x,k,t)dk, \qquad (3)$$

where the local wavenumber is defined as $\langle k \rangle_{x,t} = \partial \varphi(x,t)/\partial x$ and $\varphi(x,t)$ is the phase of the pulse, $u(x,t) = R(x,t)e^{i\varphi(x,t)}$.

This approximation, Eq. (1), is remarkably simple: other than to calculate the Wigner distribution of the initial wave, one does not have to do any further calculations—one merely substitutes $x-v_g(k)t$ for x in the initial Wigner distribution and multiplies by $e^{2\omega_I(k)t}$. Moreover, besides being easily obtained, this approximation is interesting and insightful because it is physically interpretable. It shows that the Wigner distribution of the wave evolves (approximately) in the following simple manner. At each point in phase space the point evolves with constant velocity motion where the velocity is the group velocity. It also shows in a simple way the effect of damping and, in particular, it shows the growth or decay of a particular wavenumber: for $\omega_I(k) < 0$ we have damping (energy dissipation), and for $\omega_I(k) > 0$ energy is added to the wave as it propagates (for $t > 0$).

II. WAVE PROPAGATION AND APPROXIMATION

We describe how the above approximations arise. For an initial pulse $u(x,0)$, the pulse at a later time, $u(x,t)$, is given by[1–5]

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int S(k,0)e^{ikx-i\omega(k)t} dk, \qquad (4)$$

for each mode, where $S(k,0)$ is the initial spatial spectrum,

$$S(k,0) = \frac{1}{\sqrt{2\pi}} \int u(x,0)e^{-ikx}dx \qquad (5)$$

and $\omega(k)$ is the dispersion relation, which we write in terms of its real and imaginary parts,

$$\omega(k) = \omega_R(k) + i\omega_I(k). \qquad (6)$$

As commonly defined, the group velocity, $v_g(k)$, is given by

$$v_g(k) = \frac{d\omega_R(k)}{dk}. \qquad (7)$$

The position-wavenumber Wigner distribution of the pulse is defined by[6–8]

$$W(x,k,t) = \frac{1}{2\pi} \int u^* \left(x - \frac{\lambda}{2}, t \right) u \left(x + \frac{\lambda}{2}, t \right) e^{-i\lambda k} \, d\lambda. \tag{8}$$

By substituting Eq. (4) into Eq. (8), one can show that the Wigner distribution at time t is related to the Wigner distribution of the initial pulse at $t=0$ by[9,10]

$$W(x,k,t) = \frac{1}{2\pi} \int \int W(x',k,0)$$
$$\times e^{i\theta(x'-x)} e^{i[\omega^*(k+\theta/2) - \omega(k-\theta/2)]t} \, d\theta \, dx'. \tag{9}$$

This equation is exact. Equation (1) is obtained by expanding the phase and keeping only low-order terms.[9,10]

To obtain the approximate magnitude and local wavenumber of the pulse $u(x,t)$, first note that from the exact Wigner distribution $W(x,k,t)$, one obtains the squared-magnitude and local wavenumber of the pulse via[6]

$$\int W(x,k,t) dk = |u(x,t)|^2 \tag{10}$$

and

$$\langle k \rangle_{x,t} = \frac{1}{\int W(x,k,t) dk} \int k W(x,k,t) dk = \frac{\partial \varphi(x,t)}{\partial x}, \tag{11}$$

respectively. Therefore, the approximate magnitude and local wavenumber are obtained by using the approximate Wigner distribution in the equations above.

Before illustrating the approximation with examples, we comment on the extension to multiple spatial dimensions and to multiple modes. For a three-dimensional scalar field, the Wigner distribution is defined as

$$W(\mathbf{r},\mathbf{k},t) = \left(\frac{1}{2\pi} \right)^3 \int u^* \left(\mathbf{r} - \frac{\mathbf{r}'}{2}, t \right) u \left(\mathbf{r} - \frac{\mathbf{r}'}{2}, t \right) e^{-i\mathbf{r}'\cdot\mathbf{k}} \, d\mathbf{r}', \tag{12}$$

and the approximation becomes

$$W(\mathbf{r},\mathbf{k},t) \approx e^{2\omega_I(\mathbf{k})t} W(\mathbf{r} - \mathbf{v}_g(\mathbf{k})t, \mathbf{k}, 0), \tag{13}$$

where $\mathbf{v}_g(\mathbf{k})$ is the group velocity vector.

For the multiple mode case, one has an option depending on the use of the approximation. Generally, as with the exact solution or stationary phase method, one studies one mode at a time and that is also the case with the Wigner approximation, as presented. If one considers instead the total solution (i.e., the sum of modes) then, like the autocorrelation function and spectral density, the Wigner distribution and approximation produce cross-products between the modes. One could develop a Wigner approximation that drops the cross-products, and that may or may not be more accurate, depending on whether the modes add coherently or incoherently. Also, we make the following observation regarding modes. Over the past 30 years the concept of multicomponent signals has arisen in signal processing. A multicomponent signal is one that breaks up into distinct parts in phase space/the time-frequency plane.[6,11] One of the very significant advantages of using a phase space approach to study wave propagation is that it can show in a simple way that a wave is multicomponent and allows one to extract information about each component. Since each mode typi-

cally travels with a different group velocity, the modes increasingly separate the further the wave propagates, that is as time increases.

III. EXAMPLES

A. Example 1: Quadratic dispersion

We consider first an exactly solvable case to be able to compare effectively. For the initial pulse and dispersion relation we take, respectively,

$$u(x,0) = e^{i\beta x^2/2 + ik_0 x}, \tag{14}$$

$$\omega(k) = (\gamma_R + i\gamma_I) k^2/2 = \gamma k^2/2. \tag{15}$$

The group velocity is

$$v_g(k) = \gamma_R k. \tag{16}$$

Working out the exact solution, one obtains

$$u(x,t) = \frac{1}{\sqrt{1+\gamma\beta t}} \exp \left(i \frac{\beta}{2} \frac{(x - k_0\gamma t)^2}{1 + \gamma\beta t} + ik_0 x - i\gamma k_0^2 t/2 \right) \tag{17}$$

$$= \frac{1}{\sqrt{1+\gamma\beta t}} \exp \left[\frac{i}{2\beta} \left(\frac{(\beta x + k_0)^2}{1 + \gamma\beta t} - k_0^2 \right) \right]. \tag{18}$$

The exact magnitude and local wavenumber are

$$|u(x,t)|^2 = \frac{1}{\sqrt{(1 + \gamma_R\beta t)^2 + (\gamma_I\beta t)^2}}$$
$$\times \exp \left[\gamma_I t \frac{(\beta x + k_0)^2}{(1 + \gamma_R\beta t)^2 + (\gamma_I\beta t)^2} \right], \tag{19}$$

$$\frac{\partial\varphi(x,t)}{\partial x} = \frac{1 + \gamma_R\beta t}{(1 + \gamma_R\beta t)^2 + (\gamma_I\beta t)^2} (\beta x + k_0). \tag{20}$$

For the Wigner approximation approach, the Wigner distribution of the initial pulse is readily calculated,

$$W(x,k,0) = \delta(k - \beta x - k_0). \tag{21}$$

The approximate Wigner distribution, Eq. (1), follows immediately,

$$W(x,k,t) \approx e^{\gamma_I k^2 t} \delta(k - \beta(x - \gamma_R k t) - k_0) \tag{22}$$

$$= \frac{1}{|1 + \gamma_R\beta t|} e^{\gamma_I k^2 t} \delta \left(k - \frac{\beta x + k_0}{1 + \gamma_R\beta t} \right) = W_a(x,k,t). \tag{23}$$

The approximate magnitude and local wavenumber of the pulse are

$$|u(x,t)|^2 \approx \frac{1}{|1 + \gamma_R\beta t|} \exp \left[\gamma_I t \left(\frac{\beta x + k_0}{1 + \gamma_R\beta t} \right)^2 \right] \tag{24}$$

and

$$\frac{\partial\varphi(x,t)}{\partial x} = \langle k \rangle_{x,t} \approx \frac{\int k W_a(x,k,t) dk}{\int W_a(x,k,t) dk} = \frac{\beta x + k_0}{1 + \gamma_R\beta t}. \tag{25}$$

For comparison, we also give the stationary phase approximation.[1–5] Usually the stationary phase approximation is presented in terms of real dispersion relations. Since,

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

P. J. Loughlin and L. Cohen: Letters to the Editor   1269

in our case, the dispersion relation is complex, we re-write Eq. (4) explicitly in terms of the real and imaginary parts of the dispersion relation as

$$u(x,t) = \frac{1}{\sqrt{2\pi}} \int S(k,0) e^{\omega_I(k)t} e^{ikx - i\omega_R(k)t} dk. \quad (26)$$

Applying the method of stationary phase to approximate this integral yields

$$u(x,t) \sim \frac{S(k_s,0) e^{\omega_I(k_s)t}}{\sqrt{|t\omega_R''(k_s)|}} e^{j(k_s x - \omega_R(k_s)t - (\pi/4)\mathrm{sgn}(t\omega_R''(k_s)))} \quad (27)$$

where k_s are the stationary points, which are obtained by solving $x - \omega_R'(k_s)t = 0$ for k_s. For the specific wave and dispersion relation considered in this example, we obtain

$$u(x,t) \sim \sqrt{\frac{i}{|\gamma_R \beta t|}} \exp\left[\frac{\gamma_I x^2}{2\gamma_R^2 t}\right]$$
$$\times \exp\left[i\left(\frac{x^2}{2\gamma_R t} - \frac{\left(\frac{x}{\gamma_R t} - k_0\right)^2}{2\beta} - \frac{\pi}{4}\mathrm{sgn}(t\gamma_R)\right)\right], \quad (28)$$

by which the approximate magnitude and local wavenumber are

$$|u(x,t)|^2 \approx \frac{1}{|\gamma_R \beta t|} \exp\left[\frac{\gamma_I x^2}{\gamma_R^2 t}\right], \quad (29)$$

$$\frac{\partial \varphi(x,t)}{\partial x} \approx \frac{\beta x + k_0}{\gamma_R \beta t} - \frac{\beta x}{(\gamma_R \beta t)^2}. \quad (30)$$

We now briefly compare some aspects of the approximations to the exact answer. We first note that for both the Wigner approximation and the stationary phase method, knowledge of the dispersion relation is required, as is the case, of course, to obtain the exact solution. For this example, note that if there is no damping ($\gamma_I = 0$), the Wigner approximation is exact, but the stationary phase approximation is not. When there is damping, the Wigner approximation captures the main exponential decay or growth of the pulse magnitude, and it is seen to be an improvement over the stationary phase approximation. Regions of the parameters and time can be studied in detail, but we only note here that by comparing Eq. (19) with Eq. (24), and Eq. (20) with Eq. (25), we see that the Wigner approximation is generally very good when

$$\left(\frac{\gamma_I \beta t}{1 + \gamma_R \beta t}\right)^2 \ll 1. \quad (31)$$

For example, if we consider the long time limit $t \gg 1/\gamma_R\beta$, then we see that the approximation is good when $\gamma_R \gg \gamma_I$, where we have taken all the parameters positive. For small times $t \ll 1/\gamma_R\beta$, the approximation is good when $\gamma_I\beta t \ll 1$. The stationary phase approximation is likewise accurate for $t \gg 1/\gamma_R\beta$ when $\gamma_R \gg \gamma_I$, although the Wigner approximation approaches the exact answer faster. For small times, the stationary phase approximation is not accurate. In a future paper, further discussion and exactly solvable examples will be presented, but we consider next one additional example to illustrate the simplicity of the Wigner approximation coupled with the insights into propagation that it provides.

B. Example 2: Pulse in a waveguide

Consider a pulse propagating in a two-plate waveguide, for which the dispersion relation is[1,5,12]

$$\omega(k) = c\sqrt{k^2 + \kappa_m^2} = ck\sqrt{1 + (\kappa_m/k)^2} \quad (32)$$

where $\kappa_m = m\pi/D$ corresponds to the lowest frequency $\omega_m = \kappa_m c$ per mode m that propagates in the waveguide, D is the separation distance of the plates, and c is the propagation velocity as $D \to \infty$. For the initial pulse we take

$$u(x,0) = \left(\frac{\alpha}{\pi}\right)^{1/4} e^{-\alpha x^2/2 + ik_0 x + i\beta x^2/2}, \quad (33)$$

where x is the range. The initial Wigner distribution is

$$W(x,k,0) = \frac{1}{\pi} \exp[-\alpha x^2 - (k - \beta x - k_0)^2/\alpha]. \quad (34)$$

The group velocity is $v_g = c/\sqrt{1 + (\kappa_m/k)^2}$ and therefore the approximate Wigner distribution per mode is

$$W(x,k,t) \approx \frac{1}{\pi} \exp\left[-\alpha\left(x - \frac{ct}{\sqrt{1 + (\kappa_m/k)^2}}\right)^2\right.$$
$$\left. - \frac{1}{\alpha}\left(k - k_0 - \beta\left(x - \frac{ct}{\sqrt{1 + (\kappa_m/k)^2}}\right)\right)^2\right]. \quad (35)$$

We do not give the stationary phase approximation but invite the reader to try it. In Fig. 1, we plot the initial Wigner distribution, and the approximation for the first mode ($m = 1$) at subsequent times. Note how the pulse contracts for a period before expanding, and changes from a downward chirp to an upward chirp as it propagates.

IV. CONCLUSION

We have presented an approximation for dispersive wave propagation, based on the Wigner distribution representation of the wave. From the Wigner distribution of the initial wave, one obtains an approximate Wigner representation of the wave at some later time in a very simple way. From the approximate Wigner representation, approximations of the magnitude and local wavenumber of the wave $u(x,t)$ can be obtained by integrals of the Wigner approximation. In addition to its simplicity, the Wigner approximation provides insights into the nature of dispersive propagation, and shows the effect of dispersion and damping on a local point in phase space. Also, the case of random media can be developed by considering the expected value of the Wigner distribution.[13,14]

We comment briefly on computational issues. As mentioned previously, the computational requirements of the Wigner approximation involve only the need to compute the Wigner distribution of the initial wave. We emphasize that this calculation does not involve the dispersion relation, unlike, for example, the stationary phase approximation. The approximate time evolved Wigner distribution is then obtained very simply via Eq. (1), which takes into account the
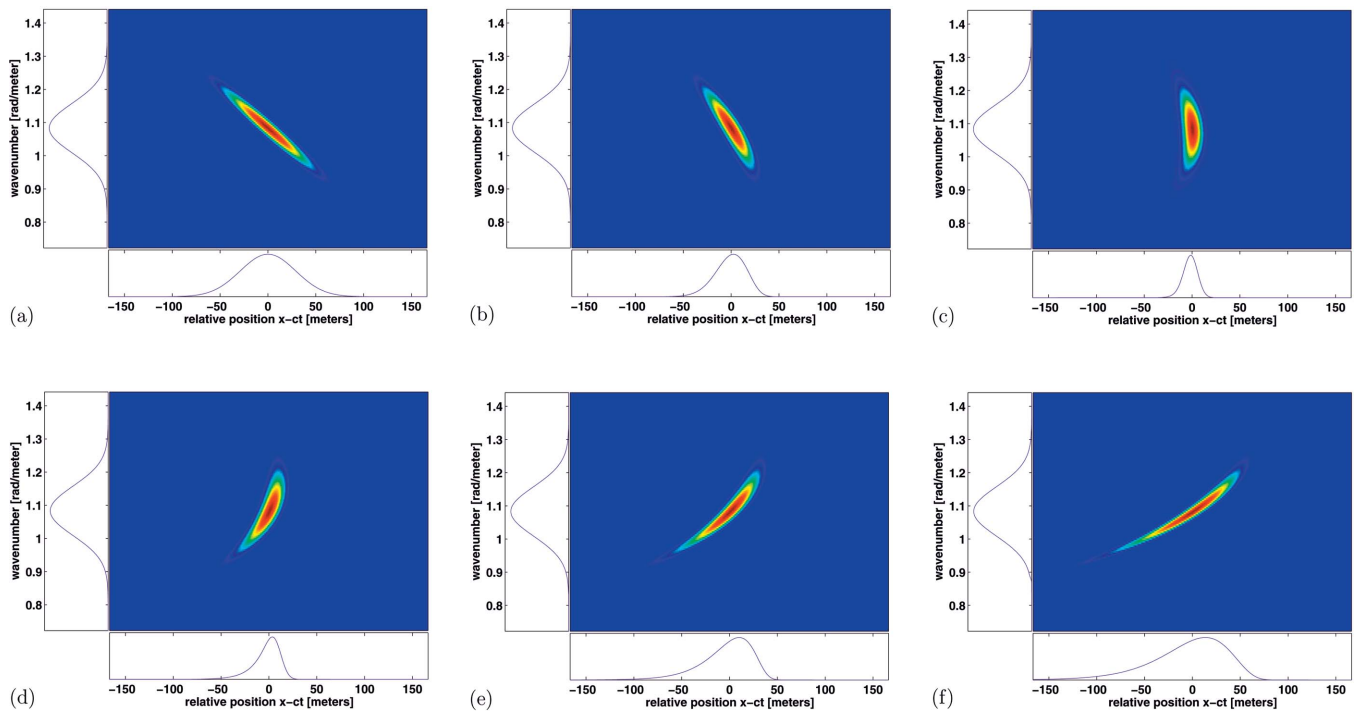
FIG. 1. (a) Exact position-wavenumber Wigner distribution (main panel) of an initial pulse ($t=0$), and then (b)–(f) an approximation for the first mode at five later times ($t=10, 20, 30, 40, 50$ s), simulating propagation of an underwater acoustic wave in a hard-bottom shallow water ocean channel of depth D $=25$ m and an isovelocity sound speed profile of $c=1450$ m/s. The lower panel in each plot shows the squared-magnitude of the pulse, which is exact for $t=0$, and approximate thereafter. The left panel shows the squared magnitude of the spatial spectrum, which is exact in all cases (the approximate Wigner distribution yields the exact magnitude-spatial spectrum). Note that the magnitude-spatial spectrum does not change as the pulse propagates, since there is no damping in this case (the dispersion relation is purely real). The parameters of the initial pulse are $k_0=1.083$ rad/m, $\alpha=1/1450$ m^{-2}, and $\beta= -0.0025$ rad/m^2, which corresponds to a pulse of nominal center frequency $f_0=250$ Hz and duration $\Delta t \approx 20$ ms. Different values for the waveguide and pulse parameters could be used to study other propagation cases.

effects of dispersion (and dissipation) on the wave in phase space. We believe that the primary advantage of the Wigner approximation, besides its simplicity, is in terms of the insights that it provides with regard to dispersive propagation, and, in particular, in showing how the wave evolves locally at each phase space point. We also mention that computational issues for higher dimensions for the standard methods become difficult but are simple for the Wigner distribution approximation.

In a future paper we discuss and compare this approximation in detail but we mention here that one reason the approximation is generally very good is that one does not have to assume that the initial spectrum is slowly varying, as, for example, is done in the stationary phase approximation. In addition, we are investigating higher-order approximations.

ACKNOWLEDGMENTS

[1] I. Tolstoy and C. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (American Institute of Physics, Woodbury, NY, 1987), Chaps. 2 and 4.

[2] K. Graff, *Wave Motion in Elastic Solids* (Oxford University Press, Oxford, 1975), Chap. 1, Sec. 1.6.

[3] J. Jackson, *Classical Electrodynamics* (Wiley, New York, 1992), Chap. 7, Secs. 7.7, 7.9, and 7.11.

[4] J. Lighthill, *Waves in Fluids* (Cambridge University Press, Cambridge, 1978), Chap. 3, Sec. 3.7.

[5] G. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974), Chap. 11, Sec. 11.2.

[6] L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

[7] G. Gaunaurd and H. Strifors, "Backscattered echoes from submerged elastic shells insonified by short realistic pulses using Wigner-type distributions," J. Acoust. Soc. Am. **92**, 2365 (1992).

[8] W. Williams, "Reduced interference distribution application and interpretation for transient acoustic events," J. Acoust. Soc. Am. **106**, 2184 (1999).

[9] L. Cohen, "Wigner distribution and pulse propagation," *45th Annual Meeting, International Society for Optical Engineering*, San Diego, July 29–August 3, 2001, Vol. 4474, pp. 20–24; "Local particle view of a propagating pulse," *15th AeroSense Symposium, International Society for Optical Engineering*, Orlando, April 18–April 20, 2001, Vol. 4391, pp. 23–34.

[10] P. Loughlin and L. Cohen, "Phase-space approach to wave propagation with dispersion and damping," *48th Annual Meeting, International Society for Optical Engineering*, 2–6 August 2004, Denver, Colorado, Vol. 5559, pp. 221–231.

[11] L. Cohen, "What is a multicomponent signal?" *IEEE Proc. ICASSP* **5**, 113–116 (1992).

[12] P. H. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 9.

[13] W. D. Mark, "Spectral analysis of the convolution and filtering of nonstationary stochastic processes," J. Sound Vib. **11**, 19–63 (1970).

[14] W. D. Mark, "Power spectrum representation for nonstationary random vibration," in *Random Vibration—Status and Recent Developments*, edited by D. H. Ielishakoff (Elsevier, Amsterdam, 1986).

Speed of sound in nitrogen as a function of temperature and pressure (L)

Axel Hagermann[a)] and John C. Zarnecki

Planetary and Space Sciences Research Institute, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

Speed of sound measurements in nitrogen by Younglove and McCarty [J. Chem. Thermodynam. **12**, 1121–1128 (1980)] are revisited and an empirical polynomial equation for the speed of sound is derived. The polynomial coefficients differ from those given by Wong and Wu [J. Acoust. Soc. Am. **102**, 650–651 (1997)] with the result that discrepancies between predicted and measured values at low temperatures are reduced. The maximal error over the complete temperature and pressure range from 80 to 350 K and 0.031 to 0.709 MPa is reduced from 5.38% to 0.78%. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000748]

I. INTRODUCTION

Recently, there has been an increase in interest in the physical properties of nitrogen gas at low temperatures in the context of planetary science. Wong and Wu reported a polynomial expression for the speed of sound in nitrogen as a function of temperature and pressure.[1] They used the values from 150 to 330 K as measured by Younglove and McCarty[2] and obtained a remarkable accuracy of 380 ppm in that region. At some low temperatures and high pressures, however, the deviation between measured and calculated values can be as high as 5.38%. In our search for a reasonably accurate function specification applicable to a wider range of temperatures and pressures, we have calculated a new polynomial fit to the complete set of Younglove and McCarty's measurements.

II. TWO-DIMENSIONAL POLYNOMIAL APPROXIMATION

We assume that the speed of sound c in nitrogen gas as a function of temperature T and pressure p can be approximated by a polynomial expression of the form

$$c(p,T) = \sum_{i=0}^{m} \left(\sum_{j=0}^{n} a_{ij} T^j \right) p^i, \qquad (1)$$

where the a_{ij} need to be found by matching the $c(p,T)$ function to experimental data. We assume throughout this paper that p is given in MPa and T in K.

Wong and Wu used a multicolumn coefficient curve matching procedure in the p, T domain, first defining coefficients A_i such that a least-squares fit is applied to the pressure-dependent measurements at temperatures T_k assuming

$$c(T_k) = \sum_{i=0}^{n} A_i p^i, \quad k = 0,1,2,\dots . \qquad (2)$$

Then the a_{ij} can be found using

$$A_i = \sum_{j=0}^{n} a_{ij} T^j. \qquad (3)$$

The function specification uses selected measurements by Younglove and McCarty at temperatures from 150 to 330 K, and is extremely accurate within this range, with a maximum deviation of 380 ppm. This is comparable with the experimental accuracy of 300 ppm. At lower and higher temperatures, however, this function specification is considerably less accurate, particularly with increasing pressure. We found deviations of more than 5% at 110 K and 1.44 MPa. An investigation revealed that the A_i column coefficients in the region from 150 to 330 K take values such that $|A_i| > |A_{i+1}|$, but this is not the case for very low and very high temperatures, explaining the higher error with increasing pressure.

We have used the full set of measurements by Younglove and McCarty. Temperatures were adjusted from the IPTS-68 to the ITS-90 system using an eighth-degree fit.[3] Given the magnitude of errors over the complete p,T range, however, we found this correction to be of minor importance. We then minimized the function of $(m+1) \times (n+1)$ variables

$$L = \sum_{\ell=1}^{237} (c_\ell - c(p_\ell, T_\ell, a_{11}, a_{12}, \dots a_{mn}))^2, \qquad (4)$$

where c_ℓ is the ℓth of the 237 measurements by Younglove and McCarty and $c(p_\ell, T_\ell, \dots)$ is the speed of sound as calculated by Eq. (1). L being a quadratic form, $\partial L / \partial a_{ij}$ can be readily derived, and the solution can be found using conjugate gradients or variants of Newton's method. We used the latter, with Wong and Wu's values as a starting guess.

―――――――――
a)Electronic mail: a.hagermann@open.ac.uk

TABLE I. Polynomial coefficients for our two-dimensional function specification for the speed of sound in nitrogen.

j	a_{1j}	a_{2j}	a_{3j}	a_{4j}
0	$6.721\,128E+01$	$-1.908\,177E+02$	$5.908\,631E+01$	$-6.511\,627E+01$
1	$1.820\,085E+00$	$2.971\,323E+00$	$-1.063\,665E+00$	$1.218\,812E+00$
2	$-5.713\,575E-03$	$-1.699\,747E-02$	$6.173\,142E-03$	$-7.949\,396E-03$
3	$1.320\,687E-05$	$4.219\,848E-05$	$-1.390\,550E-05$	$2.187\,479E-05$
4	$-1.265\,187E-08$	$-3.806\,689E-08$	$9.852\,826E-09$	$-2.165\,669E-08$

III. RESULTS

The polynomial coefficients a_{ij} for Eq. (1) with $n=4$, $m=3$ are given in Table I. The maximal deviation from Younglove and McCarty's measurements has been reduced to 0.78%. As the degree and order of our fit are only 4 and 3, respectively, some trade-offs had to be made in the midtemperature region from 150 to 330 K. Here the maximal deviation is 0.52%. In moderate temperature regimes, Wong and Wu's approximation remains more accurate. However, if we compare our fit to the measurements made by Ewing and Trusler,[4] the maximum deviation is reduced to only 0.22% in the region from 80 to 300 K. This value increases to 0.88% if we include their measurements at 373 K. In extreme conditions like those encountered in planetary science, our function specification maintains its accuracy over the whole range covered by measurements.

[1]G. S. K. Wong and L. Wu, "Variation of measured nitrogen sound speed with temperature and pressure," J. Acoust. Soc. Am. **102**, 650–651 (1997).
[2]B. A. Younglove and R. D. McCarty, "Speed-of-sound measurements for nitrogen gas at temperatures from 80 to 350 K and pressures to 1.5 MPa," J. Chem. Thermodyn. **12**, 1121–1128 (1980).
[3]R. L. Rusby, "The conversion of thermal reference values to the ITS-90," J. Chem. Thermodyn. **23**, 1153–1161 (1991).
[4]M. B. Ewing and J. P. M. Trusler, "Second acoustic virial coefficients of N2," Physica A **184**, 415–436 (1992).

Release from informational masking by time reversal of native and non-native interfering speech (L)

Koenraad S. Rhebergen[a]
Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-223, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Niek J. Versfeld[b]
Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-330, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

Wouter. A. Dreschler[c]
Department of Clinical and Experimental Audiology, Academic Medical Center, Room D2-240, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

In many studies, the influence of intelligibility of the interfering speech is avoided by reversing it in time. Usually, intelligibility with time-reversed interfering speech indeed is higher compared to that with normal interfering speech. However, due to the nature of speech, reversed speech also gives rise to increased forward masking. The latter will result in a decrease in intelligibility. Thus, differences in intelligibility as a consequence of reversing speech in time are due to two opposite effects. This paper describes a speech reception threshold (SRT) test with intelligible and unintelligible interfering speech played normally and time-reversed. With Dutch listeners, Swedish reversed interfering speech gave a rise in SRT of 2.3 dB compared with the Swedish interfering speech (played normally). The difference can be attributed to differences in forward masking. Dutch time-reversed interfering speech gave a decrease in SRT of 4.3 dB compared to (intelligible) Dutch interfering speech. The latter is the result of both a release from informational masking and an increase in forward masking. Therefore, the amount of informational masking is larger than 4.3 dB and, if one assumes similar differences in forward masking for Dutch and Swedish speech, may amount to 6.6 dB. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000751]

I. INTRODUCTION

There are various papers describing speech intelligibility in the presence of one or more interfering talkers (e.g., Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Bronkhorst, 2000; Drullman and Bronkhorst, 2000; Brungart, 2001; Brungart *et al.*, 2001; Brungart and Simpson, 2002; Summers and Molis, 2004). Remarkably, intelligibility appears to vary greatly between conditions, which in part might be due to the degree of similarity between the target speaker (here: "target" or "signal") and the interfering speakers (here: "masker" or "noise"). The more similar the target and masker are, the more the listener is confused or distracted, which in turn results in poorer performance. Differences between target and masker with respect to gender (male or female speech) or intelligibility of the interfering speech (native or foreign language) have only a small effect with respect to actual energetic masking of the target, but certainly can have a large effect on intelligibility. The phenomenon of excess masking is often labeled as informational masking (Drullman and Bronkhorst, 2000; Bronkhorst, 2000;

Brungart, 2001; Brungart *et al.*, 2001). This paper concentrates on the part of informational masking due to intelligibility of the interfering speech.

In order to study the speech intelligibility in speech-like maskers, and at the same time avoid the informational masking component due to the intelligibility of the interfering speech, many researchers use fluctuating speech-shaped noise (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992; Peters *et al.*, 1998; Versfeld and Dreschler, 2002) or play the interfering speech signal backwards (Festen and Plomp, 1990; Summers and Molis, 2004). Fluctuating speech-shaped noise is made by modulating the long term speech spectrum with the speech envelope of the interfering talker. The envelope can be extracted from broad band speech (one band), two bands (Festen and Plomp, 1990), three bands (ICRA noise; Dreschler *et al.*, 2001), or even more. The fluctuating speech-shaped noise has more or less the same intensity fluctuations in time as real speech, has the same long-term spectrum of speech, but is generally unintelligible because the fine structure is lost.

By using time-reversed speech as a masker, the spectral contents of speech are in essence untouched. The reversed speech is not intelligible and therefore this component of the informational masking by definition is removed. However, the envelope of time-reversed speech is reversed as well. The shape of temporal envelope of speech is typically dominated

---
[a] Electronic mail: k.s.rhebergen@amc.uva.nl
[b] Electronic mail: n.j.versfeld@amc.uva.nl
[c] Electronic mail: w.a.dreschler@amc.uva.nl

by plosive sounds (Rosen, 1992), and the envelope of these sounds is characterized by quick onsets (steep slope) and slow decays (shallow slope). Reversal of the speech signal thus results in temporal envelopes that have abrupt offsets. Since the auditory system does not follow the offset instantaneously, but rather displays a decay of the envelope across time, abrupt changes cannot be followed accurately, hence soft signals can be easily masked by a preceding strong signal (called forward masking).

Irino and Patterson (1996), Carlyon (1996), Stecker and Hafter (2000), and Schlauch *et al.* (2001) examined the perception of stimuli with ramped envelopes (gradual attack and abrupt decay) and damped envelopes (abrupt attack and gradual decay). The sum of these studies reveal that ramped signals are subjectively judged longer in duration and are perceived louder compared to damped signals. This is explained by the abrupt offset at a high level of a ramped sound which results in a persistence of perception, and more forward masking compared to a damped sound. The same effects could be expected by time reversed speech since its envelope is more ramped compared to the damped normally played speech (Rosen, 1992).

Forward masking has a clear effect on speech intelligibility (Festen, 1987). The recovery time from forward masking is in the range of 100–200 ms (Moore, 2003). For completeness, a second type of masking, present in human auditory perception, should be mentioned, viz., backward masking. In this case, a soft signal is masked by a louder signal that follows it. The phenomenon of backward masking is still poorly understood (Moore, 2003). The amount of backward masking obtained with practiced subjects often is little or none (Moore, 2003). Its effect on speech intelligibility is still unclear and probably not very large.

By reversal of the speech masker in time, one expects on the one hand an improvement in intelligibility of the target speech due to the fact that the masking speech becomes unintelligible (partly release from informational masking), and, on the other hand a decrease in performance due to the fact that the temporal envelope of the reversed speech causes an increase of forward masking. Festen and Plomp (1990) and Summers and Molis (2004) found a slight improvement in performance when using reversed speech instead of normal speech as a masker. This improvement is the result of both effects combined. From these experiments it is not possible to assess the individual contribution of both components.

The goal of the current experiment is to separate informational masking due to intelligibility of the interfering speech from forward masking. Additional masking due to time reversal of the speech is assessed by measuring the speech intelligibility in masking speech of a foreign language. Since the foreign speech is not intelligible (neither forward nor backward), differences in intelligibility due to time reversal must be due to differences in forward masking alone. The same experiment is repeated in native interfering speech. Differences in intelligibility due to time reversal then are the combined effect of an increase in forward masking and a decrease in informational masking. Since the speech materials are similar, and since the single effect of additional masking due to time reversal is known, the effect of informational masking due to intelligibility of the interfering speech can be estimated.

II. METHOD

A. Subjects

Eight normal-hearing subjects (3 male, 5 female) participated. Their mean age was 25 years and ranged from 21 to 39 years. Subjects were native speakers of the Dutch language. Subjects had at least high school education. Each subject had pure-tone thresholds of 15 dB HL or better at octave frequencies from 125 to 8000 Hz (ANSI S3.6, 1996).

B. Stimuli

The target speech material consisted of short every-day sentences, uttered by a male speaker (Versfeld *et al.*, 2000). The speech material comprised 39 lists of 13 sentences and was developed for a reliable measurement of speech intelligibility in noise. The speech was stored with a sampling rate of 44.1 kHz and 16 bits resolution.

Swedish and Dutch speech was used as interfering speech. The Swedish speech was developed by Hagerman (1982) and consisted of short sentences read by a female speaker. For the Dutch language, a corpus similar to that of Hagerman (1982) was developed. The Dutch sentences were uttered by a different female speaker. The Dutch speech was rescaled to obtain the same speaking rate as that of Hagerman's (1982) set, viz., 3.5 s per sentence. Scaling was done by use of the PSOLA method [Pitch-Synchronous Overlap Add, a method for manipulating the pitch and duration of an acoustic speech signal, Moulines and Charpentier, 1990]. Pseudorunning speech was obtained by concatenation of sentences in succession without pauses.

C. Procedure

Subjects were tested individually in a sound-insulated booth. The monaural speech-reception threshold was measured at the better ear for a fixed noise level of 65 dBA. Signals were played out via an Echo soundcard (Gina 24/96) on a PC at a sample frequency of 44.1 kHz and were fed through a TDT Microphone Amplifier (MA2) and a TDT Headphone Buffer (PA4) via THD 39P headphones. After the presentation of a sentence, the subject's task was to repeat the sentence he or she had just been presented. A sentence was scored as correct if all words in that sentence were repeated without any error. A list of 13 sentences, unknown to the subject, was used to estimate the level at which 50% of the sentences was reproduced without any error, the so-called speech reception threshold, or SRT. For a given condition, the first sentence of the list started far below the expected SRT. The sentence was repeated each time at a 4 dB higher level until the subject was able to reproduce it correctly. The 12 remaining sentences in that list were presented only once, following a simple up-down procedure with a step size of 2 dB. The SRT was estimated according to the procedure described by Plomp and Mimpen (1979), i.e., by taking the mean signal to noise ratio (SNR) of sentence 5 to 13 plus the
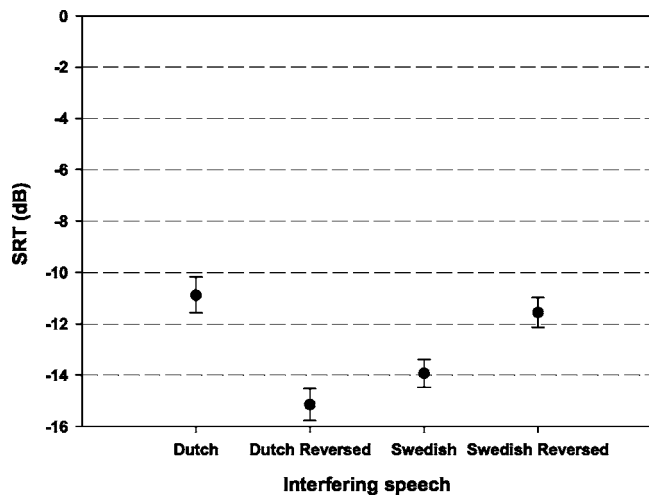
FIG. 1. Speech reception threshold (dB) as function of interfering speech type. Error bars denote the standard error of the mean.

estimated SNR that would have been used for the 14th sentence. With each sentence presentation, a random sample of the interfering speech was taken. It started 1200 ms before the start of the sentence and stopped at least 800 ms after the sentence.

In total four masking conditions were tested: Dutch masker forward, Dutch masker reversed, Swedish masker forward, and Swedish masker reversed. The experiment was partitioned into two blocks, a test and a retest. To avoid confounding of measurement condition order and sentence lists, the order of conditions and sentence lists was counterbalanced across subjects according to an 8 by 8 Latin Square method. In total, each subject received 8 lists of 13 sentences preceded by 3 practice lists.

III. RESULTS

Figure 1 shows the SRT values averaged across subjects and test-retest for each of the four conditions of the interfering speech. Error bars denote the standard error of the mean. A 4[condition]$\times$2[test/retest]$\times$8[subject] analysis of variance (ANOVA) was performed on the data set. Of the main effects, only differences in "condition" were significant ($F[3,21]=24.6, p<0.001$). The SRT of the retest was on average 1.2 dB better than the test, but this difference was not significant ($F[1,7]=4.6, p>0.05$). Also, differences between subjects were just not significant ($F[7,4]=5.6, p>0.05$). None of the interactions were significant.

The difference in SRT between the two conditions with Swedish speech was significant (Tukey's HSD test, $z=2.4$, $p=0.02$) and was on average 2.3 dB. Thus, by time reversing speech, the SRT increases (i.e., performance worsens) from -13.9 to -11.6 dB. The difference between the two conditions with Dutch interfering speech was also significant (Tukey's HSD test, $z=4.28$, $p<0.001$). Here, time reversal of the masking speech causes the SRT to improve by 4.3 dB. The latter effect was apparent in all subjects, and the result is in agreement with that of Festen and Plomp (1990). Last, the

difference between the two conditions with reversed masking speech was also significant (Tukey's HSD test, $z=3.6$, $p<0.001$).

IV. DISCUSSION

The most important finding of the present experiment is that time reversal of foreign, nonintelligible speech that is used as a masker, affects intelligibility and causes the SRT to increase by more than 2 dB. Most likely, reversal of the envelope of the speech signal results in an increased contribution of forward masking. The main conclusion to be drawn from this experiment is that reversal of a speech masker indeed introduces additional masking on top of energetic masking in the form of forward masking.

Time reversal of intelligible (Dutch) masking speech results in a decrease in the SRT of 4.3 dB. This difference is the result of the elimination of intelligibility of the masker on the one hand (enhancing speech intelligibility), and the introduction of additional forward masking on the other hand (decreasing speech intelligibility). If one assumes that the additional amount of forward masking is similar for the Swedish and Dutch masking speech, then in the present experiment the effect of informational masking due to intelligibility of the speech masker can be estimated to be 4.3 +2.3=6.6 dB.

If intelligibility of the speech masker were the sole factor determining the amount of informational masking, one would expect the SRT with Swedish reversed speech and Dutch reversed speech to be similar. In the present experiment this is not true; the difference is about 3.6 dB. Apparently, the Dutch reversed speech is a poorer masker (SRT $=-15.2$ dB) than the Swedish reversed speech (SRT $=-11.6$ dB). If Dutch interfering speech played in reverse sounds more similar to Dutch speech than Swedisch speech played either forward or in reverse, one would expect the opposite. It is interesting to note that subjects told they could not distinguish between the Swedish reversed speech and the Dutch reversed speech. Also, they misjudged the reversed Dutch speech for Swedish speech and vice versa. Therefore, the effect probably is not very large. As mentioned in Sec. I, other factors than intelligibility contribute to informational masking. From the present results, it is difficult to say which of the many potential factors contributed most; perhaps differences in intonation, rhythm, mean pitch, modulation spectrum, or differences in the long-term speech spectrum. If the interfering native and non-native speech were uttered by the same person, with the same pronounced intonation, rhythm, and pitch, one would expect the SRTs in the reversed condition to be more alike. But even then, differences might still exist due to specific characteristics of the language.

The present experiment is a first attempt to separate energetic masking from informational masking. Future research with different native and non-native interfering speech, uttered by the same person may give more insight into the effects of forward masking and informational masking on speech intelligibility.

V. CONCLUSIONS

By using speech from different languages, played both forward or time-reversed, it is possible to disentangle the effects of informational masking caused by intelligibility of the interfering speech and energetic masking. However, time reversal of speech results in an increase in temporal (forward) masking. In the present experiment, this effect is about 2.3 dB. The release from informational masking by making intelligible speech unintelligible by reversing it in time is obscured by an increase in forward masking. In the present experiment, it is shown that informational masking might be as large as 6.6 dB if one assumes the same amount of additional forward masking with Dutch and Swedish reversed speech.

ACKNOWLEDGMENTS

ANSI (**1996**). "ANSI S3.6-1996, "*American national standard methods for specification for audiometers*" (American National Standards Institute, New York).

Bronkhorst, A. W. (**2000**). "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," Acust. Acta Acust. 86, 117–128.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2002**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Carlyon, R. P. (**1996**). "Spread of excitation produced by maskers with damped and ramped envelopes," J. Acoust. Soc. Am. **99**, 3647–3655.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing aid assessment," Audiology **40**, 148–157.

Drullman, R., and Bronkhorst, A. W. (**2000**). "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235.

Festen, J. M. (**1987**). "Speech-perception threshold in a fluctuating background sound and its possible relation to temporal resolution," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht), pp. 461–466.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.

Irino, T., and Patterson, R. D. (**1996**). "Temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **99**, 2316–2331.

Moore, B. C. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, London).

Moulines, E., and Charpentier, F. (**1990**). "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453–467.

Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.

Plomp, R., and Mimpen, A. M. (**1979**). "Improving the reliability of testing the speech reception threshold for sentences," Audiology **18**, 43–52.

Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.

Schlauch, R. S., Ries, D. T., and DiGiovanni, J. J. (**2001**). "Duration discrimination and subjective duration for ramped and damped sounds," J. Acoust. Soc. Am. **109**, 2880–2887.

Stecker, G. C., and Hafter, E. R. (**2000**). "An effect of temporal asymmetry on loudness," J. Acoust. Soc. Am. **107**, 3358–3368.

Summers, V., and Molis, M. R. (**2004**). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," J. Speech Lang. Hear. Res. **47**, 245–256.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). "Method for the selection of sentence materials for efficient measurement of the speech reception threshold," J. Acoust. Soc. Am. **107**, 1671–1684.

Versfeld, N. J., and Dreschler, W. A. (**2002**). "The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners," J. Acoust. Soc. Am. **111**, 401–408.

Symmetry and Snell's law

David R. Bergman[a)]

Department of Physics, Saint Peter's College, Jersey City, New Jersey, 07306

A connection between the acoustic rays in a moving fluid medium (with a sound speed and subsonic flow each depending arbitrarily on position and time) and the null geodesics of a pseudo-Riemannian manifold allows one to derive several well-known results used in underwater and atmospheric acoustic ray theory from one single underlying symmetry principle, isometry. These results being well known in cases with Cartesian symmetry generalize to any system of curvilinear coordinates and three-dimensional fluid flows. Snell's law, range, and travel time integrals for Cartesian and cylindrically symmetry systems are derived and an illustrative example presented. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1979427]

I. INTRODUCTION

A well-known result from the study of geometric acoustics in a stationary medium states that when the sound speed depends on only one Cartesian variable, the equations for the acoustic ray may be integrated to yield a conserved quantity along the ray path. In this case the conservation of $\cos \theta / c$, known as Snell's law, can be used to predict ray paths. Over the years generalizations of this law for special circumstances including a depth-dependent moving fluid medium have appeared in the literature. In most cases the existence of cyclic variables is used to derive the result directly from the second-order ray equations. A complete derivation of a generalized Snell's law for rays in a layered moving fluid can be found in Ref. 1. Treating the acoustic rays as particle trajectories invokes an immediate connection with Noether's theorem, in which the conserved quantities along the particle trajectory are derived from the Lagrangian, or Hamiltonian, rather than from the equations of motion.

Through the purely geometric treatment of acoustic rays as geodesics of a metric space, these conservation laws emerge in a natural way by identifying isometries of an effective acoustic metric. In short an isometry is a rigid motion of a space (a mapping of the space into itself that preserves distances). Once the isometries of a manifold are found, a conserved current for each can be constructed from the infinitesimal generator of the isometry, called a Killing vector. The technique is used extensively in general relativity to describe particle motion in terms of energy conservation in an effective potential rather than by the geodesic equations. The technique is used here to derive Snell's law in a moving fluid for cases with translational and cylindrical symmetry. For a more detailed account of the use of isometries and Killing's theorem the reader is referred to Refs. 2 and 3.

By the same technique the standard ray integrals used in oceanic and atmospheric acoustics emerge. The final results in Cartesian coordinates are well known.[4–7] A salient feature of the approach presented here is that the derivations of several results are encapsulated in one common procedure,

which is applicable in any general curvilinear coordinate system. This brief paper presents the derivation of ray-path integrals in Cartesian and cylindrical coordinates via the procedure of identifying isometries of the acoustic metric along with one example with an exact solution, illustrating the connection between ray theory, differential geometry and Noether's theorem.

II. THE SPACE-TIME ACOUSTIC METRIC AND CANONICAL MOMENTUM

The acoustic rays in a fluid environment with arbitrary sound speed and subsonic current flow are identical to the null geodesics of a pseudo-Riemannian manifold. This result, which follows directly by application of the method of characteristics,[8–10] implies that the ray paths are curves of zero length in space-time:

$$ds^2 = -(c^2 - v^2)dt^2 - 2\vec{v} \cdot d\vec{x}dt + d\vec{x} \cdot d\vec{x} = 0, \qquad (1)$$

in which v and c depend on position and time with $\{x\}$ the standard Cartesian coordinates. Equation (1) defines a metric tensor, referred to in the literature as the acoustic metric, with components $\{g_{00} = v^2 - c^2, g_{0i} = g_{i0} = -v_i, g_{ij} = \delta_{ij}\}$. We define a Lagrangian via $\dot{x}^\mu \dot{x}^\nu g_{\mu\nu} = 2L$ and the Hamiltonian $p_\mu p_\nu g^{\mu\nu} = 2H$, in which the canonical momentum, $p_\mu = \partial L / \partial \dot{x}^\mu = \dot{x}^\nu g_{\mu\nu}$, is defined.[11] From the definition of L and Eq. (1) the translational momentum and energy along the ray are $\vec{p} = c\hat{n}t$ and $p_0 = -c\hat{n} \cdot \dot{\vec{x}}$, respectively,[12] where $\hat{n}$, the unit normal of the wave-front in space, is related to the ray path by $d\vec{x}/dt = c\hat{n} + \vec{v}$. The momentum is proportional to the wave-front gradient and the energy is proportional to the projection of the wave-front gradient onto the ray path velocity. Moving to general curvilinear coordinates Eq. (1) reads

$$ds^2 = -(C^2 - V^2)dt^2 - 2(V_i)dq^idt + G_{ij}dq^idq^j = 0, \qquad (2)$$

in which $C \equiv c(\vec{q}, t)$, $V_i \equiv \vec{v} \cdot \partial \vec{x} / \partial q^i$, and $G_{ij} \equiv (\partial \vec{x}/\partial q^i) \cdot (\partial \vec{x}/\partial q^j)$, all as functions of $\{q\}$.[13] The $\{V_i\}$ are the components of $\vec{v}$ projected onto the $\{q\}$ coordinate curves.

In this paradigm the acoustic rays are null geodesics in a four-dimensional space-time manifold. The geometry of the

---

[a)]Electronic mail: dbergman@spc.edu
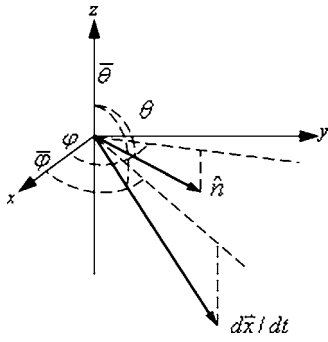
FIG. 1. Ray tangent and wavefront normal in three dimensions. Polar and azimuth angles for each are defined.

FIG. 2. Relationship between the wavefront, fluid velocity, and ray tangent in Cartesian space for two-dimensional planar rays.

manifold is determined by the environmental parameters, $c(\vec{x},t)$ and $\vec{v}(\vec{x},t)$, and described by the acoustic metric appearing in Eqs. (1) and (2). An isometry of a manifold is a symmetry indicated by the presence of a cyclic variable in the metric tensor. If $g_{\mu\nu}(x^\alpha)$ does not depend on one of the coordinates x^S, then any curve may be translated along the x^S coordinate curves with no change. From the point of view of dynamical systems theory with L defined above this cyclic coordinate leads to a quantity $p_S = g_{S\nu}\dot{x}^\nu$ which is conserved along the geodesic (ray path).

Specializing to a time-independent environment leads immediately to conservation of energy,[14] $\kappa_0 = c\hat{n}\cdot\dot{\vec{x}}$, where a minus sign has been absorbed into the definition of κ_0. If the parameters of the medium depend on only one general coordinate, say q_1, then there are two other conserved quantities corresponding to the components of translational momentum P_2 and P_3, referred to collectively here as the transverse momentum $\vec{P}_T$, $i C\hat{N}_T = \kappa$ Snell's law in vector form may be stated simply as conservation of the transverse momentum to energy ratio,

$$\frac{\vec{p}_T}{p_0} = \frac{\hat{N}_T}{C + \vec{V}\cdot\hat{N}} = \vec{\alpha}, \tag{3}$$

in which $\vec{\alpha} = \vec{\kappa}/\kappa_0$. For example, in Cartesian coordinates the assumption of a depth-dependent medium, $c(z)$, $v(z)$, leads to the transverse momentum $\vec{p}_T$, $ic\hat{n}_T = \vec{\kappa}$, and $\hat{n}_T/(c+\vec{v}\cdot\hat{n})$ =const.

III. RAY PATH GEOMETRY

Here we work exclusively in Cartesian coordinates. Equation (3) may be solved to give the direction of the wave-front normal at all points along the ray. Figure 1 shows $\hat{n}$ and $d\vec{x}/dt$ in a local Cartesian frame with polar and azimuth angles defined. In spherical coordinates $\hat{n}$ =$(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$. The ratio of n_y/n_x and the magnitude of $\vec{\kappa}$ give

$$\tan\varphi = \frac{\kappa_y}{\kappa_x}, \tag{4a}$$

$$\frac{\sin\theta}{c + \sigma\sin\theta + v_z\cos\theta} = \kappa, \tag{4b}$$

respectively, in which $\sigma = v_x\cos\varphi + v_y\sin\varphi$ and $\kappa = \sqrt{\kappa_x^2 + \kappa_y^2}$. Equation (4a) states that the wave-front propagates through the medium with a constant azimuth, as expected. Having solved Eq. (4a), Eq. (4b) is then solved for $\cos\theta$,

$$\cos\theta = \frac{-cv_z \pm (\lambda^{-1} - \sigma)\sqrt{(\lambda^{-1} - \sigma)^2 + (v_z^2 - c^2)}}{v_z^2 + (\lambda^{-1} - \sigma)^2}. \tag{5}$$

Equations (4a) and (5) determine $\hat{n}$. From this information the direction of the ray tangent is found from the relation $d\vec{x}/dt = c\hat{n} + \vec{v}$,

$$\tan\bar{\varphi} = \frac{c\sin\theta\sin\varphi + v_y}{c\sin\theta\cos\varphi + v_x}, \tag{6}$$

$$\tan\bar{\theta} = \frac{\|c\hat{n}_T + \vec{v}_T\|}{c\cos\theta + v_z}. \tag{7}$$

In underwater and atmospheric acoustics it is common to assume $v_z = 0$, $v_y = 0$ and treat only rays initially fired in the x-z plane. Making these assumptions, the angle of the wave-front normal and the tangent vector are related by[15] $c\cos\theta$ =$\|d\vec{x}/dt\|\cos\psi - v$; see Fig. 2, which in turn implies $\|d\vec{x}/dt\| = v + \sqrt{c^2 - v^2(\sin\psi)^2}$. Inserting this result into Eq. (3) leads immediately to the result

$$\frac{1}{c}\frac{\varepsilon\mu^2\cos^2\psi - \varepsilon\mu^2 + \mu\cos\psi\sqrt{1 - \varepsilon^2\mu^2\sin^2\psi}}{1 - \varepsilon^2\mu^2 + \varepsilon^2\mu^2\cos^2\psi + \varepsilon\mu\cos\psi\sqrt{1 - \varepsilon^2\mu^2\sin^2\psi}} = \alpha, \tag{8}$$

in which $\varepsilon = v/c$ is defined as the local Mach number and $\mu = c_0/c$ is the index of refraction (Kornhauser's notation is used). This result can be compared to that of Kornhauser, where use of the trigonometric identity $\cos^2\psi + \sin^2\psi = 1$ shows they are identical.

IV. RANGE AND TRAVEL-TIME INTEGRALS

A. Cartesian symmetry

Applying the formalism of Sec. II leads to the following first-order equations for the ray coordinates and travel time:

$$(c^2 - v^2)\dot{t} + \vec{v}\cdot\dot{\vec{x}}_T = \kappa_0, \tag{9}$$

$$\dot{\vec{x}}_T - \vec{v}\dot{t} = \vec{\kappa}, \tag{10}$$

where it has been assumed that $v_z=0$, $\vec{v}=\vec{v}_T$. Equations (9) and (10) combine to give

$$\dot{t} = c^{-2}(\kappa_0 - \vec{v}\cdot\vec{\kappa}), \tag{11}$$

$$\dot{\vec{x}}_T = \vec{\kappa} + c^{-2}\vec{v}(\kappa_0 - \vec{v}\cdot\vec{\kappa}). \tag{12}$$

Inserting Eqs. (11) and (12) into the null constraint leads to an equation for z,

$$\dot{z} = \pm c^{-1}\sqrt{(\kappa_0 - \vec{v}\cdot\vec{\kappa})^2 - c^2\vec{\kappa}\cdot\vec{\kappa}}. \tag{13}$$

The ratio of Eqs. (11) and (13) gives dt/dz while that of Eqs. (12) and (13) gives $d\vec{x}_T/dz$, which determine the travel-time and range coordinates of a ray as a function of depth,

$$t(z) = \int_{z_0}^{z} \frac{1 - \vec{v}\cdot\vec{\alpha}}{c\sqrt{(1-\vec{v}\cdot\vec{\alpha})^2 - c^2\alpha^2}}\, dz, \tag{14}$$

$$x_T(z) = \int_{z_0}^{z} \frac{c^2\vec{\alpha} + (1-\vec{v}\cdot\vec{\alpha})\vec{v}}{c\sqrt{(1-\vec{v}\cdot\vec{\alpha})^2 - c^2\alpha^2}}\, dz, \tag{15}$$

in which $\vec{\alpha}=\vec{\kappa}/\kappa_0$ and a factor of κ_0 has been absorbed into the definition of the ray parameter λ. When $\vec{v}=0$ Eqs. (14) and (15) reduce to the standard integrals found in most texts on oceanography; see, for example, Ref. 16.

B. Environments with cylindrical symmetry

In this section the technique is applied to systems with cylindrical symmetry, which find application in atmospheric acoustics. Let the environment be a rotating fluid with $c(r)$ and $\vec{v}=v_\varphi(r)\hat{\varphi}+v_z(r)\hat{k}$, $v^2=v_\varphi^2+v_z^2$. Expressing the Lagrangian in cylindrical coordinates

$$-(c^2-v^2)\dot{t}^2 + \dot{r}^2 + r^2\dot{\varphi}^2 + \dot{z}^2 - 2v_\varphi r\dot{\varphi}\dot{t} - 2v_z\dot{z}\dot{t} = 0, \tag{16}$$

there are three cyclic coordinates, t, z, and φ which lead to conservation of energy, linear momentum in the z direction, and angular momentum about the z axis[17]:

$$(c^2-v^2)\dot{t} + v_\varphi r\dot{\varphi} + v_z\dot{z} = 1, \tag{17}$$

$$\dot{z} - v_z\dot{t} = k, \tag{18}$$

$$r^2\dot{\varphi} - v_\varphi r\dot{t} = \alpha. \tag{19}$$

Combining these leads to

$$\dot{t} = c^{-2}(1 - r^{-1}v_\varphi\alpha - v_z k), \tag{20}$$

$$\dot{\varphi} = r^{-2}\alpha + r^{-1}c^{-2}v_\varphi(1 - r^{-1}v_\varphi\alpha - v_z k), \tag{21}$$

$$\dot{r} = \pm c^{-1}\sqrt{(1 - r^{-1}v_\varphi\alpha - v_z k)^2 - c^2(r^{-2}\alpha^2 + k^2)}. \tag{22}$$

As before these lead to the ray travel time and range as functions of r:

$$t(r) = \int_{r_0}^{r} \frac{(r - v_\varphi\alpha - rv_z k)}{c\sqrt{(r - v_\varphi\alpha - rv_z k)^2 - c^2(r^2k^2 + \alpha^2)}}\, dr, \tag{23}$$

$$\varphi(r) = \int_{r_0}^{r} \frac{(c^2\alpha + v_\varphi(r - v_\varphi\alpha - rv_z k))}{cr\sqrt{(r - v_\varphi\alpha - rv_z k)^2 - c^2(r^2k^2 + \alpha^2)}}\, dr. \tag{24}$$

The cylindrical version of Snell's law is similar in form to Eq. (3) with the exception that the rotational component is scaled by a factor of r. For two-dimensional problems in the (r,φ) plane the local horizontal is lined up with curves of constant r, θ of Sec. III (see Ref. 4) becomes the angle between $\hat{n}$ and $\hat{\varphi}$, and Snell's law becomes $\alpha=r_0\cos\theta_0/(c_0 + v_0\cos\theta_0)$.

V. EXAMPLE

As a simple example consider rays in a vortex flow described by $c=ar$ and $v_\varphi=br$, $v_z=0$ with the condition $a>b>0$ (i.e., the flow is always subsonic). The ray path is

$$t = \frac{1}{a\sqrt{1-p^2}}\ln\left(\frac{r}{r_0}\frac{\sqrt{1-p^2} + \sqrt{1-p^2-(a+bp)^2k^2r_0^2}}{\sqrt{1-p^2} + \sqrt{1-p^2-(a+bp)^2k^2r^2}}\right), \tag{25}$$

$$\varphi = (b + ap)t, \tag{26}$$

$$\lambda = -\left.\frac{1}{k^2(a+bp)}\sqrt{1-p^2-(a+bp)^2k^2r^2}\right|_{r_0}^{r}, \tag{27}$$

$$z = z_0 + k\lambda, \tag{28}$$

in which $p=\cos\theta_0$ and $\varphi_0=t_0=0$. For $k=0$ Eqs. (26) and (27) become

$$\varphi = \frac{(b+ap)}{a\sqrt{1-p^2}}\ln\left(\frac{r}{r_0}\right), \tag{29}$$

$$\lambda = \frac{1}{2}\frac{(a+bp)}{\sqrt{1-p^2}}(r^2 - r_0^2), \tag{30}$$

respectively. Rays fired in the (r,φ) plane will fly off to infinity while those fired with nonzero value of k have a radial turning point at $r_{tp}=k^{-1}(a+bp)^{-1}\sqrt{1-p^2}$, as indicated by Eq. (22), eventually spiraling in towards the z axis. Rays with $r=0$ define helices about the z axis. A careful analysis of Eq. (21) for this case indicates that φ does not change direction and rays fired against the wind with $\cos\theta_0=-\varepsilon$ travel in the radial direction, $\varphi=0$. Equations (26), (29), and (30) give a time-parametrized ray allowing wave-fronts to be easily traced with little effort. Solving Eq. (28) for $r(t)$ gives

$$r = \frac{2AB}{D^2e^{\omega t} + A^2e^{-\omega t}}, \tag{31}$$

in which $A=r_0^{-1}(B+\sqrt{B^2-D^2r_0^2})$, $B=\sqrt{1-p^2}$, $D=(a+bp)k$, and $\omega=aB$. For rays in the horizontal plane, $k=0$, a glance at Eq. (22) shows that $r=0$ when $\hat{n}=\pm\hat{\varphi}$ or $\theta=0,\pi$, indicating that rays fired in the $\pm\hat{\varphi}$ direction define circles $r=r_0$ about the center of the vortex. Sample ray traces illustrated in Figs. 3–5 show the deformation of the wave-fronts for rays in the horizontal plane, $k=0$, while Fig. 6 shows a sample three-dimensional ray trace for $k=0.1$.
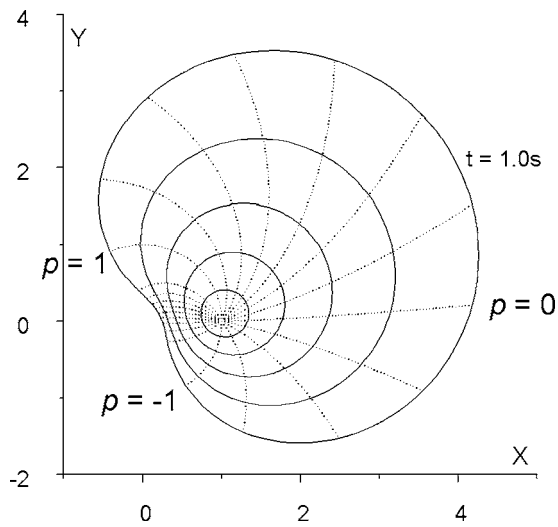
FIG. 3. Rays and wavefronts for a point source placed at (1.0 m, 0.0 m). Fluid flow is clockwise about the origin with $a=1.5$ and $b=0.5$. Wavefronts are traced out to $t=1.0$ s in increments of 0.2 s. A fan of 21 rays is shown for equal increments of $p=\cos\theta_0$. The rays marked $p=1$ and $p=-1$ define a circle about the origin.

FIG. 5. (a) Wavefronts for the same scenario as depicted in Fig. 3 for $t=20.0$ s to $t=3.0$ s. (b) Detail of the deformation near the origin.

Equation (30) describes rays launched away from the z axis, with $r_0>0$. Rays launched towards the z axis are described by Eq. (30) with $\omega\rightarrow-\omega$. Equations (26) and (27) together describe the z coordinate of a ray from its initial position to its radial turning point. The upper portion of the ray is determined by $z\rightarrow 2z_{tp}-z$.
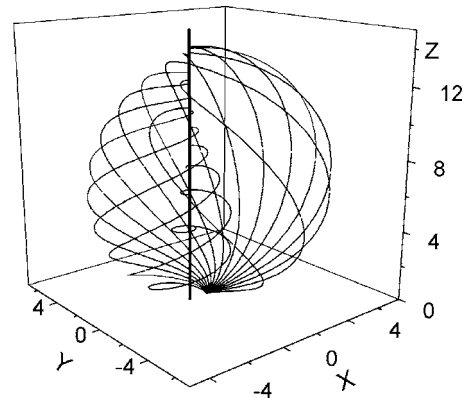
FIG. 6. Sample three-dimensional ray trace for $k=0.1$. Only rays launched away from the origin are presented in the interest of cleanliness. A vertical line is placed at the origin of the x-y plane as a reference to better illustrate the spiraling of the rays toward the z axis.

FIG. 4. (a) Wavefronts for the same scenario as depicted in Fig. 3 for $t=1.0$ s to $t=2.0$ s. (b) Detail of the deformation near the origin.

VI. CONCLUSIONS

The technique used in this letter is an ubiquitous part of mathematical physics, finding its greatest use in general relativity. The deep connection between isometry and Noether's theorem is well known in the study of dynamical systems.[18] The application of this method to acoustics serves a dual purpose in (a) encapsulating a plethora of very common, disconnected results in one easily accessible principle, geometric symmetry, while illustrating the connection between these results, and (b) providing a general technique for dealing with ray systems that do not possess Cartesian symmetry. Although only two examples were given, cylindrical and Cartesian symmetry, the method applies equally well to any system of coordinates.

ACKNOWLEDGMENTS

[1] E. T. Kornhauser, "Ray theory for moving fluids," J. Acoust. Soc. Am. **25**, 945–949 (1953).

[2] R. M. Wald, *General Relativity* (Univ. of Chicago, Chicago, 1984).

[3] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, New York, 1973).

[4] G. S. Heller, "Propagation of acoustic discontinuities in an inhomogeneous moving liquid medium," J. Acoust. Soc. Am. **25**, 950–951 (1953).

[5] P. Ugincius, "Acoustic-ray equations for a moving, inhomogeneous medium," J. Acoust. Soc. Am. **37**, 476–479 (1965).

[6] R. J. Thompson, "Ray theory for an inhomogeneous moving medium," J. Acoust. Soc. Am. **51**, 1675–1682 (1972).

[7] P. Ugincius, "Ray acoustics and Fermat's principle in a moving inhomogeneous medium," J. Acoust. Soc. Am. **51**, 1759–1763 (1972).

[8] R. Courant and D. Hilbert, *Methods of Mathematical Physics Volume II* (Wiley, New York, 1962).

[9] R. W. White, "Acoustic ray tracing in moving inhomogeneous fluids," J. Acoust. Soc. Am. **53**, 1700–1704 (1973).

[10] W. G. Unruh, "Experimental black hole evaporation?" Phys. Rev. Lett. **46**, 1351–1353 (1981).

[11] Greek indices run from 0 through 3, with 0 indicating the time direction and 1, 2, 3 corresponding to spatial directions. Roman indices run from 1 through 3 and represent only spatial components of a vector. Repeated indices imply summation.

[12] A dot appearing above a quantity represents differentiation with respect to λ, which is an affine parameter along the acoustic ray.

[13] Here we adopt a convention that lower/upper case symbols refer to quantities expressed as a function of $\{x\}/\{q\}$.

[14] Here the terms energy and momentum refer, in a generalized sense, to the canonical conjugate variables of the time and space coordinates, respectively. To have proper units the units of the metric must be redefined. These variables may also be identified with the frequency and wave number of the acoustic field (see Ref. 9).

[15] In this case we define θ to be the angle between the wavefront normal and the x axis.

[16] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, New York, 1998).

[17] We have tacitly set $\kappa_0 = 1$ in Eq. (21).

[18] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd ed. (Springer-Verlag, New York, 1989).

A time-variant impulse response method for acoustic scattering from moving two-dimensional surfaces[a)]

Richard S. Keiffer[b)]
Naval Research Laboratory, Code 7181, Stennis Space Center, Mississippi 39529-5004

Jorge C. Novarini
Planning Systems, Inc., 21294 Johnson Road, Long Beach, Mississippi 39560-9702

Robert W. Scharstein
Electrical Engineering Department, University of Alabama, Tuscaloosa, Alabama 35487-0286

A time-variant impulse response method is proposed, developed, tested, and shown to provide new insights into different scattering problems involving moving surfaces. The method proposed is general, conceptually straightforward, and can accommodate moving sources and receivers. While the time-variant impulse response method has been developed specifically in this study to extend the capability of the wedge assemblage (WA) model [e.g., R. S. Keiffer and J. C. Novarini, J. Acoust. Soc. Am. **107**, 27–39 (2000)], the framework can be applied to other scattering models including those based in the frequency domain. Calculations involving moving periodic surfaces and moving receivers are presented and compared with good results to a generalized grating equation and small wave-height approximation perturbation theory. The time-variant impulse response model is also applied to time-evolving sea surfaces, and the previously published results of Pourkaviani and Willemsen [J. Acoust. Soc. Am. **90**, 426–432 (1991)] are confirmed. It is also shown that windward-oriented backscatter geometries can lead to a Doppler spectrum that peaks at higher than expected Doppler shifts. [DOI: 10.1121/1.1992687]

I. INTRODUCTION

Previously, the details of a time-domain model for the (finite) acoustic impulse response of *static*, two-dimensional (2D), impenetrable, rough surfaces that is based on wedge diffraction theory[1] were reported. That approach, referred to as the wedge assemblage (WA) method, was shown to compare very well with a "benchmark accurate" frequency-domain solution over a broad range of grazing angles for the problem of backscattering from 2D sea surfaces. Earlier implementations of the WA method have been proven accurate for scattering from pressure-release (soft) sinusoidal surfaces,[2] and soft and hard disks.[3] Over the years since the WA method was first suggested, there have been many published papers testing this modeling approach. The interested reader can refer to the papers cited in Ref. 1, also Chaps. 11 and 12 of Medwin and Clay's book,[4] and the exact impulse response solution[5] on which the WA model is based. The work by Svensson *et al.*[6] deriving analytic expressions for the response of truncated wedges may also be of interest. More recently, a hybrid ray-WA model has been described that allows for acoustic scattering from rough surfaces over-lying refracting, range-independent media.[7] Finally, the important work by Chu,[8] and also Davis and Scharstein,[9] extending the exact Biot–Tolstoy solution to density-contrast

wedges, opens the way for the application of the WA model to certain seafloor scattering problems.

The main purpose of this paper is to describe and validate a method for extending the WA model to scattering problems in which the surface, source, and receiver are in motion. The time-variant interface scattering problem is of significant practical interest in underwater acoustics where, for example, sonar systems involving moving sources and receivers are common, and significant reverberation from time-evolving seas may be unavoidable. The approach developed in this study uses time-variant linear filter theory as a general framework, and addresses the kinematics of the problem in the time domain where the impulse response of a moving interface is the fundamental quantity that needs to be calculated. The process of scattering, in addition to its spatial dependence, is regarded as a time-variant linear filter, the complete specification of which is provided by the time-variant impulse response function. This viewpoint, when coupled to an ordered accounting of single- and multiple scattering contributions, leads to a significant simplification of the kinematics of the problem. And, it results in conceptually straightforward numerical algorithms for computing the received signal even under circumstances where the source, receiver, and scattering surface all move in complicated trajectories.

The utility of the time-variant impulse response method is not restricted to extending the WA model to time-varying scattering problems. For example, the method has been used within the framework of small wave-height perturbation

theory[10] to examine the implications of the so-called frozen surface approximation. In addition, as pointed out by Wetzel,[11] most treatments of rough surface scattering are derived from a boundary value problem in integral form based on Green's theorem in the frequency domain. While the linearity of the governing equations allows the response of a surface to an incident signal to be assembled by Fourier synthesis from single-frequency solutions, there are difficulties in the interpretation of this approach, particularly under dynamic time-varying conditions. One of the advantages of working in the time domain is that it allows for a causal-based analysis of the scattering process. Interactions can be followed in sequence as they occur. In the frequency domain, a direct identification of the sequence of events is not possible because of the intrinsic limitation imposed by the cw condition.

Finally, it should be pointed out that a moving interface causes the surrounding medium to move as well. Censor,[12] acknowledging that "strictly speaking the medium's effects and scatterer's boundary effects should be considered simultaneously," has examined this issue within the limiting circumstances of irrotational flow and slow variations in the ambient medium in space and time relative to distances and times on the order of the acoustic wavelength and period. He found that the (Doppler) effects of the medium motion are comparable to the effects of the moving interface, but become negligible as the distance from the observer to the region of the medium that is in motion is increased. On the other hand, the (Doppler) effects of interface motion on the scattered field are independent of distance.

The basic layout of the paper is as follows. In Sec. II a brief example problem is considered and the Doppler spectrum is derived to introduce the basic equations from time-variant linear filter theory and the impulse response approach. In Sec. III, the kinematics of insonifying a moving interface and the single-scatter approximation are discussed. Section III closes with details of the numerical implementation of the time-variant WA method, a discussion of sampling rates, and the numerical calculation of the bifrequency function.[13] In Sec. IV, the time-variant WA method is exercised in a series of numerical experiments that forms the basis of an initial validation of the model.

II. EXAMPLE PROBLEM: DOPPLER SPECTRUM AT A MOVING POINT RECEIVER

To briefly introduce nomenclature, establish transform conventions, and generally define terms, a simple example problem is considered in this section. Imagine a motionless medium where the sound speed is constant (c) throughout and a motionless point source located at $\bar{R}_S$ that instantaneously (at time $t=\tau$) begins injecting mass at a constant rate s_0. The impulsive pressure wave that radiates outward from the source is a solution to the wave equation

$$\nabla^2 p(\bar{r},t) - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}p(\bar{r},t) = -s_0\delta(t-\tau)\delta(\bar{r}-\bar{R}_S). \quad (1)$$

In the context of linear filter theory, the solution to Eq. (1), with due regard to the boundary conditions, is called the time-varying impulse response function. It is denoted in this paper, without explicit reference to its spatial dependency as $h(t,\tau)$. In free space, this time-variant impulse response function is

$$h(t,\tau) = \frac{s_0}{4\pi|\bar{r}-\bar{R}_S|}\delta\left(t-\tau-\frac{|\bar{r}-\bar{R}_S|}{c}\right). \quad (2)$$

The impulse response function becomes time variant when, for example, the receiver undergoes uniform rectilinear motion, $\bar{r}=\bar{R}_0+\bar{u}t$. Letting $\bar{R}=\bar{R}_0-\bar{R}_S$ and assuming that $|\bar{R}| \gg |\bar{u}T|$, that is, assuming that over the duration of the observation interval (T) the receiver travels a small distance compared to the initial source–receiver separation, then $|\bar{R}+\bar{u}t| \approx R+\hat{R}\cdot\bar{u}t$ and $|\bar{R}+\bar{u}t|^{-1}\approx R^{-1}$ are valid approximations, and the time-variant impulse response function can be accurately approximated by

$$h(t,\tau) \approx \frac{s_0}{4\pi R}\delta\left(qt-\tau-\frac{R}{c}\right), \quad (3)$$

with $R=|\bar{R}|$, $\hat{R}=R/|\bar{R}|$, and $q=1-\hat{R}\cdot\bar{u}/c$. The bifrequency function, $B(f,v)$, results from the sequential Fourier transforming the time-variant impulse response function with respect to t and τ. The following convention is adopted:

$$B(f,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(t,\tau)e^{i2\pi ft}e^{-i2\pi v\tau}dt\,d\tau. \quad (4)$$

In the case of uniform receiver motion, the "far-field" bifrequency function is

$$B(f,v) = \frac{s_0}{4\pi R}\delta(f-qv)e^{i2\pi vR/c}. \quad (5)$$

The bifrequency function has the attractive property that the frequency response due to a particular input signal can be obtained from the overlap integral[13]

$$P_{\text{out}}(f) = \int_{-\infty}^{\infty} P_{in}(v)B(f,v)dv. \quad (6)$$

Here, $P_{\text{in}}(v)$ is the Fourier transform of the input signal $p_{\text{in}}(t)$. A time-harmonic input signal having frequency $v_0 > 0$, has, in this example, the output signal

$$P_{\text{out}}(f) = \frac{s_0 e^{i2\pi Rf/qc}}{8\pi R}[\delta(f+qv_0)+\delta(f-qv_0)]. \quad (7)$$

From Eq. (7) it can be seen that the received signal is a frequency-shifted copy of the input signal. Looking at the positive received frequencies ($f>0$), the shift in received frequency is, as expected, $\eta=-[(\hat{R}\cdot\bar{u})/c]v_0$, where shift frequency η is defined as $\eta=f-v_0$. Closely related to the frequency spectrum of the output signal is the (one-sided) Doppler spectrum. It is defined for $f>0$ as twice the magnitude squared of the received frequency spectrum when expressed in terms of the shifted frequency η. The Doppler spectrum is the main quantity of interest in upcoming sections of this paper where numerical calculations are presented. There, finite duration input signals are considered. Of course, in the development of the far-field form of the impulse response

function, Eq. (3), it was already assumed that the input signal has a finite duration. Therefore, Eq. (7) represents the limiting form of the far-field output spectrum for very long-duration input signals. When the receiver is not far from the source, or if the receiver motion is not uniform rectilinear motion, a closed-form expression for the received spectrum may not be available. In that case one can always proceed numerically. If the details of the windowing function, $w(\tau)$, are important, then the frequency spectrum of the output signal is calculated from

$$P_{\text{out}}(f) = \frac{1}{2} \int_{-\infty}^{\infty} W(\upsilon)[B(f, \upsilon + \upsilon_0) + B(f, \upsilon - \upsilon_0)]d\upsilon, \qquad (8)$$

where $W(\upsilon)$ is the Fourier transform with respect to τ of the window $w(\tau)$.

The simple example developed in this section illustrates the application of time-variant linear filter theory to a moving receiver problem. It is important to note that the methodology can be easily implemented numerically to handle irregular trajectories and include source motion as well. And, in comparison to other methods suitable for that general problem [e.g., Ref. 14], the quasistatic approach just described is conceptually quite simple.

III. SINGLE-SCATTER IMPULSE RESPONSE OF A MOVING SURFACE

The following definitions are adopted regarding single and multiple scattering: the single-scatter component of the impulse response is due to those points on the interface (secondary sources) that are directly insonified by the source and in turn directly insonify the receiver. Multiple scattering is a generic term that refers to the additional scattering that follows from the reinsonification of the interface by these original and subsequent secondary sources. In most cases, the single-scatter component of the impulse response dominates and the practical limitations imposed by the additional computations needed to include multiple scattering effects are not typically justified by substantial increases in the accuracy of the predictions. Therefore, in what follows, the discussion of the approach used to extend the WA approach to time-variant scattering problems is limited to the single-scatter component of the impulse response. It should be kept in mind, however, that the methodology could be applied in a self-consistent manner to each successive order of scattering.

A. Kinematics and the single-scatter impulse response

Suppose that a moving (subsonic) point source, having instantaneous location $\bar{R}_S(t)$, stops at time $t = \tau$ and immediately begins injecting mass at a constant rate into a homogeneous (static) fluid medium. The pressure impulse that radiates outward from the source travels (with speed c) from the location of the source at the time of emission and insonifies a moving (subsonic), impenetrable, rough surface, $z_s(x, y, t)$. As the impulsive source wavefront sweeps across the time-varying surface, it intercepts or detects each point on the surface at a discrete instant of time. Call this time of detection t_d. By definition, t_d solves the equation

$$t_d - \tau - \frac{|x\hat{x} + y\hat{y} + z_s(x, y, t_d)\hat{z} - \bar{R}_S(\tau)|}{c} = 0. \qquad (9)$$

The locus of surface points detected by the impulsive source signal defines a new, effective surface, $\zeta(x, y, \tau)$, the frozen, quasistatic surface[11]

$$\zeta(x, y, \tau) = z_s(x, y, t = t_d). \qquad (10)$$

It is clear from causality that the time history of the surface prior to the arrival of the incident impulse cannot affect the scattered response. Similar considerations lead to the conclusion that the single-scatter component of the scattered response cannot be affected by the time history of the surface after the initial source insonification. In other words, the single-scatter component of the scattered response depends only on the shape of the detected surface. The effect of source and surface motion on the time variation of the impulse response function can be modeled (in the single-scatter approximation) by examining the impulse responses of a time sequence of the quasistatic surfaces, $\zeta(x, y, \tau)$.

Receiver motion can also be included. Regarding each point on the surface $\zeta(x, y, \tau)$ as a secondary source that emits a scattered impulse at $t = t_d$, let t_R be the time at which that scattered signal reaches the moving receiver at $\bar{R}_R(t_R)$

$$t_R = t_d + \frac{|x\hat{x} + y\hat{y} + \zeta(x, y, \tau)\hat{z} - \bar{R}_R(t_R)|}{c}. \qquad (11)$$

Each secondary source on $\zeta(x, y, \tau)$ emits a scattered impulse that reaches the moving receiver at a discrete, potentially different, future location $\bar{R}_R(t_R)$. The effect of receiver motion can be included in the modeling by using the correct static receiver location in the impulse response calculation of each secondary source and then summing all the impulse responses with due respect to time of arrival.

B. Numerical implementation

The details of a numerical implementation of the WA model for backscattering problems involving static 2D surfaces have been discussed previously.[1,3,4] Although it is appropriate to repeat some of that material here, the focus of this section is on the numerical implementation of the methodology being proposed to extend the WA model to time-variant problems. Begin by assuming that the time-varying surface, $z_s(x, y, t)$, is represented on a fixed uniform spatial grid ($\Delta x = \Delta y$) and that the time evolution of the surface is known at all times. Then, at any discrete horizontal location (x_i, y_j), the value of the surface height function detected by the impulse emitted by the source at time $t = \tau$ defines $\zeta(x_i, y_j, \tau)$. This quasistatic surface can be determined numerically, for example, by applying any of a number of root-finding algorithms to Eq. (9). In some cases, closed-form expressions can be derived that accurately approximate the detected time and surface [for example, see the Appendix of this paper or Ref. 10].

Finally, many surfaces of interest are modeled by filtering random numbers in the wave number domain and then Fourier transforming to the spatial domain. In the case of sea

surfaces, the time evolution of the interface is accomplished by advancing or retarding in time the phase angle of different spectral components by amounts commensurate with the gravity wave dispersion relation. For these surfaces it is not efficient to specify the entire surface at many different times in order to determine t_d and $\zeta(x_i, y_j, \tau)$ at a particular (x_i, y_j) location. To limit the number of forward and inverse two-dimensional FFTs, the strategy used in this paper is to generate, at regular time intervals, a sequence of "snapshots" of the spatially discretized surface $z_s(x_i, y_j, t)$ and then, assuming the temporal sampling rate is adequate, interpolate to obtain surface heights at intermediate times.

The preceding paragraph touches on three different methods that can be used to determine the surface detected by an impulse emitted at time, $t = \tau$. To calculate the (finite) impulse response of this surface using the WA method, the surface $\zeta(x_i, y_j, \tau)$ is "tiled" or tessellated using, for convenience, triangularly shaped facets.[1,4] These triangular facets are determined by first interpolating between any four surface heights of the square sample area to create a center surface height value. This divides the primitive area into triangularly shaped pieces. Adjacent triangular facets are used to define finite-length wedge apices. The diffracted response (P_δ) of every (nonshadowed) finite-length apex in the assemblage of wedges is calculated at discrete times, $t_i = (i-1)\Delta t$, from the Biot–Tolstoy solution and added with due respect to time of arrival

$$h(t_i, \tau) = \sum_{m=1}^{M} \varepsilon_m \frac{1}{\Delta t} \int_{t_i}^{t_i + \Delta t} P_{\delta_m}(\bar{r}_S, \bar{r}_R, t, \tau) dt. \quad (12)$$

With source wedge coordinates r_0 and θ_0, receiver coordinates r and θ, wedge angle θ_w, and offset distance Z (all of which are specific to the mth finite-length apex), the diffracted component is given by

$$P_\delta(t) = \left(\frac{-s_0 c}{8 \pi \theta_w} \right) [r r_0 \sinh(Y)]^{-1} [D_{++} + D_{--} + D_{+-} + D_{-+}], \quad (13)$$

$$Y = \cosh^{-1} \left[\frac{c^2 t^2 - (r^2 + r_0^2 + Z^2)}{2 r r_0} \right], \quad (14)$$

$$D_{\pm\pm} = \frac{\sin \left[\frac{\pi}{\theta_w} (\pi \pm \theta_0 \pm \theta) \right]}{\cosh \left[\frac{\pi Y}{\theta_w} \right] - \cos \left[\frac{\pi}{\theta_w} (\pi \pm \theta_0 \pm \theta) \right]}. \quad (15)$$

In Eq. (13), s_0 is the rate of mass injection by the source and c is the sound speed. It is understood that the mth finite-length wedge apex has a diffracted response that is zero outside of the time interval $[t_{a_m}, t_{b_m}]$. The factor ε_m can take on the value of 1.0 or 0.5 depending on the symmetry of the finite-length (physical) apex about the least time point on the infinite length (virtual) apex that contains it. Figure 1 shows a schematic of the gridded surface tessellation and gives details of the wedge specific coordinate system used in the Biot–Tolstoy solution.
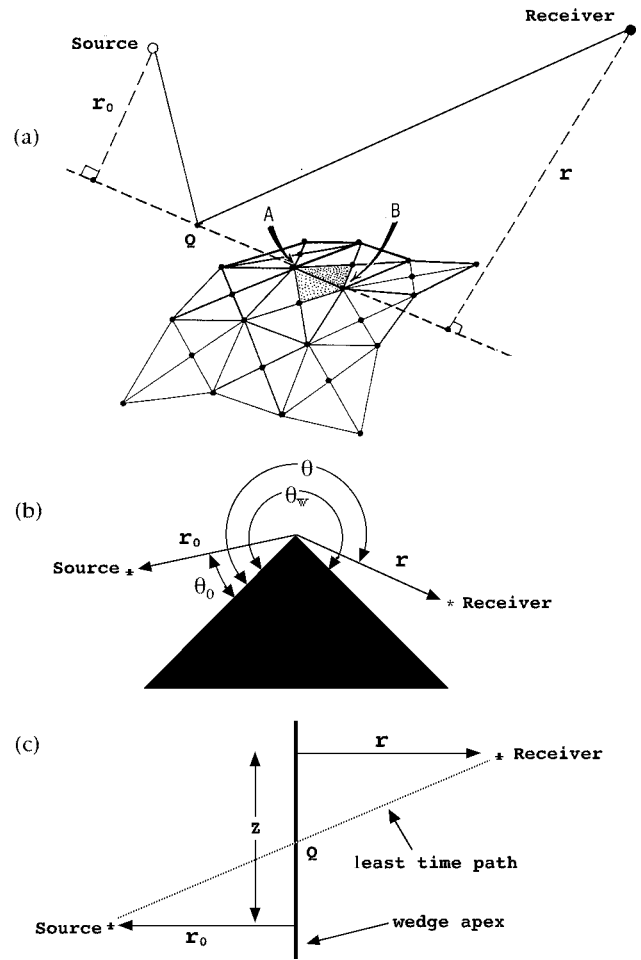
FIG. 1. (a) The decomposition of a 2D surface into triangular facets with line segment **AB** coincident with a finite-length wedge apex. **Q** is the point on the associated infinite apex for the least time of any source-to-apex-to-receiver transmission. (b) Side and (c) plan views of the wedge geometry used in the Biot–Tolstoy solution.

For a fixed source emission time τ, sampling the impulse response function in time t, at interval Δt, determines, after discrete Fourier transform (DFT), a time-variant transfer function, $H(f_i, \tau)$, and (discrete) transform variable, f_i. $H(f_i, \tau)$ is defined on principal interval, $-f_{\text{Nyq}} \leq f_i \leq f_{\text{Nyq}}$, the highest frequency of which is the Nyquist frequency for the received frequency: $f_{\text{Nyq}} = (2\Delta t)^{-1}$. The finite extent of the rough surface ultimately limits the length of impulse response function and determines the received frequency resolution, $\Delta f = 1/N\Delta t$. The discrete received frequencies are given by $f_i = -f_{\text{Nyq}} + (i-1)\Delta f$, where i is a positive integer that is no greater than $N+1$. Based on past experience with static sinusoids and sea surfaces, choosing Δt so that f_{Nyq} is about an order of magnitude larger than the highest received frequency of interest (f_{max}) is usually adequate to make the effects of aliasing acceptably small in the received frequency band of interest.

Similarly, for a fixed reception frequency, f_i, sampling the transfer function $H(f_i, \tau)$ as it varies with source emission time τ (M samples at interval $\Delta\tau$) introduces, after the corresponding DFT, another transform variable that can be identified as the discrete source frequency, $v_j = -v_{\text{Nyq}} + (j-1)\Delta v$. Here, j is a positive integer that is no greater than $M+1$. The

discrete bifrequency function, $B(f_i, v_j)$, that results after this second DFT is specified in a two-dimensional frequency space over a rectangularly shaped principal band that is limited in one dimension by $-f_{Nyq} \leq f_i \leq f_{Nyq}$ and the other by $-v_{Nyq} \leq v_j \leq v_{Nyq}$. Here, the source Nyquist frequency, $v_{Nyq} = (2\Delta\tau)^{-1}$, should be large enough to satisfactorily minimize the "folding" of high-frequency information down into the principal frequency band.

There is usually some numerical efficiency that can be gained by bandshifting or heterodyning the principal interval associated with the source frequency, $-v_{Nyq} \leq v_j \leq v_{Nyq}$, so that for each received frequency f_i the discrete bifrequency function is centered on the energetic part of the spectrum. For the problems considered in this paper where the Doppler shifts are a small fraction of the source frequency, it is expected that nearly all of the scattered energy will be centered about the line $v=f$. Multiplying the sampled transfer function, $H(f_i, \tau_j)$, by a time-harmonic signal, $\cos(2\pi f_i \tau_j)$, prior to the second DFT (involving τ and v) has the effect of specifying the discrete bifrequency function over a stair-stepped principal interval that straddles the line $v=f$. The received spectrum is obtained by holding the source frequency fixed and examining (in increments of Δf) the bifrequency function from one horizontal limit of the stair-stepped principal band to the other. The Doppler spectrum is simply the received power spectrum expressed in terms of the shift frequency variable, $\eta = f - v$.

IV. EXAMPLE CALCULATIONS AND COMPARISONS

In this section, the extension of the WA method to time-variant scattering problems is used to calculate the frequency spectrum of the scattered pressure in a variety of backscattering scenarios involving moving single- and two-component sinusoidal surfaces, time-evolving sea surfaces, moving receivers, and static point sources. Moving periodic surfaces were particularly useful for an initial evaluation of the time-variant WA method because under far-field conditions they provide discrete conditions for constructive interference, and a generalized grating equation is available to provide an independent prediction for the shifted source and received frequencies associated with different diffracted orders.

To check the energy partition among different diffracted orders, a small wave-height approximation (SWHA) perturbation theory for moving standing wave surfaces that has been modified to correctly account for the basic kinematics of the problem[10] is used as a standard (see the Appendix). In Ref. 10, it is shown that by treating the kinematics of the moving surface scattering problem in the time domain, an explicit quasistatic or frozen surface approximation naturally occurs prior to the small wave-height approximation. In the standard approach, the frozen surface approximation occurs as a result of the SWHA itself. It was also shown in Ref. 10 that the standard SWHA approach predicted (for the periodic surfaces) upshifted diffracted orders that lost energy, and downshifted diffracted orders that gained energy. This unphysical behavior is corrected in an SWHA perturbation approach that makes an explicit frozen surface approximation prior to invoking the small wave-height approximation. The

backscattered pressure fields were calculated via this new SWHA perturbation theory, which was carried out to third order in dimensionless expansion parameter [see the Appendix]. These results were used to benchmark the time-variant WA method in the limit of very small roughness, and then to track the initial agreement and eventual departure of the WA method from the SWHA results as the surface roughness was increased from very small to moderate values.

In the final section, the time-variant WA method is applied to the problem of backscattering from fully developed, time-evolving seas due to different wind speeds. Again, SWHA perturbation theory is used as a benchmark but, because in this case the calculation of higher order terms is substantially more involved, these scattered pressures are restricted to the first-order estimates. To augment this stage of the validation, rather, to corroborate discrepancies between the time-variant WA predictions and first-order perturbation theory, the previously published work of Pourkaviani and Willemsen,[15] which looked at the effect of higher-order perturbative corrections on the location of the spectral peak, is introduced. Finally, the work by Fuks[16] is specialized for the particular cases studied and used to generate estimates for the spreading of the Doppler spectrum.

A. Moving sinusoidal surfaces and moving receivers: Comparisons with the grating equation

In this first set of example calculations, the time-variant WA method was used to calculate the Doppler spectrum of the pressure field backscattered from a 512-m-long (and wide) pressure-release, long-crested, sinusoidal surface having a wavelength of 8 m, that moved with x-directed velocity, $u_x = 8$ m/s

$$z_s(x,t) = a \cos\left[\frac{2\pi}{\Lambda}(x - u_x t)\right]. \tag{16}$$

The scattered signal was detected at a distant receiver (restricted to the x–z plane) that receded from the origin with velocity (in m/s) $\bar{v} = -10\hat{x} + 5.775\hat{z}$. The sound speed in the medium containing the source and receiver, and underlying the pressure-release boundary, was constant ($c = 1500$ m/s). The mean grazing angle of incidence was $30°$. At time $t=0$, when the first finite impulse ($\Delta t = 1.953 \times 10^{-4}$ s) was emitted by the source, the receiver was collocated with the source. Subsequent impulses were emitted by the source at 0.05-s intervals for a duration of 3.5 s. Source emissions propagated to and scattered from the sinusoidal surface and were collected at the moving receiver. After padding with zeros, an FFT ($N=32\,768$) of each impulse response was performed to yield the time-varying transfer function. For each discrete received frequency, the time-dependent transfer function was multiplied by the time-harmonic signal, $\cos(2\pi f_i \tau_j)$, as discussed in the previous section, again padded with zeros, and the $\tau \rightarrow v$ FFT was performed using $M=128$. The received power spectrum expressed in terms of the (Doppler)-shifted frequency $\eta = f - v$, is shown in Fig. 2(a). The combination of zero padding and time sampling rates resulted in a source frequency resolution that equaled the

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Keiffer *et al.*: Impulse response of moving surfaces   1287
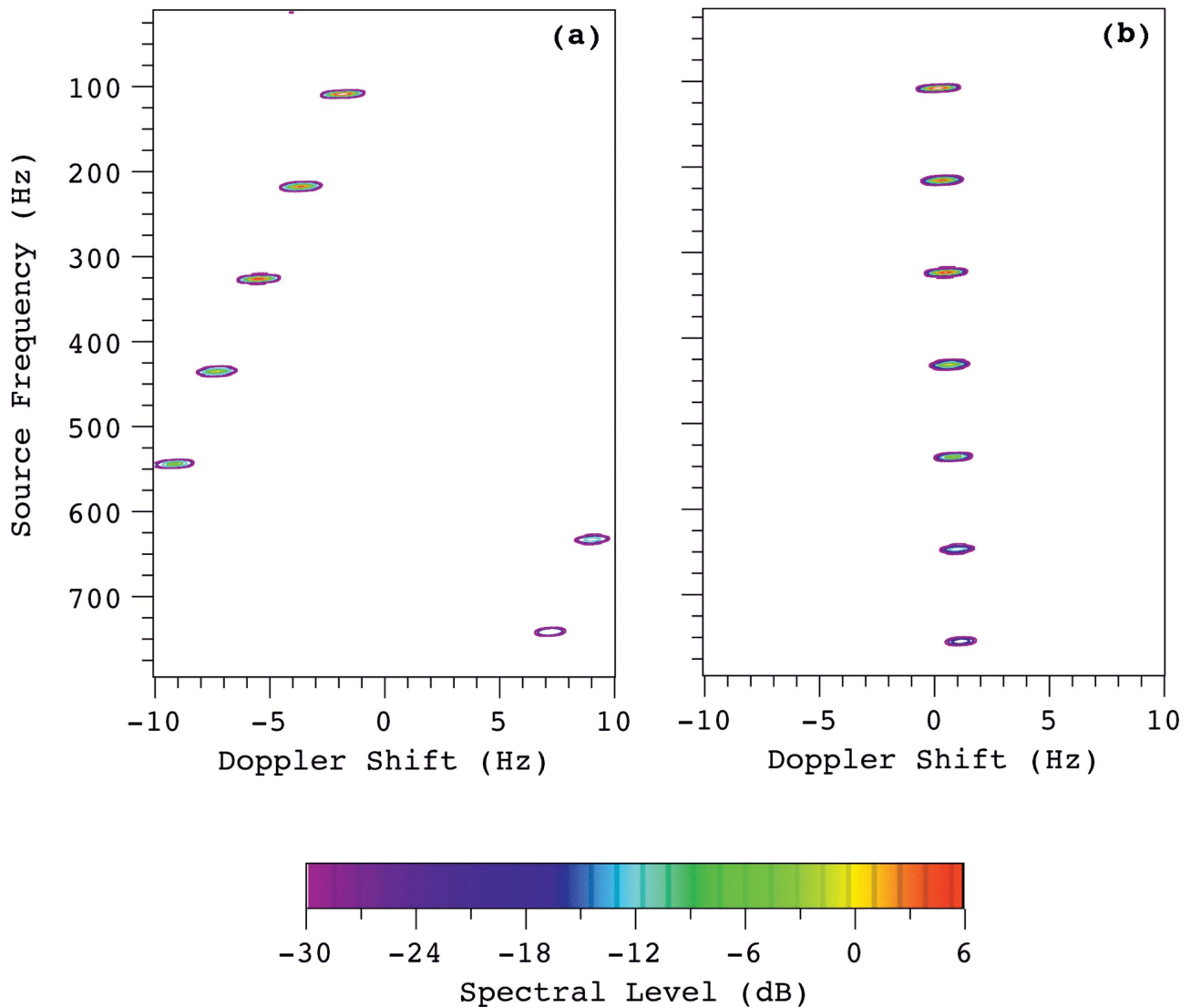
FIG. 2. Doppler spectra calculated by the time-variant WA model for backscattering (30° grazing) from a sinusoidal surface having an 8-m wavelength that is moving away from a fixed source with a horizontally directed speed of 8 m/s and detected at a receiver moving (a) away from the fixed origin of the surface and (b) toward the fixed origin of the surface.

received (and Doppler shift) frequency resolution: $\Delta\nu = \Delta\eta = 0.15\,625$ Hz. With the first diffracted order appearing at source frequency of 108.75 Hz, one can count in Fig. 2(a) five regularly spaced regions of constructive interference (diffracted orders, Bragg beams, Floquet modes, grating lobes, etc.) before the Doppler shift exceeds -10.0 Hz. At this point, diffracted orders appearing at higher source frequencies (above 600 Hz) become aliased due to an insufficient τ sampling rate. It can be seen that the most energetic of these aliased diffracted orders appear at positive Doppler shifts.

For this moving periodic interface problem, a grating equation can be derived that establishes the source and received frequencies required for constructive interference at the moving receiver[10]

$$\nu_n = n\left(\frac{c - u_x\hat{R}_{R_0} \cdot \hat{x}}{\Lambda(\hat{R}_S + \hat{R}_{R_0}) \cdot \hat{x}}\right), \qquad (17)$$

$$f_n = n\left(\frac{c + u_x\hat{R}_S \cdot \hat{x}}{c}\right)\left(\frac{c - \hat{R}_{R_0} \cdot \bar{v}}{\Lambda(\hat{R}_S + \hat{R}_{R_0}) \cdot \hat{x}}\right). \qquad (18)$$

When applied to the current problem, Eqs. (17) and (18) predict constructive interference for source frequencies that are integral multiples of 108.75 Hz, received frequencies that are integral multiples of 106.92 Hz, and received Doppler shifts, $\eta_n = f_n - \nu_n$, that are positive integral multiples of -1.83 Hz.

In Fig. 2(b), this numerical experiment is repeated with the direction of the sinusoid reversed ($u_x = -8$ m/s). In this case, the motion of the receiver is almost compensated by the motion of the surface. The result was that the source frequencies that allowed for constructive interference (107.75 Hz/order) at the receiver were slightly lower than their static values (108.25 Hz/order) and the Doppler shifts were small (0.16 Hz/order). Tables I and II compile for these two numerical experiments a comparison between WA-

TABLE I. The source frequencies (Hz) and Doppler shifts (Hz) at which diffracted orders (constructive interference) occur for the numerical experiments shown in Fig. 2(a). Comparison is between the time-variant WA model and the grating equations [Eqs. (17) and (18)].

Order no.	Time-variant WA method		Grating equations [Eqs. (17) and (18)]	
	Source freq. (Hz)	Doppler shift (Hz)	Source freq. (Hz)	Doppler shift (Hz)
1	108.75±0.16	−1.88±0.16	108.78	−1.83
2	217.66±0.16	−3.75±0.16	217.56	−3.66
3	326.25±0.16	−5.47±0.16	326.34	−5.49
4	435.00±0.16	−7.34±0.16	435.12	−7.32
5	543.91±0.16	−9.22±0.16	543.90	−9.15

predicted source and Doppler shift frequencies and the grating equation values. For all nonaliased orders, the time-variant WA model predictions were in agreement with the grating equation predictions to within the resolution of the numerical computation ±0.15 625 Hz. Before leaving this section, we note that the surface and receiver motion cause the conditions for constructive interference to change, and so both the source and received frequencies are altered as a result. Diffracted orders manifesting negative (positive) Doppler shifts are accompanied by source frequencies that are higher (lower) than the static surface values, which in this case would be 108.25 Hz/order.

B. Standing wave surfaces: Comparisons with SWHA perturbation theory

Employing a scattering geometry identical to the one used in the previous section, the time-variant WA model was used to calculate the Doppler spectrum from a series of five different standing wave surfaces

$$z_s(x,t) = 2a \cos[2\pi x/\Lambda]\cos[2\pi Ft]. \quad (19)$$

In all cases, the opposite traveling sinusoids producing the standing wave pattern had wavelengths of 8.0 m, x-directed speeds of ±10.0 m/s, $F=±1.25$ Hz, and with amplitudes (a) that ranged from 0.0125 to 0.5 m. The Doppler spectra of the pressure fields backscattered from four of these

TABLE II. The source frequencies (Hz) and Doppler shifts (Hz) at which diffracted orders (constructive interference) occur for the numerical experiments shown in Fig. 2(b). Comparison is between the time-variant WA model and the grating equations [Eqs. (17) and (18)].

Order no.	Time-variant WA method		Grating equations [Eqs. (17) and (18)]	
	Source freq. (Hz)	Doppler shift (Hz)	Source freq. (Hz)	Doppler shift (Hz)
1	107.81±0.16	0.16±0.16	107.75	0.16
2	215.63±0.16	0.31±0.16	215.51	0.33
3	323.28±0.16	0.47±0.16	323.26	0.49
4	431.09±0.16	0.63±0.16	431.01	0.65
5	538.91±0.16	0.78±0.16	538.77	0.81
6	646.72±0.16	0.94±0.16	646.52	0.98
7	754.53±0.16	1.09±0.16	754.27	1.14

surfaces are shown in Fig. 3. It should be noted that in each case, the dynamic range of the plot is constant (45 dB), but the maximum values plotted are adjusted to accommodate the most energetic diffracted order.

Beginning with Fig. 3(a), the smallest amplitude surface where $a=0.0125$ m, it can be seen that there are only three diffracted orders that appear to be detectable. One of these orders, the unshifted one, is very weak compared to the other two. Based on perturbation theory (see the Appendix), this is to be expected since, at the received frequencies (f) where these first diffracted orders appear, the dimensionless expansion parameter, ka, is very small compared to 1 ($ka \approx 0.005$). Since ka is small, the Doppler spectrum should resemble the linear superposition of the individual Doppler spectra from each of the two opposite-traveling sinusoidal components considered alone.

Now consider Fig. 3(b), where, after increasing the surface amplitude a to 0.025 m, two additional diffracted orders appear. These additional regions of constructive interference fall along lines that pass through the origin and the first diffracted orders. From the earlier calculations shown in Fig. 2, it is known that these two newly visible diffracted orders depend entirely on one sinusoid or the other. It is interesting to note that of the three diffracted orders with source frequencies near 216 Hz, the zero-shifted diffracted order, which depends on both sinusoidal components, has a scattered intensity that is 6 dB stronger than either of the other two.

In Fig. 3(c), the effect of increasing amplitude a to 0.1 m is shown. Four more diffracted orders are now visibly energized. The conclusion of this series of plots is shown in Fig. 3(d). Here, the backscattered Doppler spectrum is plotted for a standing wave composed of opposite-traveling sinusoids having surface wave-height amplitudes equal to 0.5 m. Almost a dozen diffracted orders are now detectable.

Table III contains a list of (normalized) scattered intensities from the spectral peaks of the time-variant WA data shown in Fig. 3. Also listed in Table III are the locations of the various diffracted orders, that is, the source and shift frequencies (received–source) where these peak spectral intensities occur. Calculations for the peak scattered intensities and their locations were also made using SWHA perturbation theory under the assumption that true plane waves and infinite surfaces were involved. In those calculations, the details of which are discussed in the Appendix, terms up to third order or in the dimensionless expansion parameter ka were retained. To facilitate a comparison between SWHA perturbation theory and the time-variant WA method for the partition of the scattered energy among the various diffracted orders, the following normalization scheme was developed. For both calculation methods, but independent of each other, the intensity of the most energetic diffracted order for the least rough surface ($a=0.0125$ m) was set to 0 dB and all other intensities were scaled relative to this value. Because in this smallest roughness case both the WA model and perturbation theory identified the same diffracted order as the most energetic, this self-normalization procedure allowed for an extended comparison between the two models.

In determining the validity of the time-variant WA

FIG. 3. Fixed source, fixed receiver Doppler spectra calculated by the time-variant WA model for monostatic backscattering (30° grazing) from a standing wave surface composed of oppositely traveling sinusoidal surfaces ($\Lambda=8.0$ m, $F=\pm1.25$ Hz) each having surface amplitudes of (a) 0.0125 m; (b) 0.025 m; (c) 0.1 m; and (d) 0.5 m.

Keiffer *et al.*: Impulse response of moving surfaces

TABLE III. Comparison between the predictions of the time-variant WA model and SWHA perturbation theory for the normalized scattered intensities (in dB) of diffracted orders detected in the numerical experiments shown in Fig 3. Standing wave surfaces have amplitude $2a$, are composed of oppositely traveling sinusoids having 8.0-m wavelengths, and x-directed speeds of ±10.0 m/s. Note that **** indicates that the intensity of the diffracted order is below the normalized numerical noise floor of the WA model (−56.3 dB).

Frequency (Hz)		Order		$a=0.0125$		$a=0.025$		$a=0.05$		$a=0.1$		$a=0.5$	
Source	Shift	m	n	Pert.	WA	Pert.	WA	Pert.	WA	Pert.	WA	Pert.	WA
106.4	3.75	1	2	****	****	****	****	−54.0	−44.6	−36.0	−27.2	6.0	12.4
107.6	1.25	0	1	0.0	0.0	6.0	6.0	12.0	12.1	18.0	18.2	31.0	34.6
108.9	−1.25	−1	0	−0.1	−.2	5.9	5.9	11.9	11.9	17.9	18.0	30.9	34.4
110.1	−3.75	−2	−1	****	****	****	****	−54.3	−45.2	−36.3	−27.5	5.7	12.2
215.3	2.5	0	2	−38.9	−38.6	−26.9	−26.9	−14.8	−14.7	−2.8	−2.6	25.2	25.1
216.5	0.0	−1	−1	−32.9	−32.7	−20.9	−20.9	−8.9	−8.7	3.2	3.3	31.1	31.1
217.8	−2.5	−2	0	−39.0	−38.9	−27.0	−27.0	−14.9	−14.9	−2.9	−2.8	25.0	24.9
322.9	3.75	0	3	****	****	−61.2	****	−43.2	−39.0	−25.1	−20.7	16.8	19.6
324.1	1.25	−1	2	****	****	−51.7	−51.7	−33.7	−29.2	−15.6	−11.1	26.3	29.6
325.4	−1.25	−2	1	****	****	−51.8	−51.8	33.7	−19.2	−15.6	−11.2	26.3	29.5
326.6	−3.75	−3	0	****	****	−61.3	****	−43.3	−38.2	−25.2	−20.9	16.7	19.3

method to predict the partition of scattered energy among the available diffracted orders, it is critical that there be agreement with the SWHA results in the "small roughness" regime where perturbation theory is expected to be accurate. From Table III, it can be seen that the agreement between the two models is excellent (less than 0.3-dB difference) for all detectable orders up to a surface roughness of $a=0.025$ m. Then, as expected when the surface amplitude increases, more diffracted orders become energized and the two models begin to depart from each other. This departure can be seen first in the diffracted orders at higher source frequencies of the $a=0.05$ m surface. When the surface amplitude reaches $a=0.5$ m, the strong agreement has significantly eroded. A notable exception to this trend is the second-order scattered intensities; the SWHA results track well against the time-variant WA model for all of the standing wave surfaces in Table III.

Before leaving this section, the following observation is worth noting and discussing. Table III shows that, in all cases, upshifted diffracted orders are slightly more energetic than their downshifted counterparts. While this result can be anticipated from the kinematics, the standard formulation of SWHA perturbation theory for moving surfaces would have predicted just the opposite relationship between scattered intensity and Doppler shift. As previously stated in this paper, the perturbation calculations were based on an approach in which an explicit frozen surface approximation is made before invoking the small roughness approximation [see the Appendix]. As a result, this error in which the upshifted diffracted orders are less energetic than their downshifted counterparts is corrected.

C. Application to sea surfaces

The results of the numerical experiments discussed in the previous section show that the time-variant WA method yields results that agree with SWHA perturbation theory in the limiting circumstance where that standard is known to be accurate. Those "two-component" results also demonstrate the increasingly important role that higher-order terms play as the magnitude of the dimensionless expansion parameter increases. When the scattering surface is composed of a broad band of spatial scales, the number of significant contributions to the overall Doppler spectrum that come from higher-order terms can proliferate. These higher-order terms contribute to the received spectrum with different Doppler shifts, sometimes contributing to the same locations in the two frequency spaces of the Doppler spectrum, and sometimes not. It is not at all clear how their cumulative effect will alter the regime over which first-order perturbation can be considered accurate. An implementation of higher-order perturbation theory for the moving multiscale interface scattering problem is beyond the scope of the current study. However, it is known from the published literature that the details of the surface motion can affect the validity of SWHA perturbation theory. In a study of backscattering from moderate wind speed, 2D, isotropic seas, Pourkaviani and Willemsen[15] showed that the effect of including the next few higher-order perturbative corrections in calculations for the scattered intensity was to move the peak of the Doppler spectrum away from the expected Doppler shift of the Bragg component and toward *smaller* (lower) Doppler shifts. For the assumed monostatic backscatter geometry, where k is the scattered wave number and ψ is the grazing angle, the Doppler shift associated with the Bragg-scale component is

$$\eta_{\text{Bragg}} = \frac{1}{2\pi}\sqrt{2gk\cos\psi}. \tag{20}$$

The results that follow demonstrate that this change in the location of the spectral peak away from the expected first-order Doppler shift is not, as some have suggested,[17] an artifact of scattering from isotropic seas. This phenomenon can be reproduced by the time-variant WA method for backscattering from anisotropic, 2D seas when the scattering plane is oriented crosswind to the main wind/wave direction. It is also shown, to our knowledge for the first time, that windward-oriented backscattering geometries can yield spectra that peak at Doppler shifts that *exceed* the shift of the Bragg component.
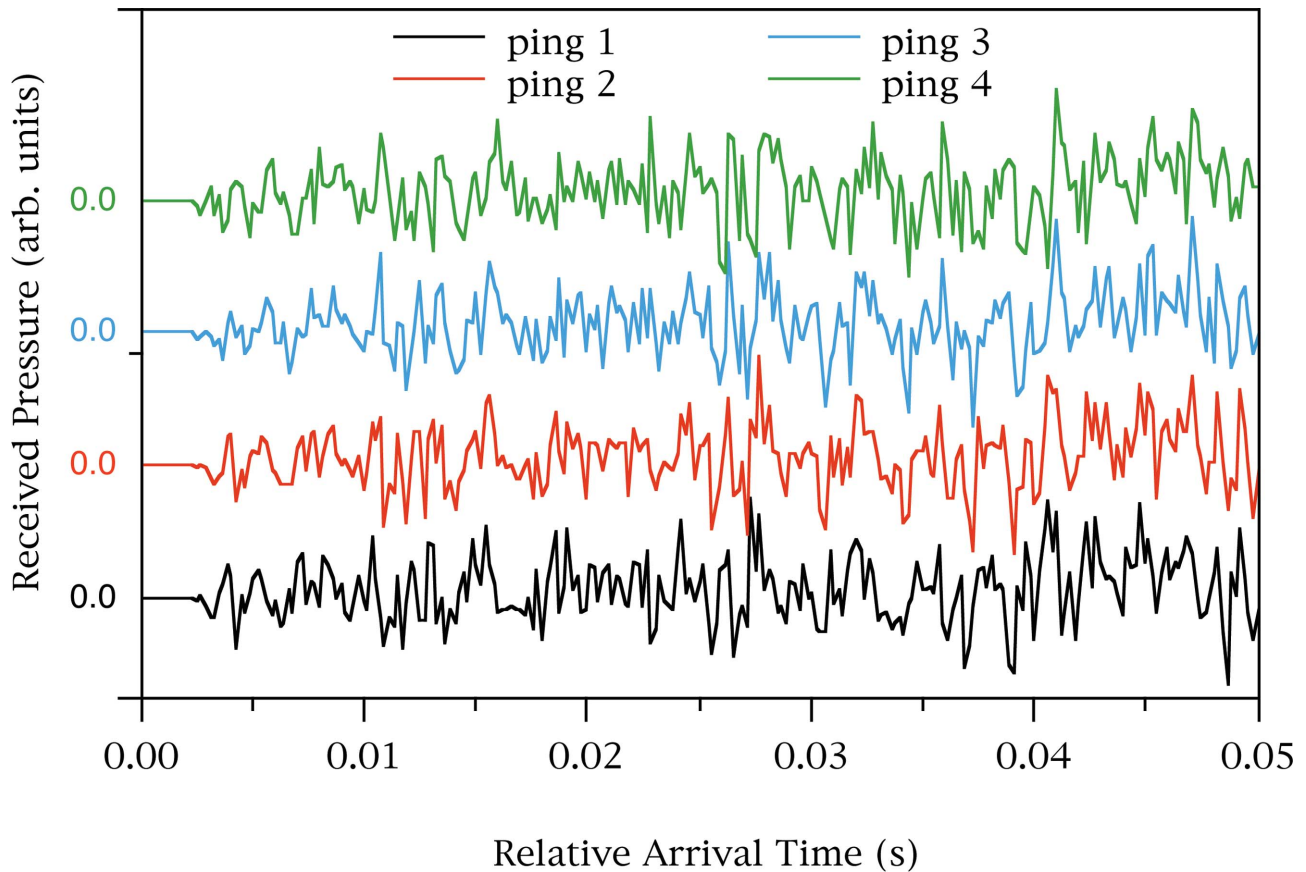
FIG. 4. Comparison of the first four impulse responses for monostatic backscattering from time-evolving seas that are traveling away from the source/receiver location. The plot shows the first 0.05 s of impulse responses that were approximately 0.3 s long. The time separation between pings was 0.1 s. The wind speed for this example was 5 m/s and the grazing angle was 30°.

The time–variant WA model was used to compute the average backscattered Doppler spectrum from ensembles of 2D time-evolving sea surfaces. The surfaces used in this study were modeled as fully developed seas having a $\cos^2(\phi/2)$ azimuthal dependence. Here, the wind direction is assumed to be into the $\phi=0°$ direction, and each spectral component of the surface was time evolved according to the dispersion relation for ocean gravity waves. Each surface realization had a Pierson–Moskowitz roughness spectrum that was due a wind speed (at 10-m height, U_{10}) of 5, 10, or 13 m/s. Surface realizations having the desired roughness spectrum were generated on a uniform 2D grid ($\Delta x=\Delta y =0.245$ m) having 1024 points on a side. The length of the surface generated was 251.3 m, which was long enough and finely sampled enough to include almost all of the expected surface variance as well as the longest waves having a significant contribution to the total variance. As in the previous examples, the sound speed in the medium containing the source and receiver was constant ($c=1500$ m/s), a monostatic geometry was assumed, and the nominal grazing/ scattering angle was 30°. Backscattering geometries oriented windward and crosswind to the main wind/wave direction were considered. A finite impulse ($\Delta t=1.953\times 10^{-4}$ s) was emitted by a distant source at 0.1-s intervals for 8.0 s. The source impulse propagated to, and scattered from, the quasi-static sea surface, and was then collected at the receiver. In Fig. 4, the beginning 0.05 s of the first four impulse re-

sponses ($\tau=0.0$, 0.1, 0.2, and 0.3 s) are shown. Note that, as expected for monostatic backscattering from seas moving away from the source/receiver, in the successive pings certain features of the time series can be seen arriving at later and later times.

After applying a Parzen window[18] to each impulse response and after padding with zeros, an FFT ($N=32\,768$) of each impulse response was performed to yield a time-varying transfer function. For each discrete received frequency, the time-dependent transfer function was then band-shifted as discussed previously, again windowed and padded with zeros, and the second FFT ($M=128$) was performed. This arrangement led to a source frequency resolution that equaled the received frequency resolution (and hence the Doppler shift resolution): $\Delta\upsilon=\Delta\eta=0.156\,25$ Hz. When identical signal processing was applied to the progression of impulse responses from an equivalent *static* surface, the spectral width of the windowed response could be measured. It was found that the scattered intensity was peaked on zero Doppler shift and dropped from this peak value with increasing Doppler shift. The scattered intensity was down by more than 4 dB for Doppler shifts greater than $\pm 0.156\,25$ Hz. Sidelobes from the Parzen window were more than 50 dB down from the mainlobe.

The average Doppler spectrum of the received signal was calculated from 15 realizations of the same sea surface spectrum. For each ensemble-averaged Doppler spectrum
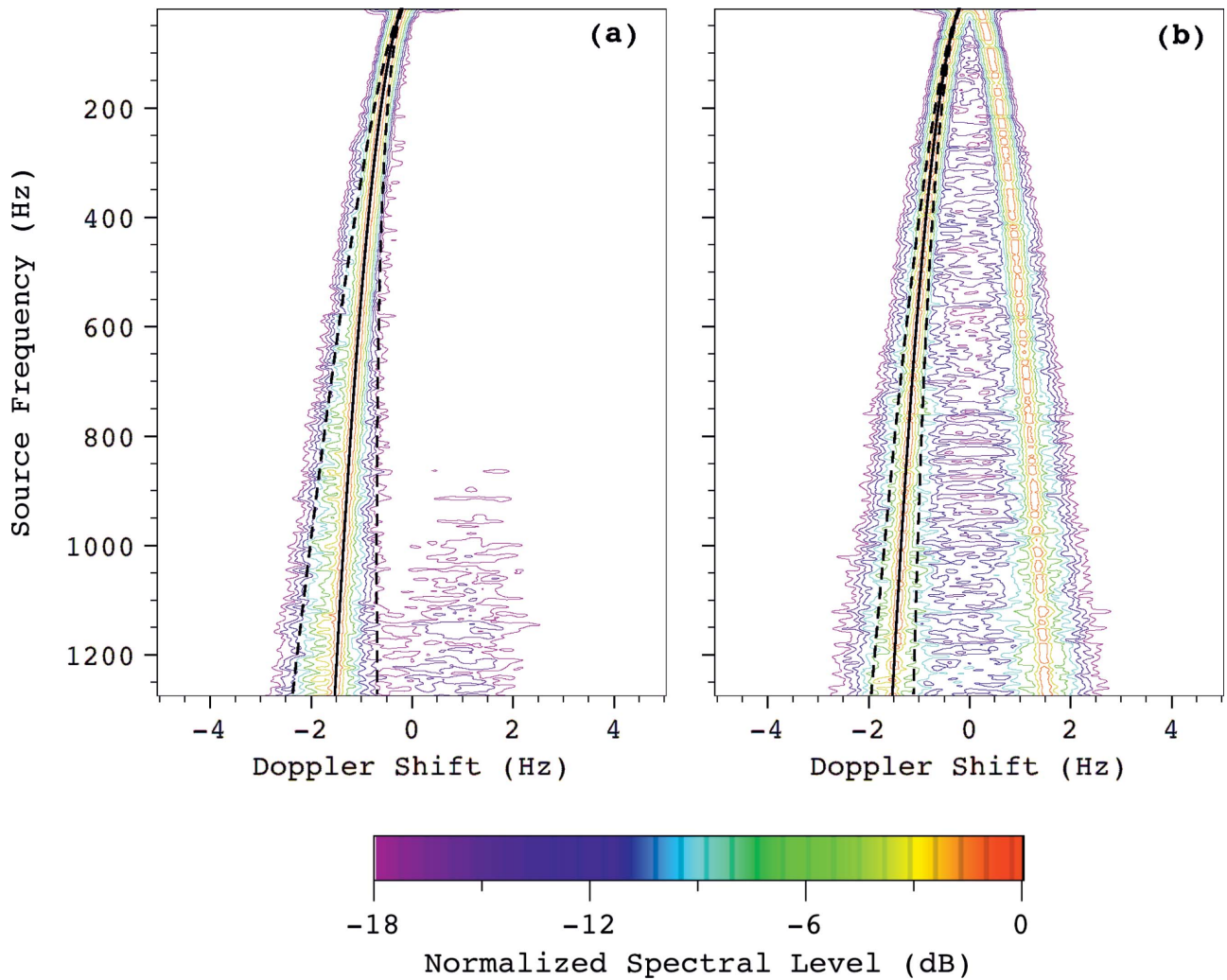
FIG. 5. Ensemble-averaged Doppler spectra calculated by the time-variant WA model for monostatic backscattering (30° grazing) from a fully developed sea due to 5-m/s winds that are oriented (a) windward and (b) crosswind to the scattering plane.

and for each source frequency, the peak level of the Doppler spectrum was set to 0 dB and all other spectral levels, occurring at other frequency shifts but having the same source frequency, were scaled accordingly. The purpose of this normalization procedure was to make more identifiable the spectral peak for each source frequency. In each example shown, the location of the expected Doppler shift from first-order perturbation theory is plotted as a function of source frequency. To augment the analysis, spectral half-width predictions based on the work of Fuks[16] are also plotted. These predictions are based on a two-scale scattering theory and purport to include broadening due to motion of the ripple and large-scale wave components. For the assumed scattering geometry and Pierson–Moskowitz wave spectrum, the spectral half-width (Δ) given by Fuks and defined as the standard deviation of the instantaneous frequency fluctuation, reduces to

$$\Delta = 1.24\left(\frac{f}{2\pi c}\right)(\sin^2 \psi + \cos^2 \psi \cos^2 \phi)U_{10}. \qquad (21)$$

In the first pair of results, the Doppler spectra for windward and crosswind azimuthal orientations are calculated for seas due to winds of 5 m/s and plotted in Figs. 5(a) and 5(b).

Superimposed on each of these plots is a solid line identifying the location of the spectral peak as predicted by first-order perturbation theory. Also drawn on each plot is a pair of dashed lines which indicate the spectral half-width according to the predictions of Fuks. This first example is the case of low wind speed and, consequently, a low to moderate surface variance ($\sigma_5^2 = 0.0248$ m^2). As expected in this situation, first-order perturbation theory predicts the same location for the spectral peak as does the time-variant WA model over the entire range of source frequencies considered (25–1260 Hz). Note that, at the maximum received frequency shown in this plot, the perturbative expansion parameter, $k\sigma_5 < 1$. According to first-order perturbation theory, the azimuthal orientation of the scattering surface should not affect the location of the spectral peak. The time-variant WA results are in agreement with this prediction, and this supports the assertion that in this situation roughness at the Bragg spatial scale is the dominant source of scattering.

There are, however, differences between the two orientations that are worth noting and perhaps indicative of the onset of significant scattering associated with non-Bragg spatial scales. Specifically, both the time-variant WA model and Fuks results predict spectral widths that depend on azimuthal

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005
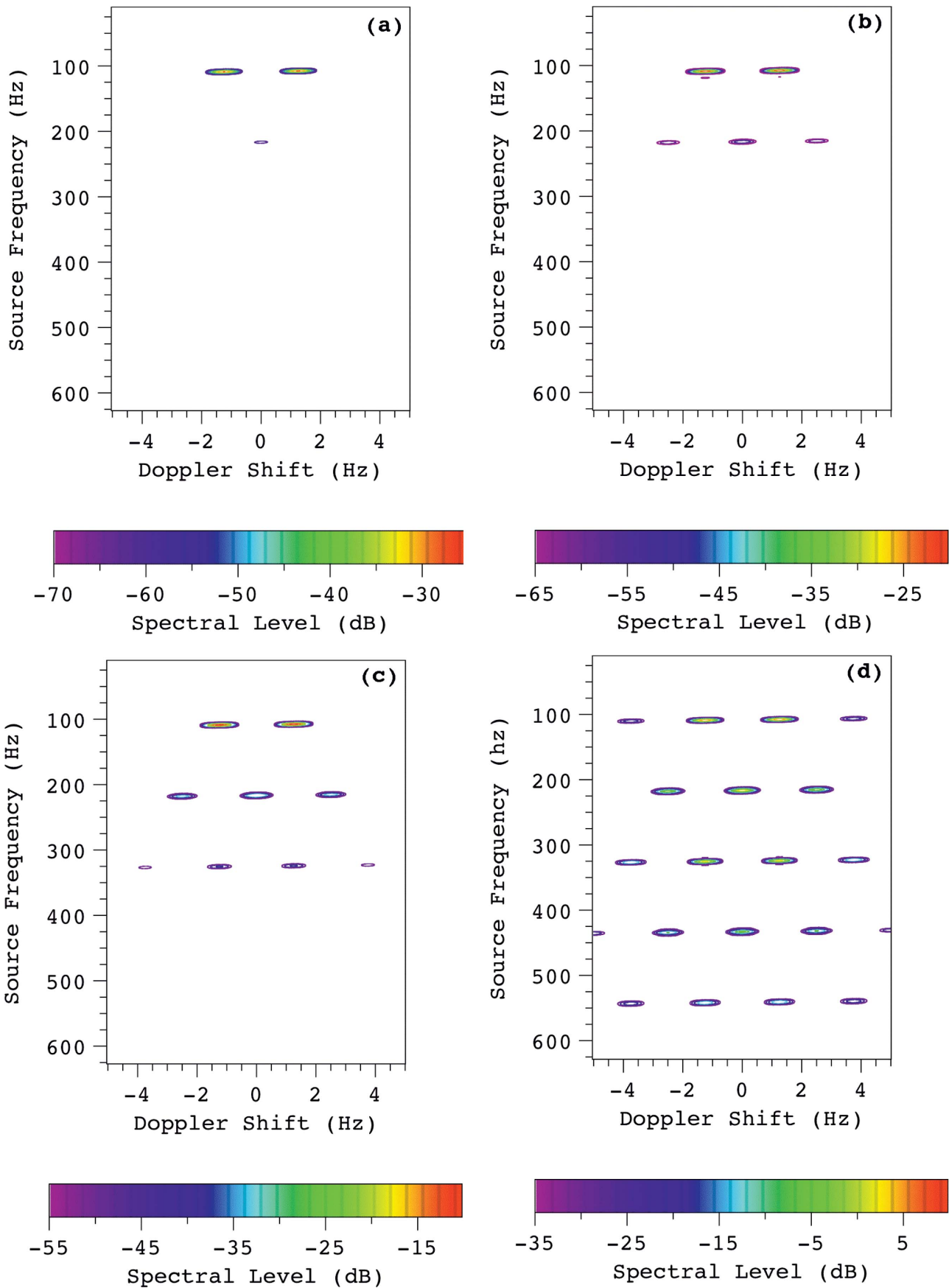
Keiffer et al.: Impulse response of moving surfaces   1293
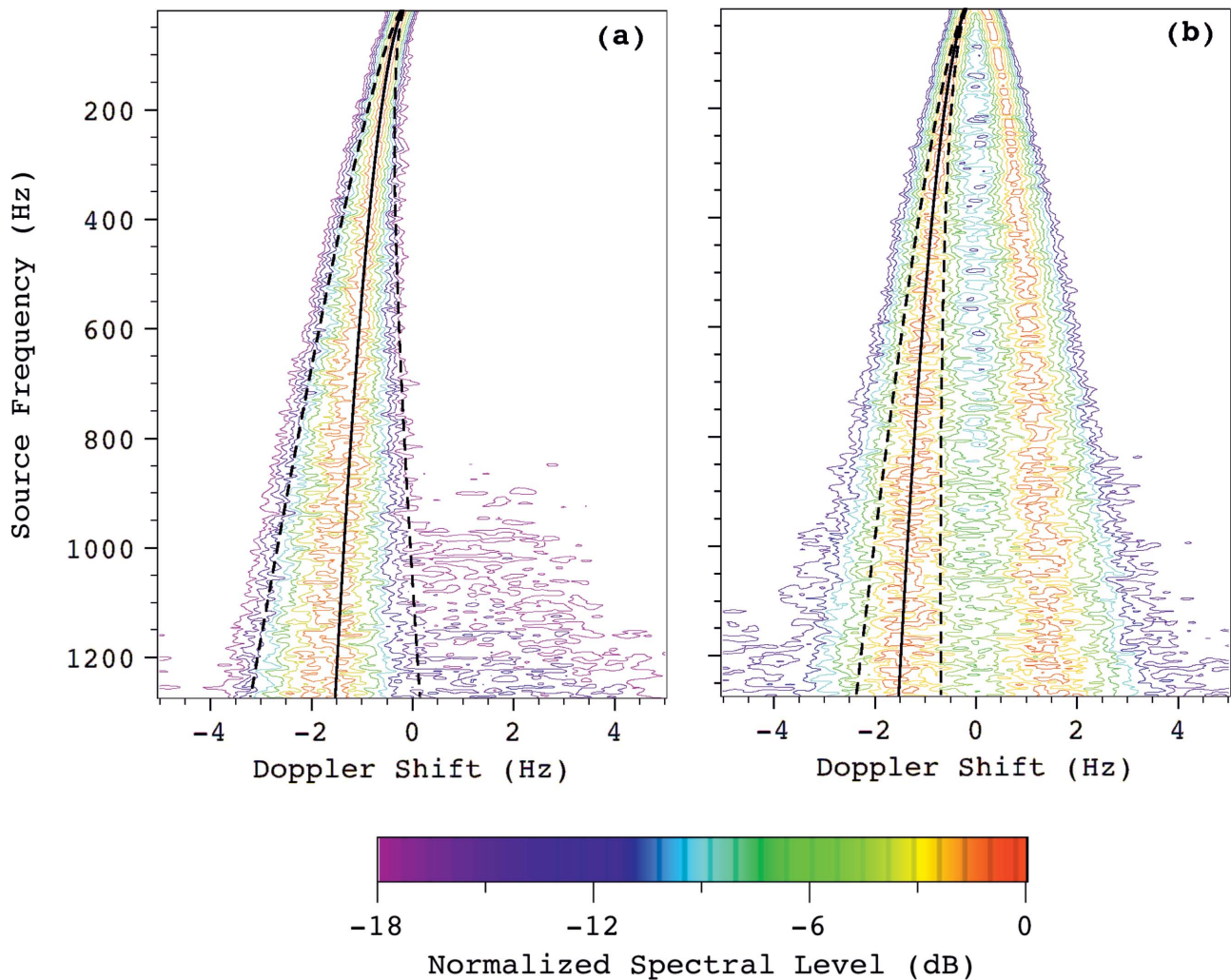
FIG. 6. Ensemble-averaged Doppler spectra calculated by the time-variant WA model for monostatic backscattering (30° grazing) from a fully developed sea due to 10-m/s winds that are oriented (a) windward and (b) crosswind to the scattering plane.

orientation. While differing in absolute terms, both of these models predict that the Doppler spectrum in the windward orientation is spread more than in the crosswind orientation. There is also a reasonable agreement between these two models for the rate at which the spectral width grows with increasing source frequency. Note that in the crosswind case the spreading is more symmetric about the expected Bragg shift. In the windward case, the spectrum spreads preferentially toward larger Doppler shifts.

Moving on to the 10-m/s example (see Fig. 6), specifically for the windward orientation, it can be seen that, for source frequencies above about 400 Hz ($k\sigma_{10} > 1.05$), the thin solid line representing the spectral peak predicted by first-order perturbation theory no longer travels down the middle of the −1-dB (red) contours of the WA data. Above about 400 Hz the WA spectral peak is consistently located at slightly greater (more negative) Doppler shifts. This is not the situation for the crosswind orientation. There, the two predictions remain in agreement over the full range of source frequencies. With regard to the spectral widths, similar observations noted for the 5-m/s surfaces apply but are somewhat more demonstrative: at any source frequency the Doppler spectrum in the windward orientation is significantly

wider than in the crosswind orientation. The spectral width predictions of Fuks remain centered on the WA results better in the crosswind case than they do for the windward orientation. Finally, the Fuks prediction for rate at which the spectral width grows as the source frequency increases seems to track the WA results reasonably well.

In the final pair of plots for this section, the backscattered Doppler spectrum from seas due to 13-m/s winds is considered. In this case the surface variance is 1.34 m², so at received frequencies greater than about 200 Hz the perturbative expansion parameter is larger than 1. Trends already present in the lower wind-speed cases are now somewhat more pronounced. In the windward oriented case [Fig. 7(a)] it can be seen that above about 300 Hz the WA prediction for the spectral peak begins to move to higher Doppler shifts than expected from first-order perturbation theory and the Bragg component. Also, the prediction of Fuks for the spectral half-width is, even more noticeably, not centered on the WA prediction. More difficult to detect by eye, but still quite evident, is the trend of the WA prediction toward lower than expected Doppler shifts in the case of the crosswind oriented scattering geometry [Fig. 7(b)]. In this geometry it is clear that the WA prediction for the spreading in the Doppler spec-

Keiffer *et al.*: Impulse response of moving surfaces
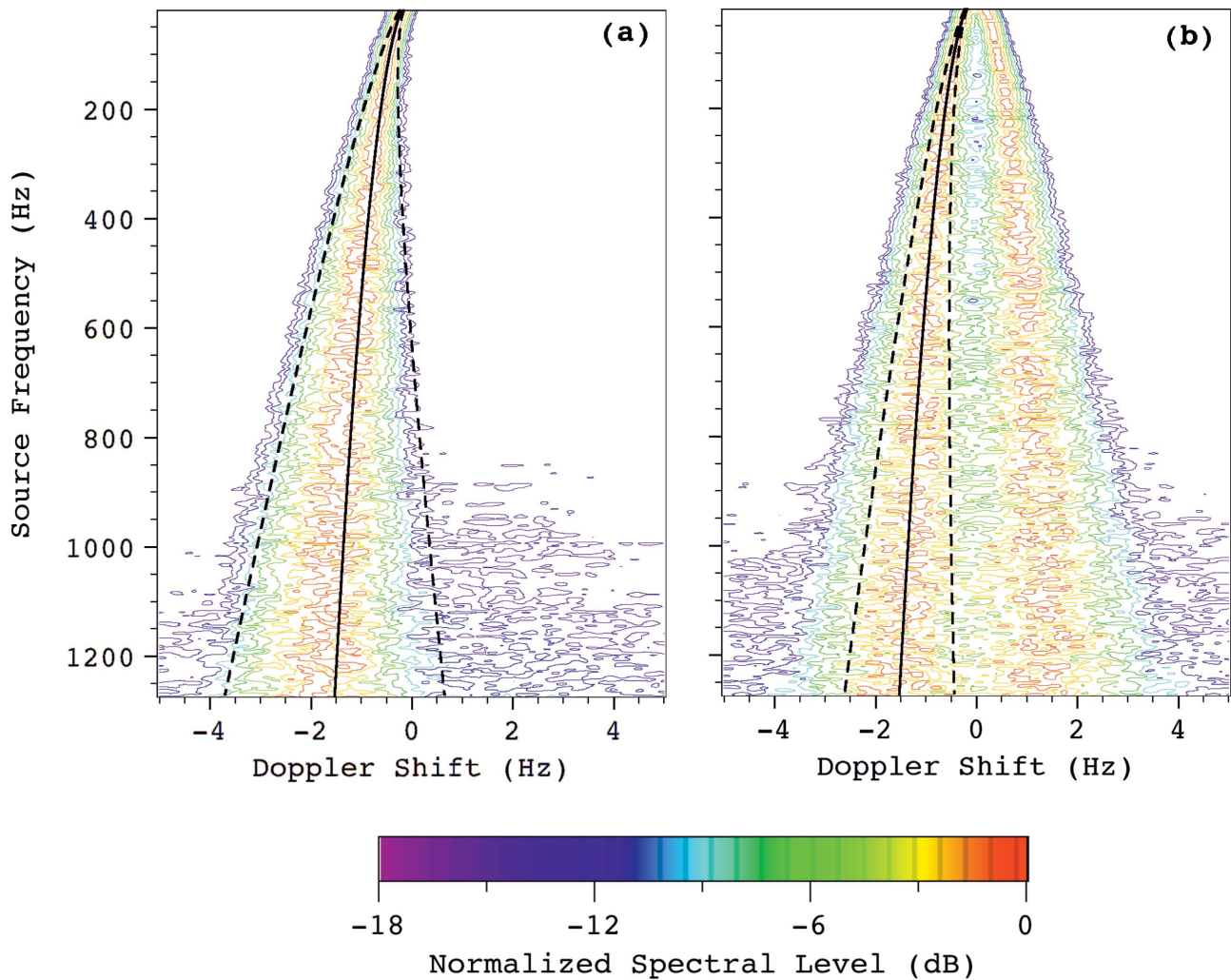
FIG. 7. Ensemble-averaged Doppler spectra calculated by the time-variant WA model for monostatic backscattering (30° grazing) from a fully developed sea due to 13-m/s winds that are oriented (a) windward and (b) crosswind to the scattering plane.

trum is preferentially towards lower Doppler shifts. At very small Doppler shifts there is evidence of significant overlap from the upshifted and downshifted components of the Doppler spectrum.

D. Identifying multiscale scattering contributions

In this last section, numerical experiments are presented that are designed to provide a closer look at the role that non-Bragg roughness scales have on the Doppler spectrum. By generating surface realizations from a sea surface roughness spectrum that is zero at wave numbers higher than $K = 1$ rad/m, one can ensure that the backscattered field that results contains no first-order perturbative contributions above about 138 Hz. That is, the scattered response can be interpreted as being due only to higher-order perturbative terms. Shown in Fig. 8 are the ensembled averaged (ten realizations) Doppler spectra for backscattering geometries oriented windward and crosswind (wind speed 5 m/s, 30° grazing angle). In the windward case [Fig. 8(a)], it can be seen that for source frequencies above 138 Hz there is sudden decay in the mainlobe of the Doppler spectrum. And, at higher frequencies, there appears to be a ghost-like replica of the lower (source) frequency part of the Doppler spectrum.

These higher-order contributions appear with Doppler shifts that are greater than the expected Bragg shift and account for the fact that in the windward orientation the spread in the Doppler spectrum is preferentially toward higher Doppler shifts. When the roughness or source frequency increases, higher-order contribution become more important and these can change the location of peak in the Doppler spectrum away from the expected Bragg shift and toward higher Doppler shifts.

Looking now at the crosswind results [Fig. 8(b)], it can be deduced that the higher-order terms contribute to the Doppler spectrum at Doppler shifts that are both higher and lower than the Bragg shift. In this orientation, there are surface wave components advancing on and receding from the source/receiver location. It can be seen that there is significant overlap near zero Doppler, and it is probably this behavior that causes the peak in the Doppler spectrum to squeeze toward lower than expected Doppler shifts.

V. CONCLUSIONS

A time-variant impulse response method has been proposed, developed, tested, and shown to provide new insights into different scattering problems involving moving surfaces.
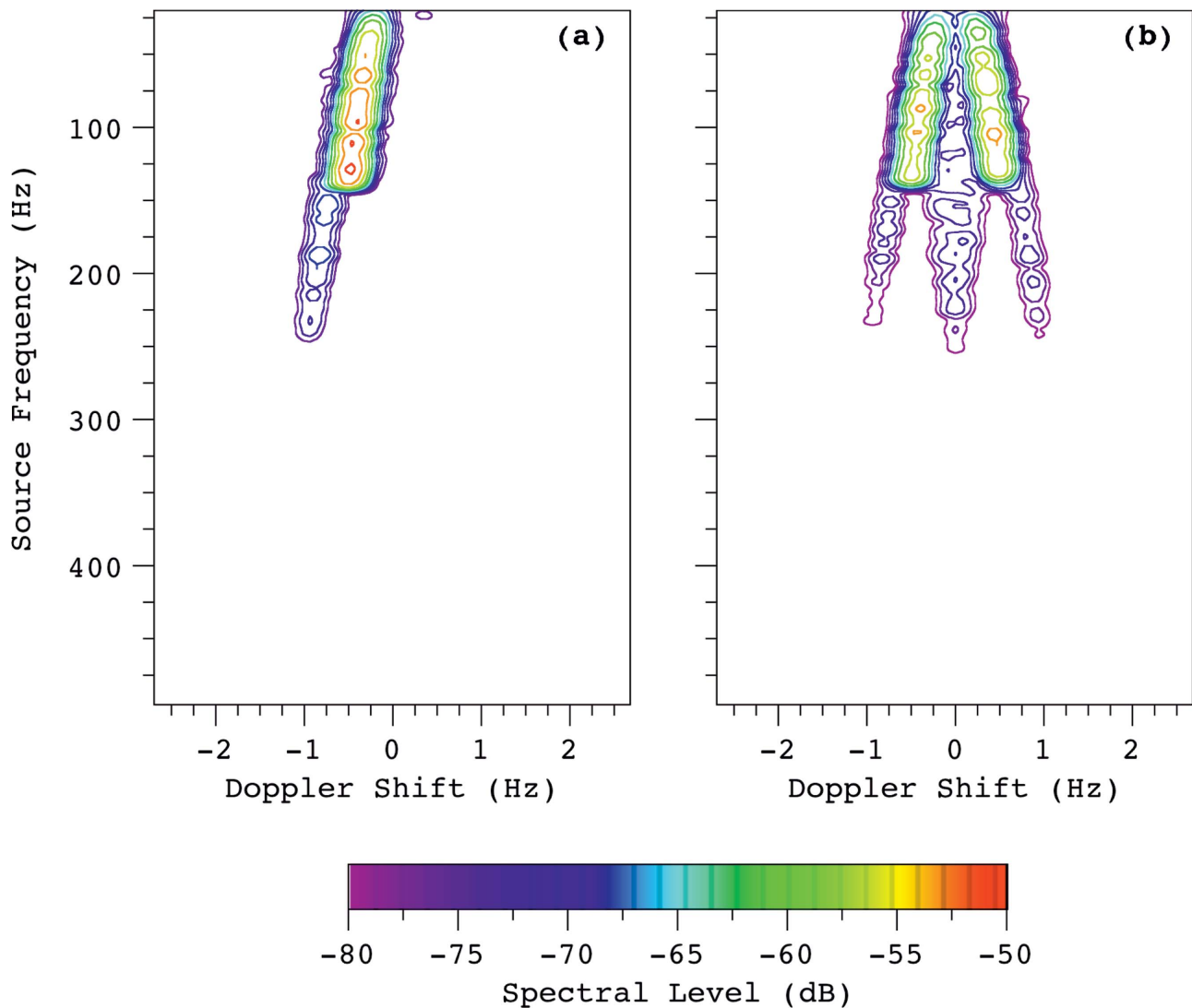
FIG. 8. Ensemble-averaged Doppler spectra calculated by the time-variant WA model for monostatic backscattering (30° grazing) from wave number-truncated, fully developed sea due to 5-m/s winds that are oriented (a) windward and (b) crosswind to the scattering plane.

The method proposed is general, conceptually straightforward, and can easily incorporate moving sources and receivers. While the time-variant impulse response method has been developed specifically in this study to extend the capability of the WA model, the framework can be applied to other scattering models including those based in the frequency domain. First, calculations involving moving periodic surfaces and moving receivers were compared with a generalized grating equation to establish that the discrete Doppler shifts expected for these types of scattering problems were being modeled correctly. Then, SWHA perturbation theory, carried out to third order in the dimensionless expansion parameter, was used as a standard to validate that this new model predicted the correct energy partition among backscattered diffracted orders. Comparisons of the SWHA results with the time-variant WA model show very good agreement in the regime where SWHA is valid, and show a reasonable departure from the SWHA results in the regime where the perturbation approach is known to be dubious. The time-variant impulse response model was applied to time-evolving sea surfaces. The earlier results of Pourkaviani and

Willemsen,[15] which indicated that under isotropic conditions the peak in the Doppler spectrum could move from the expected Bragg shift toward smaller Doppler shifts, were confirmed. It was also shown that windward-oriented backscatter geometries can lead to a Doppler spectrum that peaks at higher than expected Doppler shifts.

ACKNOWLEDGMENTS

APPENDIX: SWHA PERTURBATION THEORY FOR STANDING WAVE SURFACES

The scattering surface under consideration consists of two oppositely traveling sinusoidal waves of equal amplitude a, angular frequency $\Omega \geqslant 0$, and spatial wave number $K \geqslant 0$.

The impulse response function $h(\bar{r},t;\bar{r}_0,\tau)$, at location $\bar{r}$ and time t, is the total acoustic field due to the impulsive, point source excitation (at location $\bar{r}_0$ and emission time τ) of the homogeneous, isospeed medium together with the standing wave boundary. Here $h(\bar{r},t;\bar{r}_0,\tau)$ is a solution to the scalar wave equation

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)h(\bar{r},t;\bar{r}_0,\tau) = 0, \tag{A1}$$

subject to the Dirichlet boundary condition on the moving surface

$$h[x,z_s(x,t),t;\bar{r}_0,\tau] = 0. \tag{A2}$$

The small wave-height approximation (SWHA) perturbation solution begins with a power-series expansion of the impulse response function in powers of wave-height parameter a

$$h(t;\tau) = \sum_{p=0}^{\infty} a^p h_p(t;\tau). \tag{A3}$$

Usually, the next step is to expand the boundary conditions (A2) in a Taylor series expansion about $z=0$. This implicitly introduces the arrival time $t_0 = \tau + \sqrt{(x_0-x)^2-(y_0-y)^2-z_0^2}/c$ and the quasistatic surface, $z_s(x,t_0)$. This approach also retains in the series expansion an explicit dependence of the surface height function on t which, owing to the operational properties of derivatives of the delta function, introduces quasistatic time derivatives of the surface height function. These additional terms lead to unphysical effects[10] which can be avoided simply by making an explicit frozen surface approximation prior to the Taylor series expansion of the boundary condition. When $r_0 \gg r$ and the source is constrained to the xz plane, then $t_0 \approx \tau + (r_0/c) - [(x \sin \theta_0 \cos \phi_0)/c]$, and, in the spirit of the SWHA, the time-varying surface $z_s(x,t)$ is replaced by an equivalent quasistatic surface, $\zeta(x,\tau) = z_s(x,t_0)$. In this case of a standing wave surface

$$\zeta(x,\tau) \approx 2a\cos(Kx)\cos(\Omega t_0). \tag{A4}$$

In the frozen surface approximation, the boundary condition becomes

$$h[x,\zeta(x,\tau),t;\bar{r}_0,\tau] = 0, \tag{A5}$$

and is expressed in a Taylor series expansion about $z=0$

$$0 = h\Big|_{z=0} + \zeta(x,\tau)\left[\frac{\partial h}{\partial z}\right]_{z=0} + \frac{\zeta^2(x,\tau)}{2}\left[\frac{\partial^2 h}{\partial z^2}\right]_{z=0}$$
$$+ \frac{\zeta^3(x,\tau)}{6}\left[\frac{\partial^3 h}{\partial z^3}\right]_{z=0} + \cdots . \tag{A6}$$

Substituting Eq. (A4) into Eq. (A6), collecting like powers in surface height parameter a, and requiring that individual coefficients of a^p must vanish generates the usual system of equations that defines terms in the expansion of the impulse response function. Sequential Fourier transforms using the convention

$$B(f;v) = \int_{-\infty}^{\infty}\int_{\infty}^{\infty} h(t;\tau)\exp[i2\pi(v\tau - ft)]dt\,d\tau, \tag{A7}$$

applied to both sides of these equations yield a corresponding set of equations that defines terms in a power-series expansion of the bifrequency function. Including powers up to the third furnishes the following approximation for the bifrequency function:

$$B(f;v) \approx \frac{1}{4\pi r_0}\sum_{m=-3}^{3}\sum_{n=-3}^{3}\delta\left(v - f + m\frac{\Omega}{2\pi}\right)\exp[i(k_{ix}$$
$$+ nK)x + \gamma_{m,n}z]\sum_{p=0}^{3} a^p C_{m,n}^p, \tag{A8}$$

where $\gamma_{m,n}$ is the vertical wave number

$$\gamma_{m,n} = \left[\left(\frac{2\pi v + m\Omega}{c}\right)^2 - (k_{ix} + nK)^2\right]^{1/2}, \tag{A9}$$

and $k_{ix} = -(2\pi v/c)\sin\theta_0\cos\phi_0$. Now, using $k = 2\pi f/c$ as the scattered wave number, the nonzero coefficients in Eq. (A9) are

$$C_{0,0}^0 = -1, \tag{A10}$$

$$C_{1,1}^1 = C_{-1,1}^1 = C_{1,-1}^1 = C_{-1,-1}^1 = ik\cos\theta_0, \tag{A11}$$

$$C_{1,1}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l}\gamma_{0,0}(\gamma_{-1,-1} + \gamma_{1,-1} + \gamma_{-1,1} + \gamma_{1,1}) + \gamma_{-1,1}\gamma_{0,2}\\ + \gamma_{1,1}\gamma_{0,2} + \gamma_{1,-1}\gamma_{2,0} + \gamma_{1,1}\gamma_{2,0} + \gamma_{1,1}\gamma_{2,2}\\ - \frac{1}{2}(2\gamma_{-1,1}^2 + 2\gamma_{1,1}^2 + \gamma_{-1,-1}^2 + \gamma_{1,-1}^2) + \frac{3}{2}[k\cos\theta_0]^2\end{array}\right|, \tag{A12}$$

$$C_{1,-1}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l}\gamma_{0,0}(\gamma_{-1,-1} + \gamma_{1,-1} + \gamma_{-1,1} + \gamma_{1,1}) + \gamma_{-1,-1}\gamma_{0,-2}\\ + \gamma_{1,-1}\gamma_{0,-2} + \gamma_{1,-1}\gamma_{2,-2} + \gamma_{1,-1}\gamma_{2,0} + \gamma_{1,1}\gamma_{2,0}\\ - \frac{1}{2}(2\gamma_{-1,-1}^2 + 2\gamma_{1,-1}^2 + 2\gamma_{-1,1}^2 + \gamma_{1,1}^2) + \frac{3}{2}[k\cos\theta_0]^2\end{array}\right|, \tag{A13}$$

$$C_{1,3}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{-1,1}\gamma_{0,2} + \gamma_{1,1}\gamma_{0,2} + \gamma_{1,1}\gamma_{2,2} \\ -\frac{1}{2}(\gamma_{1,1}^2 + \gamma_{-1,1}^2) + \frac{1}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A14}$$

$$C_{1,-3}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{-1,-1}\gamma_{0,-2} + \gamma_{1,-1}\gamma_{0,-2} + \gamma_{1,-1}\gamma_{2,-2} \\ -\frac{1}{2}(\gamma_{-1,-1}^2 + \gamma_{1,-1}^2) + \frac{1}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A15}$$

$$C_{-1,1}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{0,0}(\gamma_{-1,-1} + \gamma_{1,-1} + \gamma_{-1,1} + \gamma_{1,1}) + \gamma_{-1,-1}\gamma_{-2,0} \\ + \gamma_{-1,1}\gamma_{-2,0} + \gamma_{-1,1}\gamma_{-2,2} + \gamma_{-1,1}\gamma_{0,2} + \gamma_{1,1}\gamma_{0,2} \\ -\frac{1}{2}(2\gamma_{1,1}^2 + 4\gamma_{-1,1}^2 + \gamma_{1,-1}^2 + 2\gamma_{-1,-1}^2) + \frac{3}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A16}$$

$$C_{-1,-1}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{0,0}(\gamma_{-1,-1} + \gamma_{1,-1} + \gamma_{-1,1} + \gamma_{1,1}) + \gamma_{-1,-1}\gamma_{-2,-2} \\ + \gamma_{-1,-1}\gamma_{-2,0} + \gamma_{-1,1}\gamma_{-2,0} + \gamma_{-1,-1}\gamma_{0,-2} + \gamma_{1,-1}\gamma_{0,-2} \\ -\frac{1}{2}(\gamma_{1,1}^2 + 2\gamma_{-1,1}^2 + 2\gamma_{1,-1}^2 + 4\gamma_{-1,-1}^2) + \frac{3}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A17}$$

$$C_{-1,3}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{-1,1}\gamma_{-2,2} + \gamma_{-1,1}\gamma_{0,2} + \gamma_{1,1}\gamma_{0,2} \\ -\frac{1}{2}(\gamma_{-1,1}^2 + \gamma_{1,1}^2) + \frac{1}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A18}$$

$$C_{-1,-3}^3 = \frac{k\cos\theta_0}{4i}\left|\begin{array}{l} \gamma_{-1,-1}\gamma_{0,-2} + \gamma_{1,-1}\gamma_{0,-2} + \gamma_{-1,-1}\gamma_{-2,-2} \\ -\frac{1}{2}(\gamma_{-1,-1}^2 + \gamma_{1,-1}^2) + \frac{1}{2}[k\cos\theta_0]^2 \end{array}\right|, \tag{A19}$$

$$C_{0,0}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,-1} + \gamma_{1,-1} + \gamma_{-1,1} + \gamma_{1,1}), \tag{A20}$$

$$C_{2,0}^2 = \frac{k\cos\theta_0}{2}(\gamma_{1,1} + \gamma_{1,-1}), \tag{A21}$$

$$C_{2,2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{1,1}), \tag{A22}$$

$$C_{0,2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,1} + \gamma_{1,1}), \tag{A23}$$

$$C_{-2,-2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,-1}), \tag{A24}$$

$$C_{0,-2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,-1} + \gamma_{1,-1}), \tag{A25}$$

$$C_{2,-2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{1,-1}), \tag{A26}$$

$$C_{-2,0}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,-1} + \gamma_{-1,1}), \tag{A27}$$

$$C_{-2,2}^2 = \frac{k\cos\theta_0}{2}(\gamma_{-1,1}). \tag{A28}$$

[1] R. S. Keiffer and J. C. Novarini, "A time domain rough surface scattering model based on wedge diffraction: Application to low-frequency backscattering from two-dimensional surfaces," J. Acoust. Soc. Am. **107**, 27–39 (2000).

[2] R. S. Keiffer, "On the validity of the wedge assemblage method for pressure-release sinusoids," J. Acoust. Soc. Am. **93**, 3158–3168 (1993).

[3] R. S. Keiffer, J. C. Novarini, and G. V. Norton, "The impulse response of an aperture: Numerical calculations within the framework of the wedge assemblage method," J. Acoust. Soc. Am. **95**, 3–12 (1993).

[4] H. Medwin and C. S. Clay, *Fundamentals of Acoustical Oceanography* (Academic, New York, 1998), Chaps. 11–12.

[5] M. A. Biot and I. Tolstoy, "Formulation of wave propagation in infinite media by normal coordinates with application to diffraction," J. Acoust. Soc. Am. **29**, 381–391 (1957).

[6] U. P. Svensson, R. I. Fred, and J. Vanderkooy, "An analytic secondary source model of edge diffraction impulse responses," J. Acoust. Soc. Am. **106**, 2331–2344 (1999).

[7] J. C. Novarini and R. S. Keiffer, "A hybrid model for the acoustic response of a two-dimensional rough surface to an impulse incident from a refracting medium," Appl. Acoust. **64**, 655–668 (2003).

[8] D. Chu, "Exact solution for a density contrast shallow-water wedge using normal coordinates," J. Acoust. Soc. Am. **87**, 2442–2450 (1990).

[9] A. M. J. Davis and R. W. Scharstein, "The complete extension of the Biot–Tolstoy solution to the density-contrast wedge with sample calculations," J. Acoust. Soc. Am. **101**, 1821–1835 (1997).

[10] R. S. Keiffer, J. C. Novarini, and R. W. Scharstein, "The role of the frozen surface approximation in small wave-height perturbation theory for moving surfaces," J. Acoust. Soc. Am. **113**, 1223–1229 (2003).

[11]L. B. Wetzel, "A time domain model for sea scatter," Radio Sci. **2**, 139–150 (1993).

[12]D. Censor, "Acoustical Doppler effect analysis—Is it a valid method?," J. Acoust. Soc. Am. **83**, 1223–1230 (1988).

[13]L. J. Ziomek, *Underwater Acoustics: A Linear Systems Theory Approach* (Academic, Orlando, FL, 1985), Chap. 2.

[14]G. Rosenhouse and N. Peled, "Dynamic simulation of sound field created by a motion of a monopole along a curved path and related physical phenomena," J. Comput. Acoust. **1**, 287–302 (1993).

[15]M. Pourkaviani and J. F. Willemsen, "Perturbative corrections to acoustic Doppler backscattering I. Ocean gravity waves," J. Acoust. Soc. Am. **90**, 426–432 (1991).

[16]I. M. Fuks, "Spectral width of signals scattered by a distributed sea surface," Sov. Phys. Acoust. **20**, 275–281 (1974).

[17]C. S. Hayek, I. W. Schurman, J. H. Sweeney, and C. A. Boyles, "Azimuthal dependence of Bragg scattering from the ocean surface," J. Acoust. Soc. Am. **105**, 2129–2141 (1999).

[18]E. R. Kanasewich, *Time Sequence Analysis in Geophysics*, 3rd ed. (The University of Alberta Press, Alberta, Canada, 1981), Appendix 5.

Time-lapse travel time change of multiply scattered acoustic waves

Carlos Pacheco
Center for Wave Phenomena, Department of Geophysics, Colorado School of Mines, Golden, Colorado 80401

Roel Snieder
Center for Wave Phenomena, Department of Geophysics, Colorado School of Mines, Golden, Colorado 80401

Existing techniques in correlation spectroscopy, such as coda wave interferometry and diffusing acoustic wave spectroscopy, determine the average motion of scatterers or change in the propagation velocity from the temporal change of multiply scattered sound. However, neither of them gives an indication of the spatial extent of the change in the medium. This study is an extension of the technique coda wave interferometry, where multiply scattered waves are used to determine the change in the wave field due to a localized perturbation in the propagation velocity. Here, the propagation of multiply scattered sound is described using the diffusion approximation, which allows the cross-correlation function of the unperturbed and perturbed wave fields to be related to the localized change in the propagation velocity. The technique is tested numerically for two-dimensional (2D) acoustic waves using synthetic seismograms calculated using finite-differences before and after a small perturbation in the propagation velocity has been introduced. Despite the relatively small size and magnitude of the change, multiple scattering greatly amplifies small perturbations, making changes in the phase or travel time of the wave field visible in the later-arriving waveforms. Potential applications of this technique include nondestructive evaluation of inhomogeneous materials and time-lapse monitoring of volcanoes and highly heterogeneous reservoirs. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000827]

I. INTRODUCTION

Most imaging techniques using scattered waves rely on the single scattering approximation. Seismic imaging[1] mostly uses primary reflected waves to obtain an image of the subsurface. In many physical problems, however, waves are strongly scattered and the single scattering approximation is not a valid model for the propagation of waves through the medium. In such cases we have to use a model that accounts for the multiple scattering of waves and the associated attenuation. The diffusion model has been used with success to characterize a wide range of wave phenomena in strongly scattering media.[2–6] In this model, wave energy transport acquires a diffusive character, e.g., wave energy is transported in a process similar to heat diffusion. In medical imaging, for example, diffusing near-infrared light has been used to image localized heterogeneities of tissue.[7]

In many practical applications the medium changes over time, and therefore the image of the medium changes. We then would like to obtain time-lapse measurements in order to monitor temporal changes in the medium. Examples of applications where detecting temporal changes may be useful include the monitoring of volcanoes, oil reservoirs, radioactive waste disposal sites, and fluidized suspensions. In reflection seismology, dynamic reservoir characterization provides optimal management of a reservoir, which leads to increased production. Time-lapse (4D) reflection seismic aims at inferring changes from the medium from changes in the seismic amplitudes and/or travel times from seismic reflection data that have been acquired at two different times. As an example, a 4D dataset recorded at Weyburn Field, Canada, has been used to infer time-lapse changes in the oil reservoir caused by a massive miscible CO_2 flood to enhance oil recovery.[8,9] The main goal of these 4D studies is to extract information about local changes in the reservoir using mainly the amplitude information.

The fine structure of strongly scattered waves can provide a wealth of new information in seismology, ultrasonics, acoustics, and other fields that study wave propagation in heterogeneous media.[10] Multiple scattering theory has been used to determine the number and the scattering strength of scatterers in motion inside a highly reflecting cavity, and more specifically for fish counting in a tank.[11,12] In biophysics and medical imaging, diffusing photons are now used to view body function and structure after it was found that photon transport within tissues is dominated by scattering rather than absorption.[7] Diffuse transmission spectroscopy[13] has been used to probe the structure of opaque materials such as colloids, foams, and sand, using multiply scattered photons.

Multiply scattered waves have also been used to study the dynamics of complex media and turbulent fluids. Diffuse light spectroscopy[7] has been used to measure the spatial variations in the absorption and scattering of large tissue volumes. Diffusing wave spectroscopy (DWS) is a technique in which multiply scattered light is used to study the dynamics

of colloidal suspensions;[14] this technique was adapted later to acoustic waves.[15] This technique estimates the average motion of the scatterers from the temporal fluctuations of multiply scattered sound. In this technique, the propagation of multiply scattered sound is described using the diffusion equation, which makes it possible to relate the temporal field fluctuations with the dynamics of the multiple scattering medium. Recently, coda waves have been used to study the temperature dependence of the seismic velocity in granite[16] and in a reverberant three-dimensional cavity[17] using a technique called *coda wave interferometry*. In this technique, multiply scattered waves are used to detect temporal changes in a medium by using the scattering medium as an interferometer. For small changes in the medium, estimates of this perturbation can be obtained from multiply scattered waves by a cross correlation in the time domain.[16] For larger changes, the decay of the correlation of the diffuse field is related to distortion and is an interesting system-dependent quantity.[17] Coda wave interferometry has also been used to determine the relative location of earthquakes for scattered waves.[18]

In none of those approaches has an attempt been made to determine the spatial extent of the change or local perturbation in the medium. When the random multiple scattering medium changes over time, changes in the diffuse field are related to changes in the medium. The sensitivity to perturbations in the medium enhances as the scattering order increases and thus becomes an interesting tool to detect the appearance of a very small defect in a multiple scattering sample.[19] In this study we develop a technique that relates the localized changes in the propagation velocity of the medium to the travel time change of the multiply scattered waves. As in diffuse wave spectroscopy, this technique relies on the diffusion approximation of the intensity in strongly scattering media. Therefore, we can model the mean travel time change of waves with propagation time t before and after a small and localized change in the propagation velocity has been introduced in the medium. This work is an extension of coda wave interferometry in the sense that it accounts for localized changes in the propagation velocity of the scattering medium. Thus, we are able to model the fluctuation in the phase of the multiply scattered wave field for a given localized time-lapse velocity perturbation of the multiple scattering medium. We assess the validity of our theory using finite-difference simulations of multiple scattering of acoustic waves in 2D media.

The paper is divided into five parts. Section II explains how the diffusion approximation is used to describe energy transport in strongly scattering media and its relation with a random walk process. Section III introduces coda wave interferometry and how it can be used to obtain an estimator for the mean travel time change of scattered waves. In Sec. IV we derive theoretically an expression for the mean travel time change of multiply scattered waves by using the analogy between a diffusion process and random walks. The main result of this work is the expression that relates the mean travel time change of multiply scattered waves with the localized perturbation in the propagation velocity of the medium. In Sec. V we validate the theory with numerical experiments.
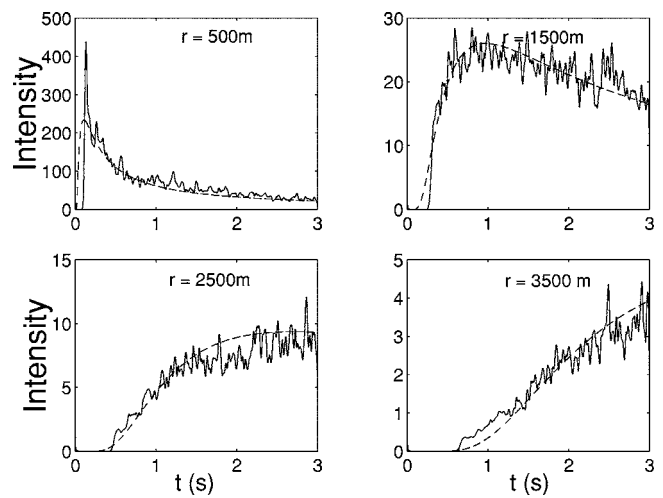
FIG. 1. Actual averaged intensities (solid line) versus best diffusion fit (dashed line) for receivers at 500, 1500, 2500, and 3500 m. The estimated value of the diffusion constant is 5.78×10^5 m^2/s.

amples using finite-difference synthetic seismograms. We found good agreement between our theory and results from the finite-difference seismograms, despite some fluctuations that are analyzed in Sec. VI. Finally, we discuss applications and limitations of the technique in Sec. VII.

II. DIFFUSION ENERGY TRANSPORT AND RANDOM WALKS

The transport of energy through a strongly scattering medium has attracted considerable attention in numerous fields of physics, such as astrophysics, optics, acoustics, solid-state physics, and heat conduction. In any of these fields, one studies a pulse of energy that propagates through the medium with an intensity $P(r,t)$. The diffusion equation describes the propagation of the average intensity in a multiple scattering medium.[20] In a two-dimensional medium of infinite extent, constant scattering properties, without intrinsic attenuation, and in the long-time limit,[21] the average intensity at **r** can be approximated by the solution of the diffusion equation

$$P(\mathbf{r},t) = \frac{1}{4\pi Dt} \exp\left[\frac{-r^2}{4Dt}\right], \qquad (1)$$

where r is the distance to the source and D is the diffusion constant. Equation (1) describes the temporal evolution of the average intensity after the waves have scattered multiple times from small-scale heterogeneities. Figure 1 shows the average intensity for receivers located 500, 1500, 2500, and 3500 m away from the source. The actual mean intensities were calculated after averaging over 100 different receiver locations after propagating a seismic wave field through a medium with random velocity fluctuations using finite differences as described in Sec. V. Figure 2 shows one of the seismograms calculated for a source–receiver distance of 2500 m. Despite the random appearance of the individual seismogram, an ensemble of such seismograms yields an average intensity that obeys the diffusion equation. The mean intensities from the finite-difference simulations are shown as solid curves, whereas the best diffu-
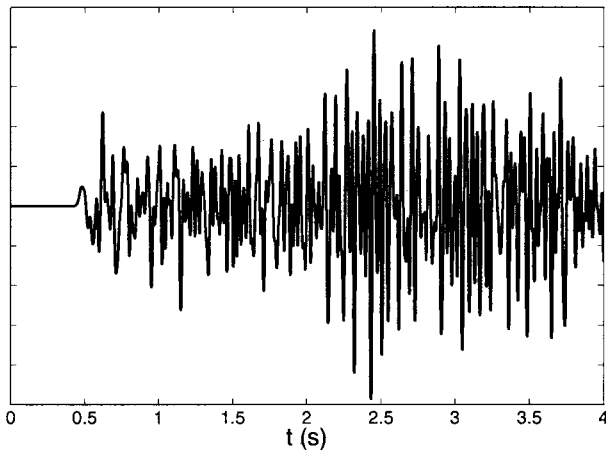
FIG. 2. Example of a finite-difference seismogram for a source–receiver distance of 2500 m after propagating a finite-bandwidth pulse through a medium with random velocity fluctuations.

FIG. 4. Observed distribution of distances for the 5000 realizations of random walks (open circles) versus the diffusion curve (solid line) for $t=3$ s.

sion fits [Eq. (1)] are shown as dashed lines. Notice the good agreement between the diffusion curves and the average intensities.

Given a seismogram such as the one shown in Fig. 2, it is extremely difficult to determine if a specific trajectory gives rise to an energy impulse at a specific time t. The problem is simplified if, instead of considering one specific trajectory, we consider the average wave field obtained after summing the contributions of all possible trajectories arriving at the receiver at time t. When scattering is strong and waves follow infinitely many trajectories, wave propagation can be considered as a random walk process. This simplifies the problem because a random walk process can on average be described with the solution of the diffusion equation. This provides us with the advantage of a simple physical picture where the diffusively scattered wave field is represented as the sum of partial waves traveling along various diffuse paths.[22]

The multiply scattered waves travel through the medium in a random walk process that is characterized in n-D (where n is the number of dimensions) by the transport mean free path l^*, the energy velocity v_e, and the diffusion coefficient $D=v_e l^*/n$.[15] Figure 3 shows three realizations of random

walk in 2D calculated numerically using $D=3.45 \times 10^5$ m²/s and $v_e=5300$ m/s. The random walks were calculated until a maximum elapsed time $t=3$ s. If we calculate many realizations, the distribution of distances r traveled from the origin at a given time t is given by Eq. (1). This is shown in Fig. 4, were we show the distribution of distances traveled for the random walk particles for 5000 different realizations of random walks starting from the origin with total elapsed time $t=3$ s. In Sec. IV we use the random walk approximation to multiple scattering to derive an expression for the mean travel time of waves in a strongly scattering medium.

III. CODA WAVE INTERFEROMETRY

When a strongly scattering medium changes, the speckle pattern of multiply scattered waves changes, which reflects the changes that occur in the interference of waves traveling with different scattering paths through the sample. Multiply scattered waves are useful in such situations, because they are increasingly sensitive with time to the perturbations in scatterer locations or perturbations in the propagation velocity of the medium. This increased sensitivity is due to the fact that waves bounce more often among scatterers as time increases, and as a result, small changes in the medium are amplified through multiple scattering.

For a small perturbation in the propagation velocity, an estimate of this perturbation can be obtained by a time-windowed cross correlation of the unperturbed and perturbed scattered waves.[23] The unperturbed wave field $u_{\text{unp}}(t)$ can be written as a summation of waves over all possible trajectories T[24]

$$u_{\text{unp}}(t) = \sum_T A_T(t), \qquad (2)$$

where a trajectory T is defined by the sequence of scatterers that a particular multiple scattering wave encounters, and $A_T(t)$ is the corresponding waveform. For diffusive wave propagation the trajectories T can be thought of as a collection of random walks.

FIG. 3. Three realizations of the random walk characterized by $D=3.45 \times 10^5$ m²/s and $v_e=5300$ m/s.

C. Pacheco and R. Snieder: Monitoring with multiply scattered waves

When we introduce a small perturbation of the propagation velocity, the dominant effect on the multiple scattering waveform is a travel time perturbation τ_T of the wave that travels along the trajectory T

$$u_{\text{per}}(t) = \sum_T A_T(t - \tau_T). \tag{3}$$

We can characterize the change in the wave field using the time-windowed cross-correlation function[23]

$$C_{\text{up}}^{(t,t_w)}(t_s) = \int_{t-t_w}^{t+t_w} u_{\text{unp}}(t')u_{\text{per}}(t' + t_s)dt', \tag{4}$$

where t denotes the center of a time window of length $2t_w$, and t_s is the lag time for the correlation. When Eqs. (2) and (3) are inserted into Eq. (4), double sums $\Sigma_{TT'}$ appear. We assume that in these double sums, the cross terms related to different trajectories $(T \neq T')$ are incoherent and average out to zero when the mean of the source signal vanishes. The contribution of the cross terms is estimated in Ref. 25. A dimensionless measure of the change of the wave field is given by the time-windowed correlation coefficient, which is given by

$$R^{(t,t_w)}(t_s) = \frac{\int_{t-t_w}^{t+t_w} u_{\text{unp}}(t')u_{\text{per}}(t' + t_s)dt'}{\left(\int_{t-t_w}^{t+t_w} u_{\text{unp}}^2(t')dt' \int_{t-t_w}^{t+t_w} u_{\text{per}}^2(t')dt'\right)^{1/2}}. \tag{5}$$

For time shifts t_s much smaller than the dominant period of the wave, a second-order Taylor expansion of the $A_T(t' + t_s - \tau_T)$ in τ_T gives[23]

$$R^{(t,t_w)}(t_s) = 1 - \frac{1}{2}\bar{\omega}^2\langle(\tau - t_s)^2\rangle_{(t,t_w)}, \tag{6}$$

where $\bar{\omega}$ is the dominant frequency of the wave. In this expression $\langle \ldots \rangle_{(t,t_w)}$ denotes the average over all waves that arrive in the time window $(t-t_w,t+t_w)$ with a weight factor that is equal to the intensity of the waves.[23] Thus, averages are taken with a weight factor that is given by the intensity of each wave. This means that in this work, the average travel time change $\langle\tau\rangle_{(t,t_w)}$ is given by an average of the travel time change of individual waves with different trajectories T arriving on the time window $(t-t_w,t+t_w)$, i.e.,

$$\langle\tau\rangle_{(t,t_w)} = \frac{\sum_T w_T(t,t_w)\tau_T(t,t_w)}{\sum_T w_T(t,t_w)}, \tag{7}$$

where the weighting factor $w_T(t,t_w)$ is given by the intensity for the diffuse waves arriving on the time window $(t-t_w,t+t_w)$.

IV. MEAN TRAVEL TIME CHANGE IN THE DIFFUSION REGIME

A. Random walk probabilities

Each random walk from the source to the receiver has an associated probability which depends on the diffusion of the intensities in the strongly scattering medium. Recalling Eq. (7), and considering waves arriving on the time window $(t$
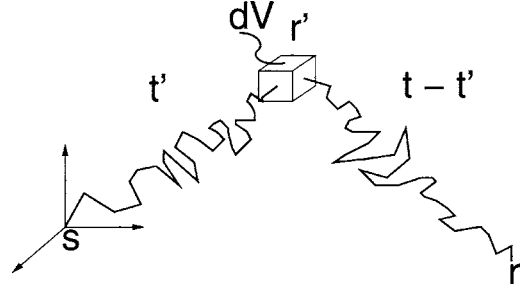
FIG. 5. A random walk particle going from source $\mathbf{s}$ at the origin to volume element dV at $\mathbf{r}'$ on time t', and then to receiver at $\mathbf{r}$ on time $t-t'$.

$-t_w,t+t_w)$, the average or mean travel time change $\langle\tau\rangle$ is a weighted average of the travel time changes of the individual trajectories T, i.e.,

$$\langle\tau\rangle = \frac{\sum_T I_T \tau_T}{\sum_T I_T}, \tag{8}$$

where I_T is the intensity or probability associated with the trajectory T, which can be calculated using the solution to the diffusion equation. In the following, we show how to derive an expression for this intensity for a given source and receiver location.

Let us assume that the seismic energy transport can be described as a diffusion process. Thus, the space and time evolution of the diffusive intensity in the medium due to an intensity impulse at the origin at time $t=0$ is given by $P(\mathbf{r},t)$, as defined in Eq. (1). In multiple scattering of waves, this diffusive energy corresponds to the ensemble averaged or mean intensity $\langle I\rangle$.

We can interpret $P(\mathbf{r},t)$ in a different way, regarding diffusion as a random walk process. In random walk theory, the product $P \times dV$ represents the probability of a particle on a random walk of visiting a volume element dV at location $\mathbf{r}$ at a given time t.[26] If, at time $t=0$, a normalized intensity impulse is generated at the source, the total energy within some region V at some later time is given by the integral

$$W(V,t) = \int_V P(r,t)dV(r). \tag{9}$$

Integration over all space gives the total energy of the system, which by the normalization is $W(t)=1$. The quantity $W(V,t)$ is equal to the probability of a particle on a random walk of visiting the volume region V at a time t.

We now consider the probability that a random walk particle leaves a source at $\mathbf{s}$ at time $t=0$, visits a volume element dV at $\mathbf{r}'$ at time t', and arrives at $\mathbf{r}$ at time t as depicted in Fig. 5. Assuming that the two paths, from source to the volume element, and from the volume element to the receiver, are independent, this probability is equal to the product of two probabilities: the probability of the particle of going from $\mathbf{s}$ to $\mathbf{r}'$ in a time t', and the probability of going from $\mathbf{r}'$ to $\mathbf{r}$ in a time $t-t'$, i.e.,

$$P(\mathbf{r}',\mathbf{r},\mathbf{s},t',t) = P(\mathbf{r}',\mathbf{s},t')P(\mathbf{r},\mathbf{r}',t-t'), \qquad (10)$$

where $P(\mathbf{r}',\mathbf{r},\mathbf{s},t',t)$ stands for the probability of all the trajectories visiting the volume element at location $\mathbf{r}'$ at time t' given that the walk started at the source location $\mathbf{s}$ and ended at the receiver location $\mathbf{r}$ at time t.

The probability of a particle to travel from the source to the receiver is given by the solution to the diffusion equation for homogeneous medium given in Eq. (1). This solution also gives the time-dependent intensity $P(\mathbf{r},\mathbf{s},t)$ at the receiver location due to a unit intensity impulse at the source. This intensity is obtained by summing the contributions to the intensity of all waves traveling with all possible diffuse trajectories going from the source to the receiver, i.e., integrating Eq. (10) over all possible volume elements $dV(\mathbf{r}')$

$$P(\mathbf{r},\mathbf{s},t) = \int_V P(\mathbf{r}',\mathbf{r},\mathbf{s},t',t)dV(\mathbf{r}')$$

$$= \int_V P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',\mathbf{s},t')dV(\mathbf{r}'). \qquad (11)$$

The integrand $P(\mathbf{r}',\mathbf{r},\mathbf{s},t',t)$ contains the contributions to the intensity of all paths which are initiated at the source location $\mathbf{s}$, visit the location $\mathbf{r}'$, and end at the receiver located at $\mathbf{r}$. By integrating over all the volume where scattering occurs, we are summing the contributions over all possible trajectories from the source to the receiver. Note that in Eq. (11) no integration over t' is performed. Equation (11) is a restatement of the Chapman–Kolmogorov equation[26]

$$\int_V P(\mathbf{r},\mathbf{r}',t)P(\mathbf{r}',\mathbf{s},t')dV(\mathbf{r}') = P(\mathbf{r},\mathbf{s},t+t'), \qquad (12)$$

which states that a random walk process starting at $t=0$ at location $\mathbf{s}$ reaches $\mathbf{r}$ at t via one of the possible values $\mathbf{r}'$ at a intermediate time t'.

The quantity $P(\mathbf{r}',\mathbf{s},t')$ in Eq. (11) is the diffuse intensity at a time t' at $\mathbf{r}'$ due to an impulse source at $\mathbf{s}$ activated at time $t=0$, and $P(\mathbf{r},\mathbf{r}',t-t')$ is the intensity at $\mathbf{r}$ at time t due to an impulse source at $\mathbf{r}'$ on a time $t-t'$. Equation (11) holds for all times $0 < t' < t$.[26] If we integrate both sides of Eq. (11) over time t' over the interval $0 < t' < t$, we obtain

$$tP(\mathbf{r},\mathbf{s},t) = \int_V \int_0^t P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',\mathbf{s},t')dt'\, dV(\mathbf{r}'),$$
$$(13)$$

where we can identify the time integral $\int_0^t P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',\mathbf{s},t')dt'$ as the time convolution of the two intensities: one at location $\mathbf{r}'$ due to an impulse intensity at $\mathbf{s}$ and the other at location $\mathbf{r}$ due to an impulse intensity at $\mathbf{r}'$. If we divide both sides of Eq. (13) by $P(\mathbf{r},\mathbf{s},t)$, we arrive at the following integral representation for the travel time of the diffuse wave field:

$$t = \frac{1}{P(\mathbf{r},\mathbf{s},t)} \int_V \int_0^t P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',\mathbf{s},t')dt'\, dV(\mathbf{r}').$$
$$(14)$$
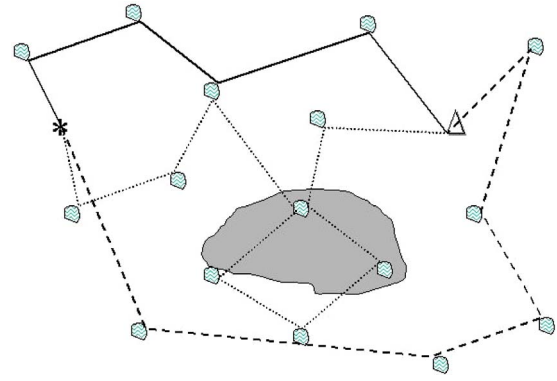
FIG. 6. (Color online) Sensitivity kernel K in 2D for source (asterisk) and receiver (triangle) separated 3000 m and $t=2$ s. The diffusion constant is 5.8×10^5 m^2/s.

We have obtained in Eq. (14) a representation for the time t at $\mathbf{r}$ in terms of the diffuse intensity due to an impulse source at $\mathbf{s}$. The time t corresponds to the time of propagation of the diffuse intensity. If we define the kernel $K(\mathbf{r}',\mathbf{r},\mathbf{s},t)$ as

$$K(\mathbf{r}',\mathbf{r},\mathbf{s},t) = \frac{1}{P(\mathbf{r},\mathbf{s},t)} \int_0^t P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',\mathbf{s},t')dt',$$
$$(15)$$

we can express the travel time t as the following volume integral:

$$t = \int_V K(\mathbf{r}',\mathbf{r},\mathbf{s},t)dV(\mathbf{r}'), \qquad (16)$$

where $K(\mathbf{r}',\mathbf{r},\mathbf{s},t)$ represents the time of flight distribution of multiply scattered waves started at the source at location $\mathbf{s}$, visiting location $\mathbf{r}'$ and detected on a receiver at location $\mathbf{r}$. Figure 6 shows a plot of the sensitivity kernel calculated for a source and receiver separated 3000 m, $t=1$ s, and $D=5.8 \times 10^5$ m^2/s. Note that the kernel vanishes outside the area of an ellipse with foci at the source and receiver location.

B. Integral representation for the mean travel time change of the diffuse wave field

When the scatterers in a multiple scattering material move, or when the propagation velocity of the medium changes, the diffuse wave field changes. In diffusing acoustic wave spectroscopy,[15] the fluctuations of the multiply scattered wave field are measured and analyzed to provide a sensitive technique for probing the dynamics of the scatterers. Here, we use a similar approach, considering the phase changes in the diffuse wave field which arise due to a spatially localized change in the propagation velocity.

We perturb the slowness in a finite region of the medium as shown in Fig. 7. We work under the assumption that the perturbation is weak so that the scattering coefficient does not change, and the waveform for each scattering path stays approximately the same. Also, the scattering paths remain unchanged so that the only difference between the unperturbed and the perturbed field is a small travel time perturbation, τ_T. If the mean slowness of the medium is denoted by
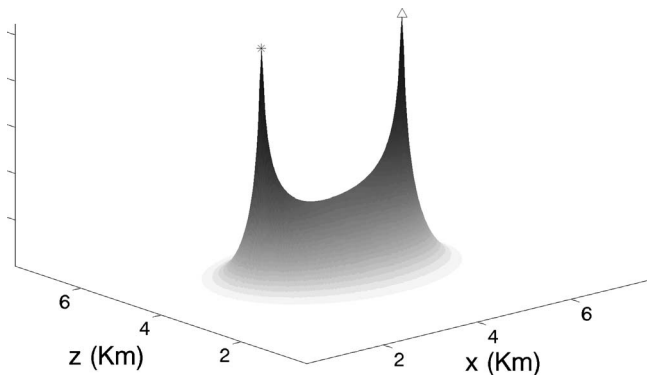
FIG. 7. Multiple scattering paths and a localized perturbation. After the perturbation the paths remain the same. The source location is represented by an asterisk and the receiver by a triangle.

s, we calculate the mean length $\langle L(t) \rangle$ of the multiple scattering paths at time t by dividing Eq. (16) by s

$$\langle L(t) \rangle = \int_V \frac{1}{s} K(\mathbf{r}', t) dV(\mathbf{r}'), \qquad (17)$$

where we have placed s inside the integral since it is assumed to be independent of location. For notation simplicity we have removed the explicit dependence of K on the source and receiver location. A perturbation in the slowness gives rise to a perturbation in the travel time or phase of the wave field, i.e.,

$$t + \delta_t = \int_V K(\mathbf{r}', t) \frac{(s + \delta_s)}{s} (\mathbf{r}') dV(\mathbf{r}'), \qquad (18)$$

where δ_s is the perturbation in the background slowness. Since in the integral in Eq. (18) we are averaging the contribution to the travel time perturbation of all volume elements, δ_t becomes the average or mean travel time change that we define as $\langle \tau(t) \rangle$. We have assumed that the perturbation in the slowness is small ($\delta s / s \ll 1$) so that the travel time perturbation depends only linearly on the slowness perturbation. The average or mean travel time change for the multiple scattering paths with path length L is then

$$\langle \tau(t) \rangle = \int_V K(\mathbf{r}', t) \frac{\delta s}{s} (\mathbf{r}') dV(\mathbf{r}'), \qquad (19)$$

where $\langle \tau(t) \rangle$ is the mean travel time change of the multiply scattering waves with travel time t and path length $\langle L(t) \rangle$ due to the relative slowness perturbation $\delta s / s$. Note that this average is weighted by the intensity, since the integration kernel $K(\mathbf{r}', t)$ represents the intensity of diffuse trajectories going through location $\mathbf{r}'$ with total travel time t, as needed in the averaging in Eq. (8).

To calculate the mean travel time change for a particular source and receiver configuration, we need to integrate the kernel K weighted by the slowness perturbation $\delta s / s$ over the volume where scattering takes place. The kernel K is given by the time convolution in Eq. (15). For a source and receiver at different locations, this time convolution does in general not have an analytical solution so it must be evaluated numerically. For the special case of coincident source

and receiver, an analytical solution for this convolution can be obtained. In the Appendix we calculate the kernel K for coincident source and receiver for two- and three-dimensional media. In 3D, the kernel K is given by

$$K_{3D}(\mathbf{r}, t) = \frac{1}{2\pi D r} \exp\left[\frac{-r^2}{Dt} \right]. \qquad (20)$$

From Eq. (20) we see that the main contributions to the travel time change come from paths located close to the coincident source and receiver location. Also, we can see that for a fixed distance r the integration kernel K increases with time t. In multiple scattering, the effective distance traveled by diffuse waves is proportional to the square root of distance ($r \sim \sqrt{4Dt}$). For smaller value of the diffusion coefficient, the stronger the scattering is and the smaller the effective distance traveled by diffuse waves (wave paths become more localized around the source and receiver location). Thus, with increased time, diffuse waves sample the same region multiple times and consequently the travel time change increases with time.

For the special case of 2D wave propagation the integration kernel K is for coincident source and receiver (see the Appendix for derivation)

$$K_{2D}(\mathbf{r}, t) = \frac{1}{2\pi D} \exp\left[\frac{-r^2}{2Dt} \right] K_0\left[\frac{r^2}{2Dt} \right], \qquad (21)$$

where K_0 is the modified Bessel function of the second kind. If we insert Eq. (21) into Eq. (19) and integrate over area instead of volume, we obtain

$$\langle \tau(t) \rangle = \frac{1}{2\pi D} \int_A \exp\left[\frac{-r^2}{2Dt} \right] K_0\left[\frac{r^2}{2Dt} \right] \frac{\delta s}{s} (\mathbf{r}) dA(\mathbf{r}), \qquad (22)$$

where r is the distance from the slowness perturbation $\delta s / s (\mathbf{r})$ element to the coincident source and receiver location. In Eq. (22), we have obtained an explicit expression relating the mean travel time change $\langle \tau \rangle$ of the diffuse waves to the localized slowness perturbation $[\delta s / s(\mathbf{r})]$ in a multiple scattering medium for coincident source and receiver. In general, for a given perturbation in slowness, we can predict the mean travel time change $\langle \tau \rangle$ for any source and receiver configuration by numerically calculating the convolution in Eq. (19).

V. TRAVEL TIME CHANGE FROM FINITE-DIFFERENCE SYNTHETICS

To generate synthetic seismograms for our study of multiple scattering, we use a fourth-order 2D acoustic finite-difference code that propagates a finite-duration pulse through a specified velocity field. Following Ref. 27, we model the 2D velocity field as a constant-background model with added random velocity fluctuations that constitute the scatterers (see Fig. 8). The total velocity field can be decomposed as

$$v(\mathbf{r}) = v_0 + v_r(\mathbf{r}), \qquad (23)$$

where v_0 is the background velocity and v_r are the random velocity fluctuations. The velocity fluctuations are characterized by a Gaussian autocorrelation function with correlation
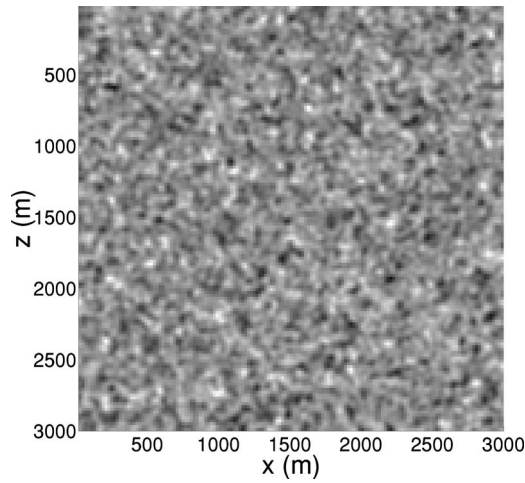
FIG. 8. 2D velocity model with random velocity fluctuations added to it. The correlation length of the velocity fluctuations is 40 m.

distance a, with zero mean and standard deviation σ (see Fig. 8 for a realization of the velocity model). The autocorrelation function of the velocity fluctuations v_r has the form

$$\langle v_r(\mathbf{r}')v_r(\mathbf{r}+\mathbf{r}')\rangle = \sigma^2 \exp\left[\frac{-r^2}{a^2}\right]. \tag{24}$$

The synthetic seismograms were created by transmitting a bandlimited pulse with a dominant wavelength of 240 m. The grid size used was 20 m and the autocorrelation a length was set to 40 m, which is much smaller than the wavelength. The mean velocity v_0 is equal to 6000 m/s. To ensure strong scattering, we created a velocity field with a relative standard deviation of 25% about the mean velocity value. To test the validity of the diffusion approximation for our numerical model, we performed a numerical experiment where we placed a source in the middle of the model and recorded the seismograms on an array of receivers around the source. Figure 9 shows 100 synthetic seismograms computed at a distance of 250 m (left) and 3000 m (right) from the source. Note the strength of the multiple scattered arrivals after the highly attenuated ballistic arrival, especially for a distance of 3000 m from the source. The average intensity (which was

FIG. 10. Slowness perturbation added to the initial velocity model to create the perturbed velocity model. The side length of the square is 3000 m and the magnitude of the perturbation is $\delta s/s=0.005$. The source is shown as a star and the receivers as triangles.

shown in Fig. 1) was obtained by averaging the squared envelope of all calculated waveforms at a given distance to the source.

We test our theory with the finite-difference simulations of acoustic waves in the multiple scattering regime before and after a localized slowness perturbation has been introduced in the model. We perturb the random velocity model which represents the unperturbed medium by adding a localized slowness perturbation as shown in Fig. 10. The value of the relative slowness perturbation is $\delta s/s=0.005$.

We first analyze the unperturbed and perturbed synthetic seismograms for the receiver **R1** located on the perturbed region of our model 2000 m away from the source. Figure 11 shows both the unperturbed and perturbed seismograms for receiver **R1**. The waveforms consist of diffuse or multiply scattered waves that have followed a multitude of paths from the source to the receiver. Notice the strength of the coda waves for late times. At first sight there seems to be no substantial difference between the wave field before and after the perturbation. However, zooming at around 2 s (see Fig.

FIG. 9. Seismograms recorded at 250 m (left) and 3000 m (right) from the source.
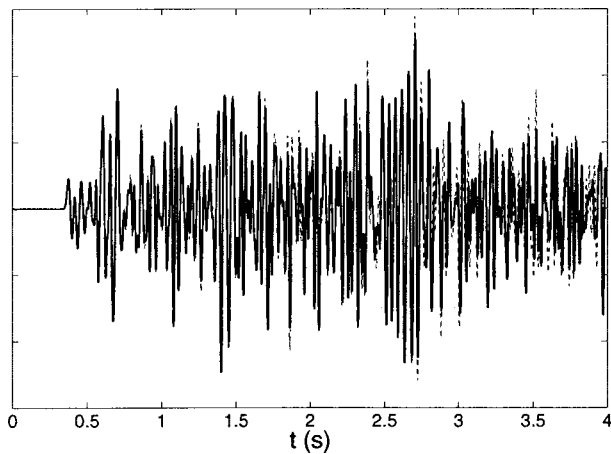
FIG. 11. Unperturbed (solid) and perturbed (dashed) synthetic seismograms recorded at the receiver **R1** located 2000 m away from the source.

FIG. 13. (Color online) Theoretical versus measured mean travel time change for receiver **R1** located 2000 m away from the source.

12) we see that the unperturbed seismogram lags in time with respect to the perturbed seismogram. More careful inspection indicates that the behavior of the time lags with travel time is systematic, i.e., the lag is increasing with travel time t.

We calculate the theoretical mean travel time change using Eq. (19) for receiver **R1** at several propagation times t and compare the result with the mean travel time change estimated from the synthetic seismograms using the time-windowed cross-correlation technique. Figure 13 shows good agreement between the theoretical and the measured mean travel time change $\langle \tau \rangle(t)$.

We also calculated the mean travel time changes for the receivers **R2** and **R3** located 5500 m away from the source. Even though the distance to the source is the same for these receivers, their locations with respect to the perturbation is different and therefore the mean travel time change is different for the two receivers. The theoretical travel time change for receiver **R2** is almost twice the travel time change for receiver **R3** (see Fig. 14). Also note that we have obtained a good agreement between the theoretical and estimated mean travel time change. This sensitivity of the mean travel time

change to the relative source and receiver locations with respect to the perturbation can be exploited in an inversion scheme to estimate the propagation velocity change from measurements of the mean travel time change at different receiver locations. Notice also that for all receivers there are fluctuations of the measured travel time change about the theoretical value. We explore the origin of these fluctuations and how to minimize them in the following.

VI. FLUCTUATIONS OF THE MEAN TRAVEL TIME CHANGE

We measured the mean travel time change from the synthetic seismograms using a time-windowed cross-correlation technique. The measured mean travel time change measured in this way is approximately an average of the time lags for the many scattering events on that time window. The fluctuations seen on the estimated mean travel time change are due to a number of factors, of which the most important is the size of the time window. We measure time lags of a continuous signal. Special care must be taken not to choose too small a time window to ensure stability of the estimation in the time-windowed cross correlation. The larger the time

FIG. 12. Zoom of the unperturbed (solid curve) and perturbed (dashed curve) synthetic seismograms recorded at the receiver **R1** located 2000 m away from the source. Note the time lags between the unperturbed and perturbed seismograms at different times.
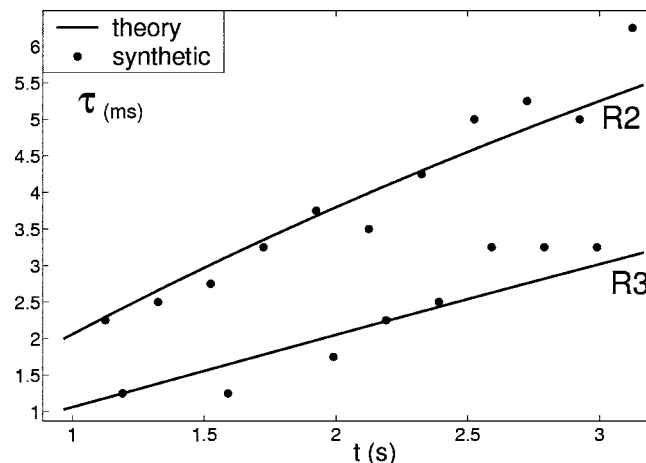
FIG. 14. Theoretical versus measured mean travel time change for receivers **R2** and **R3** located 5500 m away from the source.
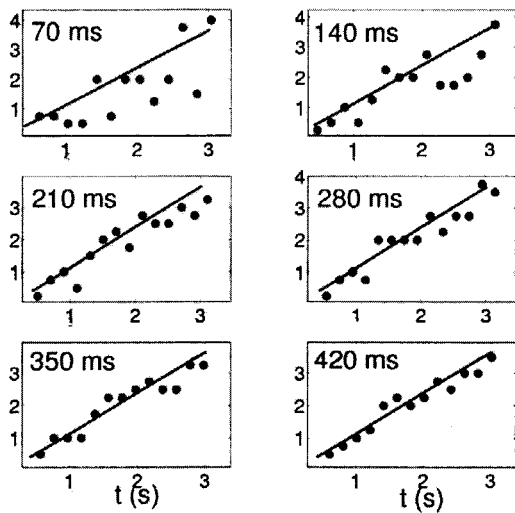
FIG. 15. Mean travel time change $\langle\tau\rangle(t)$ estimated from the synthetic seismograms for receiver **R1** using different window lengths on the cross correlation. The mean travel time change is in milliseconds. Note the reduction in the fluctuations of the measured mean travel time change (circles) around the theoretical value (solid line) for larger window lengths. The dominant period is 40 ms. The window length is shown in the upper-left corner of each plot.

$$\langle\tau(t)\rangle = \int_V K(r,t) \times \frac{\delta s}{s}(r) \times dV$$

FIG. 16. (Color online) Schematic view of the tomographic problem in matrix form. The sensitivity kernel K relates linearly the unknown parameters $\delta s/s$ with the measured traveltime change $\langle\tau\rangle(t)$. Each row in the matrix operator corresponds to a different receiver location.

window the more averaging of travel time perturbations occurs. Thus, choosing a larger window may help minimize the fluctuations of the observed mean travel time change.

We can reduce the fluctuations of the estimated mean travel time change $\langle\tau\rangle(t)$ by choosing a larger time window length $2t_w$ on the time-windowed cross correlation. Reference 25 shows that increasing the time window length decreases the magnitude of the cross terms on the average intensity $\langle I \rangle$, and therefore the fluctuations of the time lags measured in the cross correlation. Larger contributions from the cross terms to the average intensity imply larger departures from the diffusive behavior.

The decrease of the fluctuations with the increase of the time window length $2t_w$ can be appreciated in Fig. 15, where we plot the measured mean travel time change (circles) versus the theoretical mean travel time change (solid line) using different window lengths on the time-windowed cross correlation for the receiver **R1** of example 1 of the previous section. We see that the fluctuations of the measured travel time change decrease with increasing window length $2t_w$. For this synthetic example, the dominant period T_{dom} is 40 ms. The length of the time window in Fig. 15 varies from 70 to 420 ms.

VII. CONCLUSIONS

We have developed a technique that relates the mean travel time changes to the local changes in the propagation velocity of the medium. The theory is formulated by means of the diffusion approximation of the multiple scattering wave field. The sensitivity kernel $K(r,t)$ describes the dependence of the mean travel time change with the source and receiver location, the relative slowness perturbation, and the diffusion constant of the medium.

Equation (19) relates the mean travel time change $\langle\tau\rangle \times (t)$ at time t with a localized slowness perturbation $\delta s/s(\mathbf{r})$.

Note that this equation can be used as the basis of a standard linear inverse problem, as there is a linearized relation between the data (mean travel time change for different times t, and different source/receiver pairs) and the unknown parameters of the medium (localized slowness perturbation) which we want to retrieve. In the previous section we showed that the mean travel time change is a function of time and of the source and receiver locations. This dependence of the mean travel time change on the relative source and receiver location with respect to the localized slowness perturbation can be exploited in the inversion scheme for the spatial location of the slowness change.

We have developed our technique for homogeneous media. In practice, detectors will most likely need to be placed on the surface of the medium, so that boundary conditions will have to be taken into account. Also, the medium may contain layers, so we will have to account for heterogeneity as well. The problem of analyzing the precise effect of a contrast of wave velocities and/or scattering strengths between them will arise. Under the assumption of an almost isotropic scattering field[28] we determined the boundary conditions for the diffusion and radiative transfer equations for layered media. Undoubtedly all this will make our kernels more complicated.

Future work includes developing an inversion scheme to obtain the shape and magnitude of the slowness perturbation from the measured travel time changes at different receiver locations. This is similar to the transmission tomography problem, with the added complication that we are not only taking first or minimum time arrival but all multiple arrivals due to the interference of multiply scattered waves. The problem is simplified due to the fact that we are not using explicitly the paths as in transmission tomography but rather using the average wave field obtained with the diffusion approximation. Instead of calculating the travel time change with infinitely many line integrals, we calculate the mean travel time change using a much simpler volume integral that

C. Pacheco and R. Snieder: Monitoring with multiply scattered waves

contains the contributions of all possible paths. For a given time t and for a fixed source location we can set up the inverse problem in matrix form using different receivers as is shown in Fig. 16. The matrix operator (calculated from the time of flight distribution K) which multiplies the unknown parameter vector [slowness perturbation $\delta s/s(\mathbf{r})$ as a function of position] generates the data (mean travel time change for different receiver locations). Each row in the matrix corresponds to an observation of mean travel time change for a specific source–receiver pair at a discrete number of times t.

This technique can be used to detect temporal changes in highly heterogeneous material or reservoirs given that the diffusion approximation of the intensities is valid.

ACKNOWLEDGMENTS

APPENDIX A: SENSITIVITY KERNEL K FOR COINCIDENT SOURCE AND RECEIVER

We start from the expression

$$K(\mathbf{r}',\mathbf{s},\mathbf{r},t) = \int_0^t \frac{P(\mathbf{r},\mathbf{r}',t-t')P(\mathbf{r}',s,t')}{P(\mathbf{r},t)} dt', \qquad (A1)$$

where $P(\mathbf{r},t)$ is the intensity at the receiver located at $\mathbf{r}$ due to a normalized impulse source at the origin at time $t=0$, and is given by Eq. (1). For homogeneous media it depends only on time and on the distance r between source and receiver. In 2D it is equal to

$$P(\mathbf{r},t) = \frac{1}{4\pi Dt} \exp\left[\frac{-r^2}{4Dt}\right]. \qquad (A2)$$

The time convolution is given by

$$P(|\mathbf{r}'-\mathbf{s}|) * P(|\mathbf{r}-\mathbf{r}'|) = \int_0^t P(|\mathbf{r}'-\mathbf{s}|,t')P(|\mathbf{r}-\mathbf{r}'|,t$$
$$-t')dt'. \qquad (A3)$$

Substituting Eq. (A2) into Eq. (A3) gives for coincident source and receiver $(\mathbf{r}=\mathbf{s}=\mathbf{0})$

$$P(\mathbf{r}',t) = \int_0^t \frac{\exp\left[\dfrac{-r'^2}{4Dt'}\right]}{4\pi Dt'} \frac{\exp\left[\dfrac{-r'^2}{4D(t-t')}\right]}{4\pi D(t-t')} dt'. \qquad (A4)$$

As there is symmetry around $t/2$, we can write after renaming $\mathbf{r}'$ as $\mathbf{r}$

$$P(\mathbf{r},t) = \frac{2}{(4\pi D)^2} \int_0^{t/2} \frac{\exp\left[\dfrac{-r^2}{4D}\dfrac{t}{t'(t-t')}\right] dt'}{t'(t-t')}. \qquad (A5)$$

We apply the following changes of variables: $\epsilon = 1/[t'(t-t')]$ and then $\nu = r^2t[\epsilon-(4/t^2)]$ to obtain

$$P(\mathbf{r},t) = \frac{2}{(4\pi D)^2} e^{[-r^2/t]} \int_0^{\infty} \frac{e^{-\nu}d\nu}{\sqrt{\nu^2 + \dfrac{\nu r^2}{Dt}}}. \qquad (A6)$$

This integral can be calculated identifying it as a integral of the type

$$\int_0^{\infty} \frac{e^{-px}dx}{\sqrt{x(x+a)}} = e^{ap/2}K_0\left(\frac{ap}{2}\right), \qquad (A7)$$

for $a>0$ and $p>0$. The solution to the integral in Eq. (A7) can be found in Ref. 29. Setting $p=1$ and $a=r^2/Dt$, expression (A7) becomes

$$(P*P)(\mathbf{r},t) = \frac{2}{(4\pi D)^2 t} \exp\left[\frac{-r^2}{2Dt}\right] K_0\left[\frac{r^2}{2Dt}\right], \qquad (A8)$$

where K_0 is the modified Bessel function of the second kind. Substituting Eqs. (A4) and (A2) into Eq. (A1), we arrive at the expression for the kernel $K(\mathbf{r},t)$ in two dimensions for coincident source and receiver

$$K_{2D}(\mathbf{r},t) = \frac{1}{2\pi D} \exp\left[\frac{-r^2}{2Dt}\right] K_0\left[\frac{r^2}{2Dt}\right]. \qquad (A9)$$

To obtain the sensitivity kernel K in three-dimensional media, we substitute Eq. (A2) with the Green's function for the diffusion equation in 3D

$$P(\mathbf{r},t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left[\frac{-r^2}{4Dt}\right]. \qquad (A10)$$

Solving the integral defined in Eq. (A1) using the same changes of variables as in the 2D case, we obtain the expression for the kernel $K(\mathbf{r},t)$ in three dimensions for coincident source and receiver

$$K_{3D}(\mathbf{r},t) = \frac{1}{2\pi Dr} \exp\left[\frac{-r^2}{Dt}\right]. \qquad (A11)$$

[1] J. Claerbout, *Imaging the Earth's Interior* (Blackwell, Cambridge, MA, 1985).
[2] J. P. Wesley, "Diffusion of seismic energy in the near range," J. Geophys. Res. **70**, 5099–5106 (1965).
[3] Y. Kopnichev, "The role of multiple scattering in the formation of a seismogram's tail," Izv., Acad. Sci., USSR, Phys. Solid Earth **13**, 394–398 (1977).
[4] S. A. Shapiro and G. Kneib, "Seismic attenuation by scattering: Theory and numerical results," Geophys. J. Int. **114**, 373–391 (1993).
[5] J. Page, H. Schriemer, A. Bailey, and D. Weitz, "Experimental test of the diffusion approximation for multiply scattered waves," Phys. Rev. E **52**(3), 3106–3114 (1995).
[6] H. Schriemer, M. Cowan, J. Page, Z. Liu, and D. Weitz, "Energy velocity of diffusing waves in strongly scattering media," Phys. Rev. Lett. **79**(17), 3166–3191 (1997).
[7] A. Yodh and B. Chance, "Spectroscopy and imaging with diffusing light," Phys. Today **48**, 34–40 (1995).
[8] G. Li, "4D seismic monitoring of CO2 flood in a thin fractured carbonate reservoir," The Leading Edge **22**(7), 690 (2003).
[9] T. Davis, M. Terrel, R. Cardona, R. Benson, R. Kendall, and R. Winarsky, "Multicomponent seismic characterization and monitoring of the CO2 flood at Weyburn Field, Saskatchewan," The Leading Edge **22**(7), 696 (2003).
[10] K. van Wijk, J. A. Scales, and J. Tromp, "Analysis of strong scattering at the microscale," J. Acoust. Soc. Am. **115**(3), 1006–1011 (2004).
[11] J. De Rosny and P. Roux, "Multiple scattering in a reflecting cavity: Application to fish counting in a tank," J. Acoust. Soc. Am. **109**(6), 2587–

2597 (2001).

[12] J. de Rosny, P. Roux, and M. Fink, "Field fluctuation spectroscopy in a reverberant cavity with moving scatterers," Phys. Rev. Lett. **90**(9), 094302 (2003).

[13] P. Lemieux, M. Vera, and D. Durian, "Diffusing-light spectroscopies beyond the diffusion limit: The role of ballistic transport and anisotropic scattering," Phys. Rev. E **57**(4), 4498–4515 (1998).

[14] D. Weitz and D. Pine, "Diffusing wave spectroscopy," in *Dynamic Light Scattering, The Method and Some Applications*, edited by W. Brown (Clarendon, Oxford, 1993), pp. 652–720.

[15] M. Cowan, I. Jones, J. Page, and D. Weitz, "Diffusing acoustic wave spectroscopy," Phys. Rev. E **65**(066605), 1–11 (2002).

[16] R. Snieder, A. Grêt, H. Douma, and J. Scales, "Coda wave interferometry for estimating nonlinear behavior in seismic velocity," Science **295**(22), 2253–2255 (2002).

[17] O. Lobkis and R. Weaver, "Coda-wave interferometry in finite solids: Recovery of P-toS conversion rates in an elastodynamic billiard," Phys. Rev. Lett. **90**(25), 254302 (2003).

[18] M. Snieder and M. Vrijlandt, "Constraining relative source locations with coda wave interferometry: Theory and application to earthquake doublets in the Hayward Fault, California," J. Geophys. Res. (submitted).

[19] A. Tourin, A. Derode, and M. Fink, "Sensitivity to perturbations of a time-reversed acoustic wave in a multiple scattering medium," Phys. Rev. Lett. **87**(27), 274301 (2001).

[20] A. Tourin, M. Fink, and A. Derode, "Multiple scattering of sound," Waves Random Media **10**, R31–R60 (2000).

[21] J. Paasschens, "Solution of the time-dependent Boltzmann equation," Phys. Rev. E **53**, 1135–1141 (1997).

[22] S. Skipetrov and R. Maynard, "Diffuse waves in nonlinear disordered media," in *Wave Scattering in Complex Media: From Theory to Applications*, Vol. **107** of Nato Science Series, edited by B. van Tiggelen and S. Skipetrov (Kluwer Academic, Dordrecht, 2003), pp. 75–97.

[23] R. Snieder, "Coda wave interferometry and the equilibration of energy in elastic media," Phys. Rev. E **66**, 046615–046618 (2002).

[24] R. Snieder, "Imaging and averaging in complex media," in *Diffuse Waves in Complex Media*, edited by J. Fouque (Kluwer Academic, Raleigh, NC, 1999), pp. 405–454.

[25] R. Snieder, "Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).

[26] G. Roepstoff, *Path Integral Approach to Quantum Physics: An Introduction* (Springer, New York, 1994).

[27] A. Frankel and R. Clayton, "A finite-difference simulation of wave propagation in two dimensional random media," Bull. Seismol. Soc. Am. **74**(6), 2167–2186 (1984).

[28] L. Margerin, M. Campillo, and B. van Tiggelen, "Radiative transfer and diffusion of waves in a layered medium: New insight into coda q," Geophys. J. Int. **134**, 596–612 (1998).

[29] J. Gradshteyn and I. Ryzhik, in *Table of Integrals, Series, and Products*, 7th ed., edited by Y. V. Geronimus and M. Tseytlin (Academic, New York, 1973).

On the sound field of an oscillating disk in a finite open and closed circular baffle

Tim Mellow
Nokia UK Ltd., Farnborough, Hants GU14 0NG, England

Leo Kärkkäinen
Nokia Research Center, Helsinki, Finland

Equations describing the radiation characteristics of a rigid disk in a finite open baffle are derived using a method similar to that used by Streng for a circular membrane based upon the dipole part of the Kirchhoff–Helmholtz boundary integral formula. In this case, however, a power series solution to the radiation integral is derived in order to eliminate the need for numerical integration. Hence, a set of simultaneous equations is obtained by simply equating the coefficients of the power series, which leads to two mathematical functions, one real and one imaginary, that can be applied to any radial velocity distribution. This provides an alternative method to obtain the sound scattered by a disk or the complementary hole in an infinite resilient screen according to Babinet's principle. Using the principle of superposition (or Gutin concept), it is shown how the sound radiation characteristics of a disk radiating from just one side can be obtained by combining the radiation field of a disk in a finite baffle with that of a disk in an infinite baffle. This one-sided radiator may be interpreted as a disk in a thin, circular enclosure. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000828]

I. INTRODUCTION

The radiation characteristics of elementary sources are invaluable in providing lumped parameters for acoustics models based upon equivalent electrical circuits. Furthermore, they serve an important educational purpose in illustrating fundamental radiation and diffraction theory. They may also be used to provide benchmarks for boundary or finite element modeling (BEM/FEM). This can provide us with much information about the required element size and what kind of meshing geometry to use.

In addition to describing the radiation characteristics of a disk in an open or closed circular baffle, a more general aim of this paper is to present a simple unified approach to the problem of flat, axially symmetric sound sources based upon the Green's function in cylindrical coordinates and the Kirchhoff–Helmholtz boundary integral formula.[1] The latter is a general equation that describes the spatial distribution of the pressure within and on the surface bounding an acoustic medium. However, in the case of surfaces radiating into free space, the volume integral term can be omitted and the Green's function for an unbounded medium $g(\mathbf{r}|\mathbf{r}_0)$ is used so that the pressure distribution is described by

$$\widetilde{p}(\mathbf{r}) = \int\int g(\mathbf{r}|\mathbf{r}_0)\frac{\partial}{\partial n_0}\widetilde{p}(\mathbf{r}_0) - \widetilde{p}(\mathbf{r}_0)\frac{\partial}{\partial n_0}g(\mathbf{r}|\mathbf{r}_0)dS_0, \quad (1)$$

where the first term (or *monopole* part) is the integral of the inward-pointed normal gradient of the boundary values of $p(\mathbf{r}_0)$ and $g(\mathbf{r}|\mathbf{r}_0)$ over the surface, and the second term (or *dipole* part) is the integral of the boundary values of $p(\mathbf{r}_0)$ and the inward-pointed normal gradient of $g(\mathbf{r}|\mathbf{r}_0)$ over the surface.

Many of the classical analytical results reproduced during the course of this paper are best known through the works of Beranek[2] and Olson.[3] Fifty years ago, it must have been a formidable task calculating these results without the benefits of modern computing power. Even today, the task of reproducing the results of the original papers,[4,5] from which they were derived, is perhaps not so trivial unless the reader is well versed in the mathematical methods used. Hence, it is useful to derive a compact set of equations suitable for modern computer applications. Bouwkamp[6] and Spence[7] independently derived the first rigorous solutions to the problem of the freely suspended disk using iterative methods based upon oblate spheroidal wave functions, while Meixner and Fritze[8] calculated its near-field pressure response. Nimura and Watanabe[9] calculated the radiation impedance and directivity of a rigid disk in a finite circular open baffle for lower values of ka. Pachner[10] described a calculation method using spherical Bessel and Legendre functions and used the superposition of fields, although no results were presented. Crane[11] calculated the radiation impedance of a circular piston in a closed circular baffle directly by applying the boundary conditions of a sound field resulting from the superposition of a disk in a finite open baffle upon a disk in an infinite baffle. In theory, it should be possible to extract the open baffle impedance from these results simply by doubling them and then subtracting the infinite baffle impedance. However, given the approximate nature of this calculation method and the fact that the real open baffle impedance is so small at low frequencies, the result is inevitably wildly inaccurate. In the last few decades, there has been much interest in time-domain techniques.
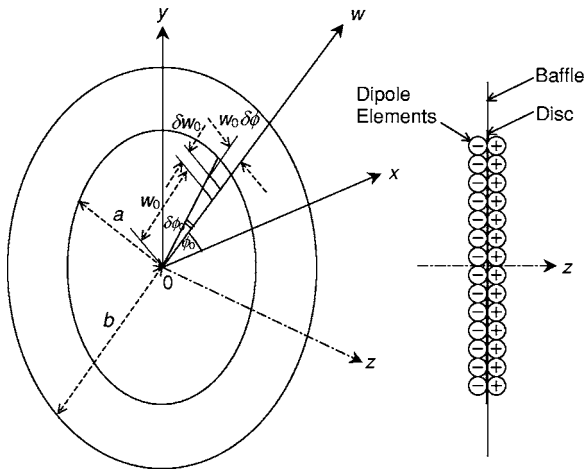
FIG. 1. Rigid disk in a finite open-back circular baffle.

Previously, one of the present authors[12] has used the least mean square (LMS) algorithm to tackle this problem, but found it to be numerically unsatisfactory for larger baffle sizes. Instead, the velocity of the radiator and its baffle is described, in this paper, by a step function using a Bessel function series, which produces a membrane-like expression. This allows a set of simultaneous equations to be obtained by simply equating the coefficients of the power series.

II. RIGID DISK IN A FINITE OPEN-BACK CIRCULAR BAFFLE

A. Boundary conditions

The disk shown in Fig. 1 lies in the xy plane, with its center at the origin and oscillates in the z direction with velocity $\tilde{u}_0$, thus radiating sound from both sides. The radius of the disk is a and it is assumed to be infinitesimally thin. It is surrounded by an infinitesimally thin circular baffle, the inner and outer radii of which are a and b, respectively. If $b=a$, the problem reduces to that of a *rigid disk in free space*. The area of each surface element is given by

$$\delta S_0 = w_0 \delta w_0 \delta \phi_0. \tag{2}$$

The pressure field on one side of the xy plane is the symmetrical "negative" of that on the other, so that

$$p(w,z) = -p(w,-z). \tag{3}$$

Consequently, there is a Dirichlet boundary condition in the plane of the disk where these equal and opposite fields meet.

$$p(w,0) = 0, \quad b < w \leq \infty \tag{4}$$

On the front and rear surfaces of the baffle, there is a Neumann boundary condition

$$\frac{\partial}{\partial z} \tilde{p}(w,z)|_{z=0\pm} = 0, \quad a < w \leq b. \tag{5}$$

Also, on the front and rear surfaces of the disk, there is the coupling condition

$$\frac{\partial}{\partial z} \tilde{p}(w,z)|_{z=0\pm} = -ik\rho c \tilde{u}_0(w), \quad 0 \leq w \leq a \tag{6}$$

where ρ is the density of air or any other surrounding medium, and c is the speed of sound in that medium. The annotation $\sim$ denotes a harmonically time-varying quantity where we have suppressed the term $e^{i\omega t}$. In order to tackle this problem, we shall use the second term or *dipole* surface integral part of Eq. (1). However, some prior expression for the frontal surface pressure distribution $\tilde{p}_+(w_0)$ is needed. Also, because the disk can radiate from both sides, the rear surface pressure distribution $\tilde{p}_-(w_0)$ must be included too, where $\tilde{p}_+(w_0) = -\tilde{p}_-(w_0)$. Streng[13,14] showed that the surface pressure distribution for any flat, axially symmetric unbaffled source (or sink), based upon Bouwkamp's solution[6] to the free-space wave equation in oblate spheroidal coordinates, could be written as

$$\tilde{p}_+(w_0) = -\tilde{p}_-(w_0) = \sum_{m=0}^{\infty} \tilde{A}_m \left(1 - \frac{w_0^2}{b^2}\right)^{m+(1/2)}. \tag{7}$$

where $\tilde{A}_m$ are the as yet unknown power series coefficients.

B. Solution of the free-space wave equation

The pressure distribution, in accordance with the Huygens–Fresnel principle, is given by the second term or *dipole* part of Eq. (1) written in cylindrical coordinates as follows:

$$\tilde{p}(w,z) = -\int_0^{2\pi} \int_0^b (\tilde{p}_+(w_0) - \tilde{p}_-(w_0))$$
$$\times \frac{\partial}{\partial z_0} g(w,z|w_0,z_0)|_{z_0=0+} w_0 \, dw_0 \, d\phi_0, \tag{8}$$

where the Green's function $g(w|w_0)$ is the solution to the following free-space wave equation in the presence of a monopole point source located at (w_0,z_0) on the surface of the disk:

$$(\nabla^2 + k^2) g(w,z|w_0,z_0) = -\delta(w-w_0, z-z_0), \tag{9}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial w^2} + \frac{1}{w}\frac{\partial}{\partial w} + \frac{\partial^2}{\partial z^2}, \tag{10}$$

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c}. \tag{11}$$

However, the pressure produced at each point (w,z) in space by each dipole element is defined in the integral of Eq. (8) by the product of the surface pressure, the inward-pointed normal gradient of the Green's function, and the area of each element given by Eq. (2). A solution to Eq. (9) is the free-space Green's function in cylindrical coordinates,[1] also known as the Lamb or Sommerfeld integral, which is given by

$$g(w,z|w_0,z_0) = \frac{i}{4\pi} \int_0^\infty J_0(\mu w) J_0(\mu w_0) \frac{\mu}{\sigma} e^{-i\sigma|z-z_0|} d\mu,$$

$$(12)$$

where

$$\sigma = \sqrt{k^2 - \mu^2}. \tag{13}$$

Due to axial symmetry, only the zeroth term of the original power series is shown in Eq. (12). The normal gradient of the Green's function at the surface is given by

$$\frac{\partial}{\partial z_0} g(w,z|w_0,z_0)\big|_{z_0=0+}$$

$$= \frac{1}{4\pi} \int_0^\infty J_0(\mu w) J_0(\mu w_0) \mu e^{-i\sigma z} d\mu. \tag{14}$$

Inserting this together with Eq. (7) in Eq. (8) and integrating over the surface of the disk and its baffle yields

$$\tilde{p}(w,z) = -b \sum_{m=0}^\infty \tilde{A}_m 2^{m+(1/2)} \Gamma\left(m + \frac{3}{2}\right) \int_0^\infty \left(\frac{1}{\mu b}\right)^{m+(1/2)}$$

$$\times J_{m+(3/2)}(\mu b) J_0(\mu w) e^{-i\sigma z} d\mu, \tag{15}$$

where the following identity[15] has been used:

$$\int_0^a w_0 \left(1 - \frac{w_0^2}{b^2}\right)^{m+(1/2)} J_0(\mu w_0) dw_0$$

$$= b^2 \int_0^1 t(1-t^2)^{m+(1/2)} J_0(\mu b t) dt$$

$$= b^2 2^{m+(1/2)} \Gamma\left(m + \frac{3}{2}\right) \left(\frac{1}{\mu b}\right)^{m+(3/2)} J_{m+(3/2)}(\mu b), \tag{16}$$

where $t = w_0/b$. Applying the boundary conditions of Eq. (5) and Eq. (6) leads to

$$\frac{\partial}{\partial z} \tilde{p}(w,z)\big|_{z=0\pm}$$

$$= ib \sum_{m=0}^\infty \tilde{A}_m 2^{m+(1/2)} \Gamma\left(m + \frac{3}{2}\right) \int_0^\infty \left(\frac{1}{\mu b}\right)^{m+(1/2)}$$

$$\times J_{m+(3/2)}(\mu b) J_0(\mu w) \sigma d\mu = -ik\rho c \tilde{u}_0 \Phi(w), \tag{17}$$

where

$$\Phi(w) = \begin{cases} 1, & 0 \le w \le a, \\ 0, & a < w \le b. \end{cases} \tag{18}$$

C. Series representation of the velocity distribution

Let the Bessel series have the form

$$\Phi(w) = \sum_{n=1}^\infty a_n J_0(j_{0n} w/b), \tag{19}$$

where j_{0n} is the nth zero of J_0 such that $J_0(j_{0n})=0$. Multiplication of Eq. (19) by the normalizing function $J_0(j_{0m}w/b)$ and integration over w gives
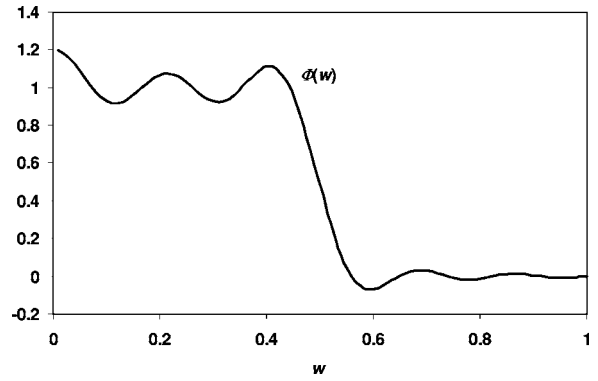
FIG. 2. Bessel series representation of the surface velocity distribution.

$$\int_0^b \Phi(w) J_0(j_{0m}w/b) w \, dw$$

$$= \int_0^a J_0(j_{0m}w/b) w \, dw$$

$$= \sum_{n=1}^\infty a_n \int_0^b J_0(j_{0n}w/b) J_0(j_{0m}w/b) w \, dw. \tag{20}$$

Integrating over w yields

$$\frac{ab}{j_{0m}} J_1(j_{0m}a/b) = \frac{b^2}{2} \sum_{n=1}^\infty a_n \delta_{mn} J_1^2(j_{0n}), \tag{21}$$

where the following identities[15] have been employed:

$$\int_0^a J_0(j_{0m}w/b) w \, dw = \int_0^1 J_0(j_{0m}ay/b) y \, dy$$

$$= ab J_1(j_{0m}a/b)/j_{0m}, \tag{22}$$

and

$$\int_0^1 J_0(\alpha x) J_0(\beta x) x \, dx = \begin{cases} 0, & \alpha \ne \beta, \\ J_1^2(\alpha)/2, & \alpha = \beta, \end{cases} \tag{23}$$

where $y = w/a$, $x = w/b$ and α, β are zeros of J_0. Hence

$$a_n = 2 \frac{a J_1(j_{0n}a/b)}{b j_{0n} J_1^2(j_{0n})}. \tag{24}$$

A series expansion[15] of $J_0(j_{0n}w/b)$ is given by

$$J_0(j_{0n}w/b) = \sum_{q=0}^\infty \frac{(-1)^q}{(q!)^2} \left(\frac{j_{0n}}{2}\right)^{2q} \left(\frac{w}{b}\right)^{2q}. \tag{25}$$

Substituting Eq. (24) and Eq. (25) in Eq. (19) yields

$$\Phi(w) = \frac{a}{b} \sum_{n=1}^N \sum_{q=0}^Q \frac{(-1)^q J_1(j_{0n}a/b)}{(q!)^2 J_1^2(j_{0n})} \left(\frac{j_{0n}}{2}\right)^{2q-1} \left(\frac{w}{b}\right)^{2q}, \tag{26}$$

where both infinite power series limits have been truncated. The result is plotted in Fig. 2 for $a = \frac{1}{2}$, $b = 1$, $N = 10$, and $Q = 50$.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

T. Mellow and L. Kärkkäinen: Disk in a circular baffle   1313

D. Formulation of the coupled problem

Equation (17) can be written more simply as

$$\sum_{m=0}^{\infty} \tau_m I_m(w,k) = -\Phi(w), \tag{27}$$

which needs to be solved for the normalized power series coefficients τ_m, where

$$\tau_m = \frac{\widetilde{A}_m}{(m+3/2)kb\rho c\widetilde{u}_0}, \tag{28}$$

and

$$I_m(w,k) = I_{mR}(w,k) + iI_{mI}(w,k), \tag{29}$$

where the real part of the integral in Eq. (29) is given by

$$I_{mR}(w,k) = b^2\Gamma\left(m+\frac{5}{2}\right)\int_0^k \left(\frac{2}{\mu b}\right)^{m+(1/2)}$$

$$\times \sqrt{k^2-\mu^2}J_{m+(3/2)}(\mu b)J_0(\mu w)d\mu, \tag{30}$$

and the imaginary part is given by

$$I_{mI}(w,k) = b^2\Gamma\left(m+\frac{5}{2}\right)\int_k^{\infty} \left(\frac{2}{\mu b}\right)^{m+(1/2)}$$

$$\times \sqrt{\mu^2-k^2}J_{m+(3/2)}(\mu b)J_0(\mu w)d\mu. \tag{31}$$

E. Solution of the real integral

Substitution of $\mu=k\sin\vartheta$ in Eq. (30) gives

$$I_{mR}(w,k) = 4\Gamma\left(m+\frac{5}{2}\right)\left(\frac{2}{kb}\right)^{m-(3/2)}\int_0^{\pi/2} \frac{\cos^2\vartheta}{(\sin\vartheta)^{m+1/2}}$$

$$\times J_0(kw\sin\vartheta)J_{m+(3/2)}(kb\sin\vartheta)d\vartheta. \tag{32}$$

A series expansion[15] of $J_{m+3/2}(kb\sin\vartheta)$ is given by

$$J_{m+(3/2)}(kb\sin\vartheta)$$

$$= \sum_{q=0}^{Q} \left(\frac{kb}{2}\right)^{2q+m+(3/2)} \frac{(-1)^q(\sin\vartheta)^{2q+m+3/2}}{q!\Gamma(q+m+5/2)}. \tag{33}$$

Replacing j_{0n}/b in Eq. (25) with $k\sin\vartheta$, and substituting this together with Eq. (33) in Eq. (32) gives

$$I_{mR}(w,k) = 4\sum_{q=0}^{Q}\sum_{r=0}^{R} \frac{(-1)^{q+r}\Gamma(m+5/2)}{(q!)^2r!\Gamma(r+m+5/2)}\left(\frac{w}{b}\right)^{2q}$$

$$\times \left(\frac{kb}{2}\right)^{2(q+r)+3}\int_0^{\pi/2} \cos^2\vartheta(\sin\vartheta)^{2(q+r)+1}d\vartheta. \tag{34}$$

Solution of the integral in Eq. (34) is enabled by use of the following identity:[16]

$$\int_0^{\pi/2} \cos^2\vartheta(\sin\vartheta)^{2(q+r)+1}d\vartheta = \frac{\sqrt{\pi}\Gamma(q+r+1)}{4\Gamma(q+r+5/2)}. \tag{35}$$

Evaluating the integral over ϑ yields

$$I_{mR}(w,k) = \sum_{q=0}^{Q}\sum_{r=0}^{R} \frac{\sqrt{\pi}(-1)^{q+r}\Gamma(m+5/2)\Gamma(q+r+1)}{(q!)^2r!\Gamma(r+m+5/2)\Gamma(q+r+5/2)}$$

$$\times \left(\frac{kb}{2}\right)^{2(q+r)+3}\left(\frac{w}{b}\right)^{2q}. \tag{36}$$

F. Solution of the imaginary integral

In order to change the infinite limit of Eq. (31) to a finite one, Streng[13] replaced the Bessel function $J_{q+3/2}$ with Hankel functions $H_{q+3/2}^{(1)}+H_{q+3/2}^{(2)}$. By applying contour integration theory, together with the substitution $\mu=ke^{i\vartheta}$, he showed that the imaginary integral Eq. (31) could be expressed as

$$I_{mI}(w,k) = \Re\left(4i\left(\frac{2}{kb}\right)^{m-(3/2)}\Gamma\left(m+\frac{5}{2}\right)\right.$$

$$\times \int_0^{\pi/2} e^{i[(1/2)-m]\vartheta}\sqrt{e^{2i\vartheta}-1}J_0(kwe^{i\vartheta})$$

$$\left.\times \{J_{m+(3/2)}(kbe^{i\vartheta}) + iY_{m+(3/2)}(kbe^{i\vartheta})\}d\vartheta\right). \tag{37}$$

A series expansion[15] of the Neumann function is given by

$$Y_{m+(3/2)}(kbe^{i\vartheta}) = \sum_{q=0}^{Q} \left(\frac{kb}{2}\right)^{2q-m-(3/2)} \frac{(-1)^{q+m}e^{i\vartheta(2q-m-3/2)}}{q!\Gamma(q-m-1/2)}. \tag{38}$$

Replacing j_{0n}/b in Eq. (25) and $k\sin\vartheta$ in Eq. (33) with $ke^{i\vartheta}$, and substituting these together with Eq. (38) in Eq. (37) gives

$$I_{mI}(w,k) = -4\sum_{q=0}^{Q}\sum_{r=0}^{R}\Re\left(\frac{(-1)^{q+r}\Gamma(m+5/2)}{(q!)^2r!\Gamma(r+m+5/2)}\left(\frac{w}{b}\right)^{2q}\right.$$

$$\times \left(\frac{kb}{2}\right)^{2(q+r)+3}i\int_0^{\pi/2} \sqrt{e^{2i\vartheta}-1}e^{2i\vartheta(q+r+1)}d\vartheta$$

$$- \frac{(-1)^{q+r+m}\Gamma(m+5/2)}{(q!)^2r!\Gamma(r-m-1/2)}\left(\frac{w}{b}\right)^{2q}\left(\frac{kb}{2}\right)^{2(q+r-m)}$$

$$\left.\times \int_0^{\pi/2} \sqrt{e^{2i\vartheta}-1}e^{2i\vartheta[q+r-m-(1/2)]}d\vartheta\right). \tag{39}$$

Solution of the integrals in Eq. (37) is enabled by use of the following identity:[16]

$$\int_0^{\pi/2} \sqrt{e^{2i\vartheta}-1}e^{2i\gamma\vartheta}d\vartheta = \frac{1}{2\gamma}\left({}_2F_1\left(-\frac{1}{2},\gamma;\gamma+1;-1\right)e^{i\pi\gamma}\right.$$

$$\left.- \frac{\sqrt{\pi}\Gamma(\gamma+1)}{2\Gamma(\gamma+3/2)}\right). \tag{40}$$

Evaluating the integrals over ϑ yields

$$I_{ml}(w,k) = 4\Gamma\left(m + \frac{5}{2}\right)\mathfrak{R}\left(\sum_{q=0}^{Q}\sum_{r=0}^{R} F_Y(q,r,m)\left(\frac{kb}{2}\right)^{2(q+r-m)}\right.$$

$$\left.\times\left(\frac{w}{b}\right)^{2q} - iF_J(q,r,m)\left(\frac{kb}{2}\right)^{2(q+r)+3}\left(\frac{w}{b}\right)^{2q}\right),$$

$$(41)$$

where the subfunctions F_Y and F_J are given by

$$F_Y(q,r,m) = (-1)^{q+r+m}\left(\frac{{}_2F_1(-1/2,\alpha;\alpha+1;-1)e^{i\pi\alpha}}{2\alpha(q!)^2 r!\Gamma(r-m-1/2)}\right.$$

$$\left.-\frac{\sqrt{\pi}\Gamma(\alpha)}{4(q!)^2 r!\Gamma(r-m-1/2)\Gamma(\alpha+3/2)}\right), \quad (42)$$

$$F_J(q,r,m) = (-1)^{q+r}\left(\frac{{}_2F_1(-1/2,\beta;\beta+1;-1)e^{i\pi\beta}}{2\beta(q!)^2 r!\Gamma(r+m+5/2)}\right.$$

$$\left.-\frac{\sqrt{\pi}\Gamma(\beta)}{4(q!)^2 r!\Gamma(r+m+5/2)\Gamma(\beta+3/2)}\right), \quad (43)$$

where

$$\alpha = q + r - m - 1/2, \tag{44}$$

$$\beta = q + r + 1. \tag{45}$$

However, for integer values of q and r, $iF_J(q,r,m)$ is purely imaginary and therefore makes no contribution to the real part of $I_{ml}(w,k)$. Similarly, the $e^{i\pi(q+r-m-1/2)}$ term of F_Y (q,r,m) is also purely imaginary and can therefore be excluded. Thus, the final result can be written

$$I_{ml}(w,k)$$

$$= -\sum_{q=0}^{Q}\sum_{r=0}^{R}\frac{\sqrt{\pi}(-1)^{q+r+m}\Gamma(m+5/2)\Gamma(q+r-m-1/2)}{(q!)^2 r!\Gamma(r-m-1/2)\Gamma(q+r-m+1)}$$

$$\times\left(\frac{kb}{2}\right)^{2(q+r-m)}\left(\frac{w}{b}\right)^{2q}. \tag{46}$$

G. Calculation of the power series coefficients (final set of simultaneous equations)

Truncating the infinite power series in Eq. (27) to order M and equating the coefficients of $(w/b)^{2q}$ yields the final set of M simultaneous equations in τ_m as follows:

$$\sum_{m=0}^{M}({}_m\mathbf{B}_q(kb) - i\,{}_m\mathbf{S}_q(kb))\tau_m = -\Phi_q, \tag{47}$$

where $\mathbf{B}$ shall be named the *Bouwkamp* function as defined by

$${}_m\mathbf{B}_q(kb) = \sqrt{\pi}\sum_{r=0}^{M}\frac{(-1)^{q+r}\Gamma(m+5/2)\Gamma(q+r+1)}{r!(q!)^2\Gamma(r+m+5/2)\Gamma(q+r+5/2)}$$

$$\times\left(\frac{kb}{2}\right)^{2(q+r)+3}, \tag{48}$$

and $\mathbf{S}$ shall be named the *Streng* function as defined by

$${}_m\mathbf{S}_q(kb) = \sqrt{\pi}\sum_{r=0}^{M}\frac{(-1)^{q+r+m}\Gamma(m+5/2)\Gamma(q+r-m-1/2)}{r!(q!)^2\Gamma(r-m-1/2)\Gamma(q+r-m+1)}$$

$$\times\left(\frac{kb}{2}\right)^{2(q+r-m)}, \tag{49}$$

and

$$\Phi_q = \frac{a}{b}\sum_{n=1}^{N}\frac{(-1)^q J_1(j_{0n}a/b)}{(q!)^2 J_1^2(j_{0n})}\left(\frac{j_{0n}}{2}\right)^{2q-1}, \tag{50}$$

which is solved for $q = 0, 1, 2, \ldots, M-1, M$. In the case of $b = a$, the problem reduces to that of *a rigid disk in free space*, in which case

$$\Phi_q = \delta_{q0}, \tag{51}$$

where δ_{q0} is the Kronecker delta function. In the case of $b = 8a$, calculations were performed with $M = 200$, $N = 40$, and 140 digits of precision. These values were reduced in proportion to b/a for smaller baffle sizes.

H. Surface pressure

From Eq. (28), it follows that

$$\tilde{A}_m = \tau_m(m + 3/2)kb\rho c\tilde{u}_0. \tag{52}$$

After substituting this in Eq. (7), the surface pressure can be written as

$$\tilde{p}_+(w_0) = kb\rho c\tilde{u}_0\sum_{m=0}^{M}\left(m+\frac{3}{2}\right)\tau_m\left(1-\frac{w_0^2}{b^2}\right)^{m+(1/2)}. \tag{53}$$

I. Radiation impedance

The total force $\tilde{F}$ acting upon the disk can be found by integrating the pressure from Eq. (53) over its surface on *both* sides as follows:

$$\tilde{F} = -\int_0^{2\pi}\int_0^a 2\tilde{p}_+(w_0)w_0\,dw_0\,d\phi_0$$

$$= -2\pi kb^3\rho c\tilde{u}_0\sum_{m=0}^{M}\tau_m\left\{1-\left(1-\frac{a^2}{b^2}\right)^{m+(3/2)}\right\}. \tag{54}$$

The acoustic radiation impedance z_{ar} is then given by

$$z_{ar} = \frac{\tilde{F}}{S\tilde{U}_0} = \frac{\tilde{F}}{S^2\tilde{u}_0} = \frac{2\rho c}{S}(R_R + iX_R), \tag{55}$$

where $\tilde{U}_0$ is the total volume velocity produced by the disk and S is the area of the disk given by $S = \pi a^2$. R_R is the normalized radiation *resistance* given by

$$R_R = -kb\frac{b^2}{a^2}\sum_{m=0}^{M}\mathfrak{R}(\tau_m)\left\{1-\left(1-\frac{a^2}{b^2}\right)^{m+(3/2)}\right\}, \tag{56}$$

and X_R is the normalized radiation *reactance* given by
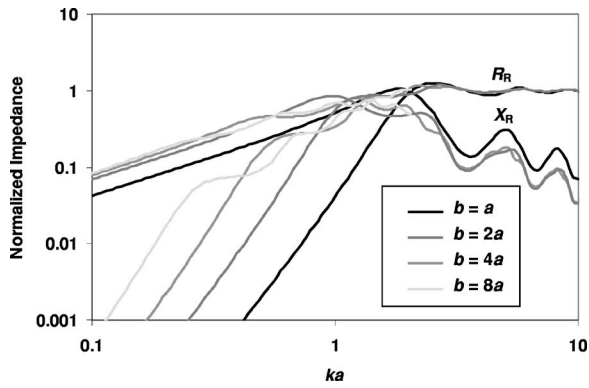
FIG. 3. Normalized radiation impedance of a disk in a finite open-back baffle.

$$X_R = -kb\frac{b^2}{a^2}\sum_{m=0}^{M} \Im(\tau_m)\left\{1 - \left(1 - \frac{a^2}{b^2}\right)^{m+(3/2)}\right\}. \quad (57)$$

This result is plotted in Fig. 3. By using standard curve-fitting methods, the following asymptotic expressions can be written:

$$R_R \approx 0.068\left(\frac{b^2}{a^2} - 0.56\right)k^4a^4, \quad kb < 0.5, \quad (58)$$

$$X_R \approx \frac{8(3b - 2a)}{3\pi(3b - a)}ka, \quad kb < 0.5. \quad (59)$$

J. Near-field pressure response

Applying expression for $\widetilde{A}_m$ in Eq. (52) to Eq. (15), the normalized near-field pressure can be written

$$\frac{\widetilde{p}(w,z)}{\rho c\widetilde{u}_0} = -kb^2\sum_{m=0}^{\infty} \tau_m\Gamma\left(m + \frac{5}{2}\right)$$
$$\times \int_0^{\infty}\left(\frac{2}{\mu b}\right)^{m+(1/2)} J_{m+(3/2)}(\mu b)J_0(\mu w)e^{-i\sigma z}d\mu. \quad (60)$$

By substituting $\mu = kt$, Eq. (60) can be written

$$\frac{\widetilde{p}(w,z)}{\rho c\widetilde{u}_0} = -4\sum_{m=0}^{M} \tau_m\Gamma\left(m + \frac{5}{2}\right)\left(\frac{2}{kb}\right)^{m-(3/2)}$$
$$\times \left(\int_0^1 \frac{1}{t^{m+(1/2)}}e^{-ikz\sqrt{1-t^2}}J_{m+(3/2)}(kbt)J_0(kwt)dt\right.$$
$$+ \left.\int_1^{\infty} \frac{1}{t^{m+(1/2)}}e^{-kz\sqrt{t^2-1}}J_{m+(3/2)}(kbt)J_0(kwt)dt\right). \quad (61)$$

The infinite integral in Eq. (61) converges so rapidly that the infinite limit can be replaced with a finite value of say, 50, without significant loss of accuracy. The result is shown in Fig. 4 for $b=2a$ with various values of ka.

$ka = 1$

$ka = 3$

$ka = 5$

$ka = 10$
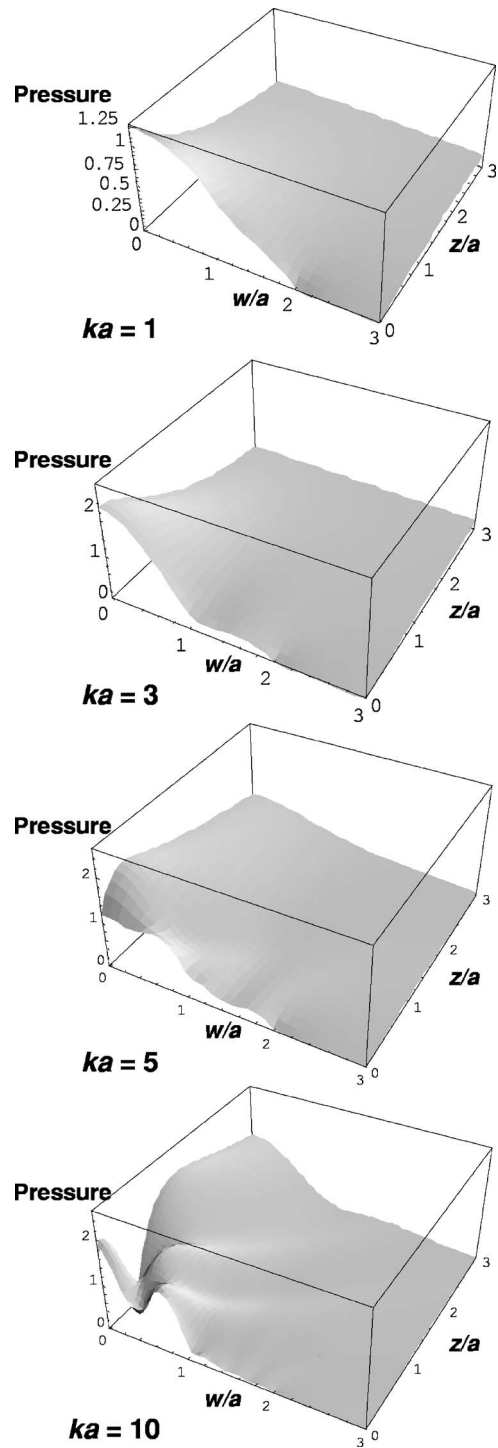
FIG. 4. Near-field pressure response of a disk in a finite open-back baffle with $b=2a$.

K. Far-field pressure response

In the case of the far-field response, it is more convenient to use spherical coordinates so that the far-field polar responses can be obtained directly. Rayleigh's far-field approximation[1] is ideal for this purpose

$$g(r,\theta,\phi|w_0,\phi_0,z_0) = \frac{1}{4\pi r}e^{-ik[r-w_0\sin\theta\cos(\phi-\phi_0)-z_0\cos\theta]}. \quad (62)$$

Again, the second term or *dipole* part of Eq. (1) is used, but this time in spherical coordinates

$$\tilde{p}(r,\theta,\phi) = -\int_0^{2\pi}\int_0^b (\tilde{p}_+(w_0) - \tilde{p}_-(w_0))$$

$$\times \frac{\partial}{\partial z_0} g(r,\theta,\phi|w_0,\phi_0,z_0)|_{z_0=0+} w_0\, dw_0\, d\phi_0. \tag{63}$$

However, it is necessary to find the surface normal gradient of the far-field Green's function given in Eq. (62)

$$\frac{\partial}{\partial z_0} g(r,\theta,\phi|w_0,\phi_0,z_0)|_{z_0=0+} = ik\cos\theta\frac{1}{4\pi r}e^{-ikr}$$

$$\times e^{ikw_0 \sin\theta \cos(\phi-\phi_0)}. \tag{64}$$

As the disk is axially symmetric, any reference angle ϕ may be chosen in Eq. (64). Therefore, the problem can be simplified by letting $\phi=\pi/2$ so that $\cos(\phi-\phi_0)=\sin\phi_0$. Inserting Eq. (64) and Eq. (53) into Eq. (63) and integrating over the surface of the disk yields

$$\tilde{p}(r,\theta) = -ik\rho c S\tilde{u}_0\frac{1}{2\pi r}e^{-ikr}D(\theta), \tag{65}$$

where S is the area of the disk, given by $S=\pi a^2$, and the following identities[15] have been used:

$$\frac{1}{2\pi}\int_0^{2\pi} e^{it\sin\phi_0}d\phi_0 = J_0(t), \tag{66}$$

where $t=kw_0\sin\theta$ together with Eq. (16), where $\mu=k\sin\theta$. The directivity function $D(\theta)$ is given by

$$D(\theta) = -kb\frac{b^2}{a^2}\cos\theta\sum_{m=0}^{M}\tau_m\Gamma\left(m+\frac{5}{2}\right)\left(\frac{2}{kb\sin\theta}\right)^{m+(3/2)}$$

$$\times J_{m+(3/2)}(kb\sin\theta). \tag{67}$$

The on-axis pressure is evaluated using $\theta=0$ in Eq. (64) before inserting it into Eq. (63). This results in an integral that is similar to the one for the radiation impedance in equation Eq. (54). Hence

$$D(0) = -kb\frac{b^2}{a^2}\sum_{m=0}^{M}\tau_m. \tag{68}$$

It is worth noting that in the unbaffled case, where $b=a$, the on-axis response is simply defined by $D(0)=R_R+iX_R$. Again, using standard curve-fitting methods, the following asymptotic expression can be written:

$$D(0) \approx i0.66\left(\frac{b}{a}-0.3\right)ka, \quad kb<0.5. \tag{69}$$

The normalized on-axis response is shown in Fig. 5, where the normalized SPL is given by

$$\text{SPL}_{\text{Norm}} = 20\log_{10}|D(0)|. \tag{70}$$

The normalized directivity function $20\log_{10}(|D(\theta)|/|D(0)|)$ is plotted in Fig. 6 with $b=2a$ for various values of ka and in Fig. 7 with $ka=\pi/2$ for various values of b/a.
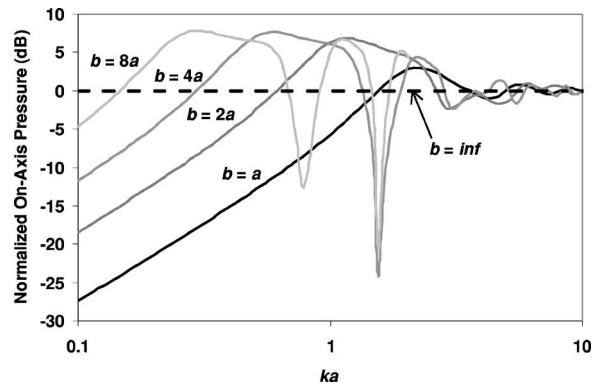
FIG. 5. Normalized far-field on-axis response of a disk in a finite open-back baffle with constant acceleration.
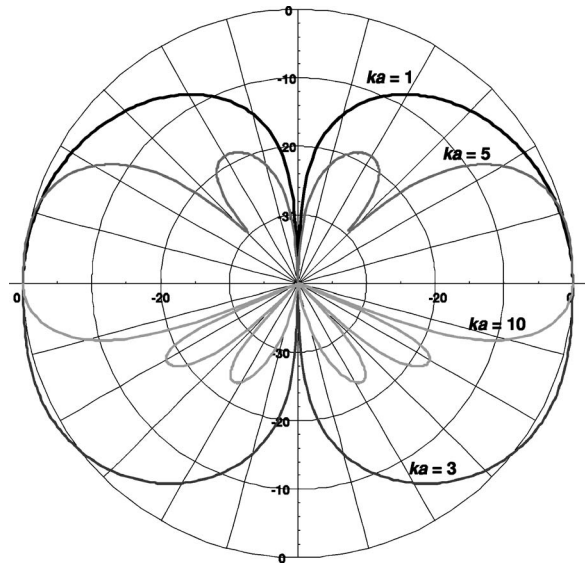
FIG. 6. Normalized far-field directivity function of a disk in a finite open-back baffle with $b=2a$.
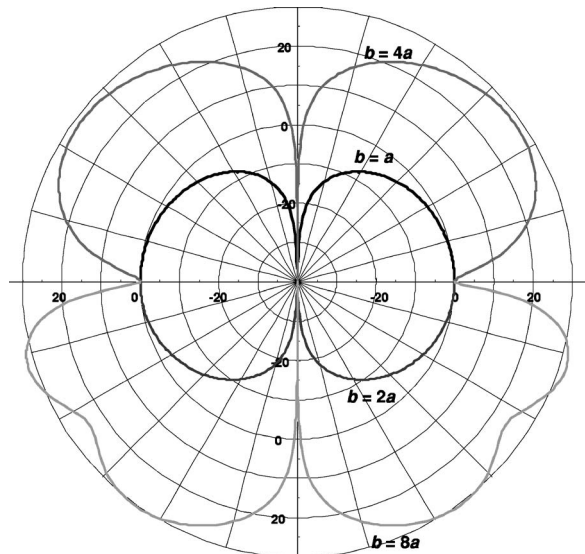
FIG. 7. Normalized far-field directivity function of a disk in a finite open-back baffle with $ka=\pi/2$.
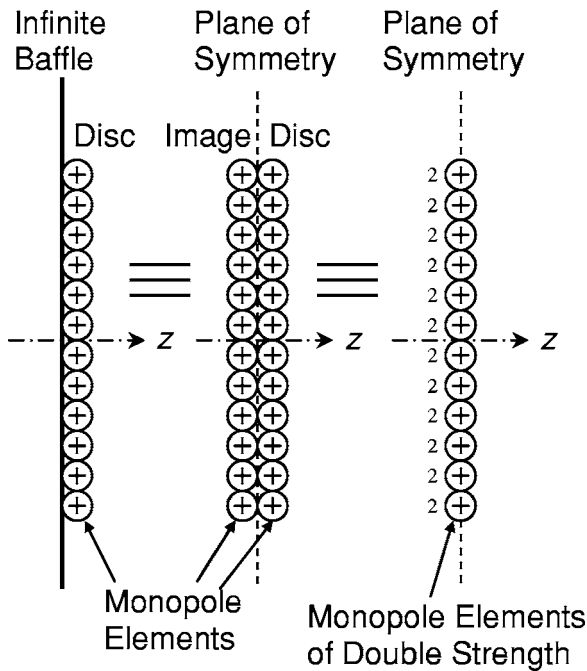
FIG. 8. Rigid disk in an infinite baffle.

III. RIGID DISK IN AN INFINITE BAFFLE

A. Boundary conditions

The original derivation of the radiation impedance by Lord Rayleigh[17] over 100 years ago used an ingenious coordinate system to simplify the problem. Stenzel[18] derived the near-field pressure response using the simple omnidirectional Green's function. Since equations describing the radiation characteristics of a disk in an infinite baffle are needed for Sec. IV, it is perhaps worth providing a brief review of King's method,[19] which is consistent with Sec. II as it uses the Green's function in cylindrical coordinates.

The disk shown in Fig. 1 is now mounted in an infinite baffle in the xy plane and oscillates in the z direction with a harmonically time-dependent velocity $\tilde{u}_0$. The area of each surface element is given by Eq. (2). However, the dipole elements are replaced with *monopole* elements as shown in Fig. 8. The monopole source elements and their images together form the disk source. Since they are coincident in the plane of the baffle, they combine to form elements of double strength. Hence, the disk in an infinite baffle can be modeled as a "breathing" disk in free space. It may also be considered as a pulsating sphere of the same radius compressed into the plane of the disk. Due to the symmetry of the pressure fields on either side of the baffle, there is the following Neumann boundary condition on its surface:

$$\frac{\partial}{\partial z}\tilde{p}(w,z)\big|_{z=0+} = 0, \quad a < w \leq \infty, \tag{71}$$

and on the surface of the disk there is the coupling condition

$$\frac{\partial}{\partial z}\tilde{p}(w,z)\big|_{z=0+} = -ik\rho c\tilde{u}_0, \quad 0 \leq w \leq a, \tag{72}$$

where ρ is the density of air or any other surrounding medium and c is the speed of sound in that medium.

B. Near-field pressure

The pressure distribution, in accordance with the Huygens–Fresnel principle, is given by the first term or *monopole* part of Eq. (1) taking into account the double-layer source

$$\tilde{p}(w,z) = 2\int_0^{2\pi}\int_0^a g(w,z|w_0,z_0)$$

$$\times \frac{\partial}{\partial z_0}\tilde{p}(w_0,z_0)\big|_{z_0=0+}w_0\,dw_0\,d\phi_0. \tag{73}$$

Inserting the Green's function of Eq. (12), together with Eq. (72), into Eq. (73), while integrating over the surface of the disk yields

$$\frac{\tilde{p}(w,z)}{\rho c\tilde{u}_0} = -ka\int_0^\infty J_1(\mu a)J_0(\mu w)\frac{1}{\sigma}e^{-i\sigma z}d\mu, \tag{74}$$

where we have used Eq. (22) with $y=w_0/a$ and $\mu=j_{0m}/b$. By substituting $\mu=kt$, Eq. (74) can be written

$$\frac{\tilde{p}(w,z)}{\rho c\tilde{u}_0} = -ka\left(\int_0^1 \frac{1}{\sqrt{1-t^2}}e^{-ikz\sqrt{1-t^2}}J_1(kat)J_0(kwt)dt\right.$$

$$\left. + i\int_1^\infty \frac{1}{\sqrt{t^2-1}}e^{-kz\sqrt{t^2-1}}J_1(kat)J_0(kwt)dt\right). \tag{75}$$

The infinite integral in Eq. (75) converges so rapidly that the infinite limit can be replaced with a finite value of, say, 50 without significant loss of accuracy. This result is shown in Fig. 9 for various values of ka.

C. Radiation impedance

The total force $\tilde{F}$ acting upon the disk can be found by integrating the pressure from Eq. (74) over its surface as follows:

$$\tilde{F} = -\int_0^{2\pi}\int_0^a \tilde{p}(w,z)\big|_{z=0+}w\,dw\,d\phi = 2\pi ka^2\rho c\tilde{u}_0$$

$$\times \left(\int_0^k \frac{J_1^2(\mu a)}{\mu\sqrt{k^2-\mu^2}}d\mu - i\int_k^\infty \frac{J_1^2(\mu a)}{\mu\sqrt{\mu^2-k^2}}d\mu\right), \tag{76}$$

where Eq. (22) has again been used. Using King's solutions[19] to the integrals over μ in Eq. (76) leads to the following expression for the acoustic radiation impedance:

$$z_{ar} = \frac{\tilde{F}}{S\tilde{U}_0} = \frac{\tilde{F}}{S^2\tilde{u}_0} = (R_R - iX_R)\frac{\rho c}{S}, \tag{77}$$

where $\tilde{U}_0$ is the total volume velocity produced by the disk and S is its surface area given by $S = \pi a^2$. R_R is the normalized radiation *resistance* given by

ka = 1

ka = 3

ka = 5

ka = 10
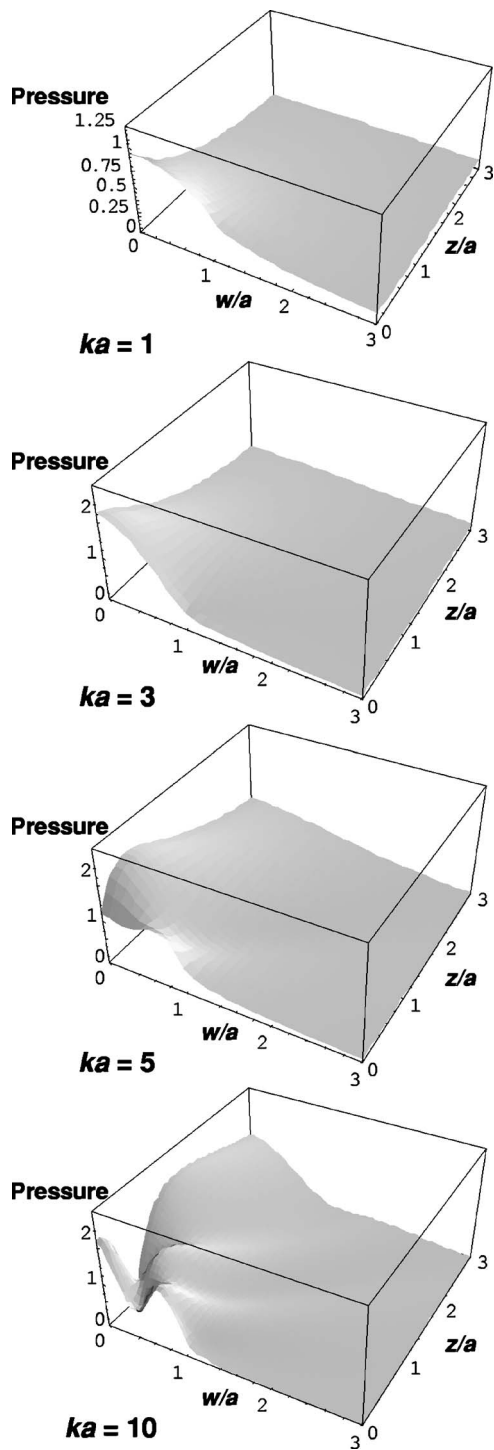
FIG. 9. Near-field pressure response of a disk in an infinite baffle.

$$R_R = 1 - \frac{J_1(2ka)}{ka}, \tag{78}$$

and X_R is the normalized radiation *reactance* given by

$$X_R = \frac{\mathbf{H}_1(2ka)}{ka}, \tag{79}$$

where J_1 is the first-order Bessel function and $\mathbf{H}_1$ is the first-order Struve function. This result is plotted in Fig. 10. The following well-known asymptotic expressions can also be written:
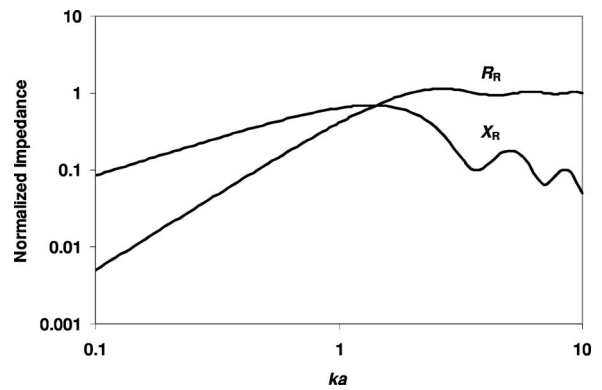
FIG. 10. Normalized radiation impedance of a disk in an infinite baffle.

$$R_R \approx k^2 a^2/2, \quad ka < 0.5, \tag{80}$$

$$X_R \approx 8ka/(3\pi), \quad ka < 0.5. \tag{81}$$

D. Far-field response

Again, the first term or *monopole* part of the K-H boundary integral formula (1) is used, but this time in spherical coordinates

$$\tilde{p}(r,\theta,\phi) = 2 \int_0^{2\pi} \int_0^a g\,(r,\theta,\phi|w_0,\phi_0,z_0)\big|_{z_0=0+}$$

$$\times \frac{\partial}{\partial z_0} \tilde{p}(w_0,\phi_0,z_0)\big|_{z_0=0+} w_0\,dw_0\,d\phi_0. \tag{82}$$

Equation (62) and Eq. (72) can now be inserted into Eq. (82), again letting $n=0$ and $\phi=\pi/2$ so that $\cos(\phi-\phi_0)=\sin\phi_0$. Integrating over the surface of the disk while using Eq. (66) and Eq. (22) with $j_{0m}/b=k\sin\theta$ yields

$$\tilde{p}(r,\theta) = -ik\rho c S\tilde{u}_0 \frac{1}{2\pi r} e^{-ikr} D(\theta), \tag{83}$$

where S is the area of the disk given by $S=\pi a^2$. The directivity function $D(\theta)$ is given by

$$D(\theta) = \frac{2J_1(ka\sin\theta)}{ka\sin\theta}. \tag{84}$$

The normalized directivity function $20\log_{10}|D(\theta)|$ is plotted in Fig. 11 for various values of ka. The on-axis pressure is evaluated using $\theta=0$ in Eq. (62) before inserting it into Eq. (82), which leads to

$$D(0) = 1. \tag{85}$$

The final result is most commonly written

$$\tilde{p}(r,0) = -i\omega\rho\tilde{U}_0 \frac{1}{2\pi r} e^{-ikr}, \tag{86}$$

where the volume velocity is given by $\tilde{U}_0 = S\tilde{u}_0$.
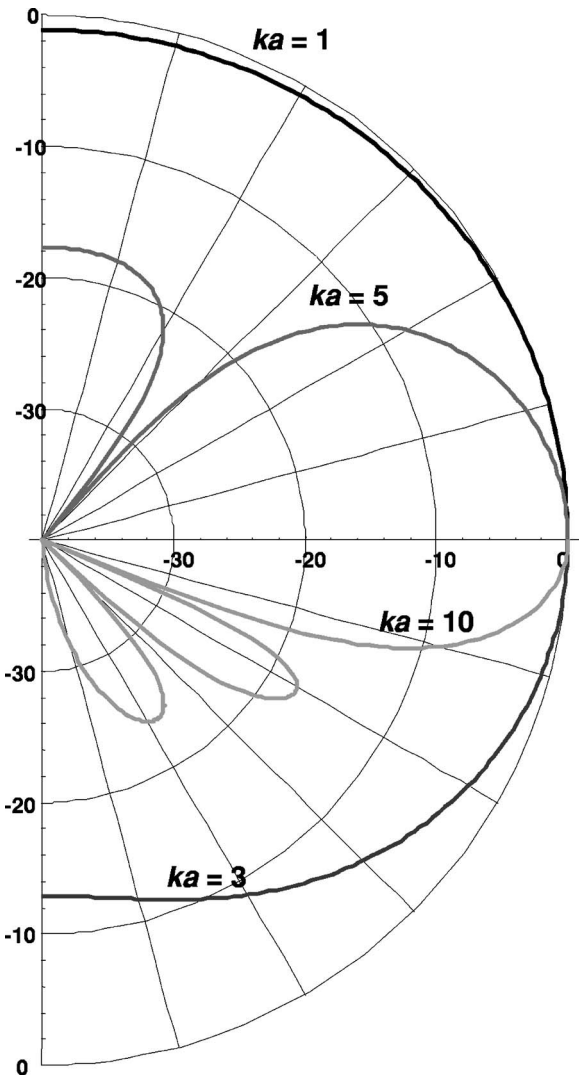
FIG. 11. Normalized far-field directivity function of a disk in an infinite baffle.

FIG. 12. Rigid disk in a finite closed-back circular baffle.

IV. RIGID DISK IN A FINITE, CLOSED-BACK CIRCULAR BAFFLE

A. Boundary conditions

The disk shown in Fig. 1, together with its circular baffle, lies in the xy plane and oscillates in the z direction with velocity $\tilde{u}_0$. The radius of the disk a and the inner and outer radii of the baffle are a and b, respectively. However, the configuration is now modified so that sound from the rear of the disk is blocked by a small cylindrical enclosure of radius b and depth h. The model is valid providing $h \leqslant a/4$, although the radiation impedance is virtually the same as that of a *rigid disk at the end of a flanged infinite tube* (or an unflanged tube[5] in the case of $b=a$). The area of each surface element is given by Eq. (2).

B. Near-field pressure

It is shown in Fig. 12 that the pressure distribution of the closed-back radiator is the sum of the outputs of the same radiator with an open back and in an infinite baffle. In other words, it uses the full K-H boundary integral of Eq. (1) in-

cluding both the dipole and monopole parts. Hence, the near-field pressure distribution can be obtained by summing Eq. (61) and Eq. (75). However, in order to normalize the response, the resulting pressure is halved. The result is shown in Fig. 13 with $b=2a$ for various values of ka.
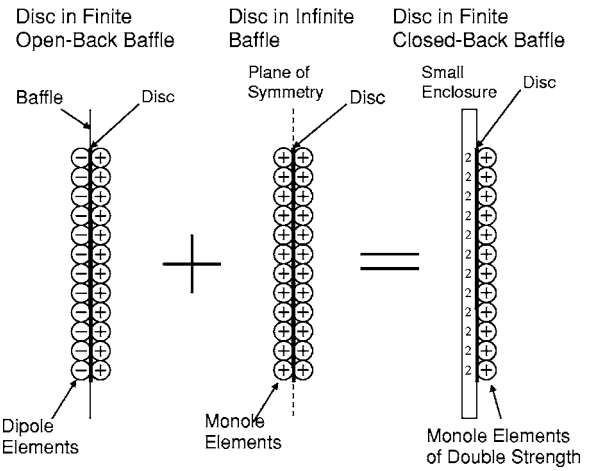
C. Radiation impedance

The same argument can be applied to the radiation impedance too, which is simply proportional to the sum of the surface pressures. The real part of the normalized radiation impedance can thus be obtained by combining Eq. (56) and Eq. (78) as follows:

$$R_R = \frac{1}{2}\left(1 - \frac{J_1(2ka)}{ka}\right.$$
$$\left. - kb\frac{b^2}{a^2}\sum_{m=0}^{M}\Re(\tau_m)\left\{1 - \left(1 - \frac{a^2}{b^2}\right)^{m+(3/2)}\right\}\right). \quad (87)$$

Similarly, the imaginary part can be obtained by combining Eq. (57) and Eq. (79)

$$X_R = \frac{1}{2}\left(\frac{\mathbf{H}_1(2ka)}{ka}\right.$$
$$\left. - kb\frac{b^2}{a^2}\sum_{m=0}^{M}\Im(\tau_m)\left\{1 - \left(1 - \frac{a^2}{b^2}\right)^{m+(3/2)}\right\}\right). \quad (88)$$

This result is plotted in Fig. 14. The asymptotic expressions can be written

$$R_R \approx k^2 a^2/4, \quad ka < 0.5, \quad (89)$$

$$X_R \approx \frac{4}{\pi}\left(\frac{2b-a}{3b-a}\right)ka, \quad ka < 0.5. \quad (90)$$

D. Far-field response

The far-field response takes on the same form as that in both a finite open baffle given by Eq. (65) and an infinite baffle given by Eq. (83)

Pressure

ka = 1

Pressure

ka = 3

Pressure

ka = 5

Pressure

ka = 10
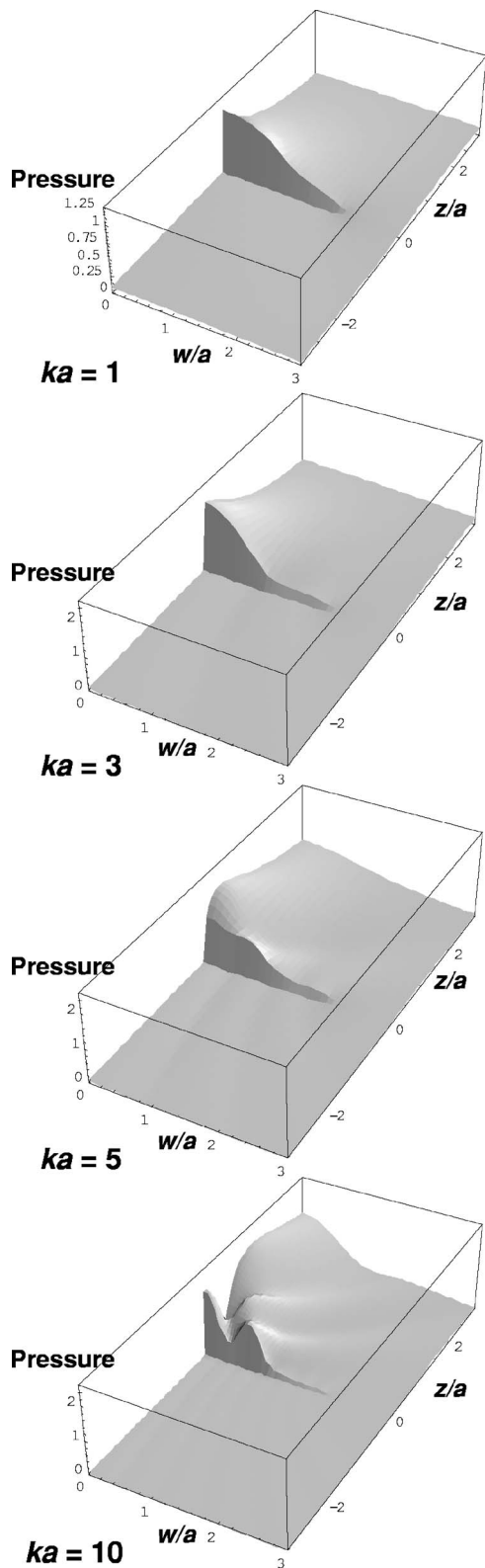
FIG. 13. Near-field pressure response of a disk in a finite closed-back baffle with $b=2a$.

$$\tilde{p}(r,\theta) = -ik\rho cS\tilde{u}_0 \frac{1}{2\pi r} e^{-ikr} D(\theta), \qquad (91)$$

where S is the area of the disk given by $S=\pi a^2$. The direc-
tivity function $D(\theta)$ is obtained by combining the directivity
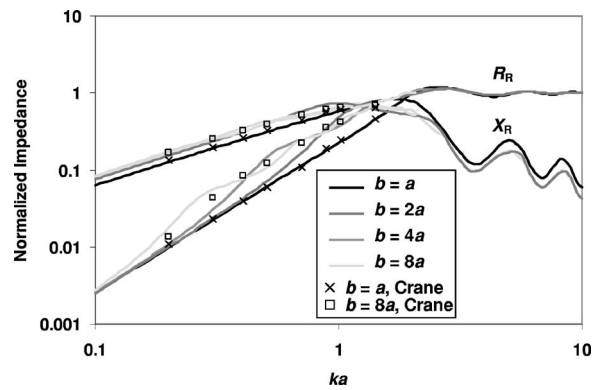
FIG. 14. Normalized radiation impedance of a disk in a finite closed-back baffle.

function in a finite open baffle given by Eq. (67) with that in
an infinite baffle given by Eq. (84)

$$D(\theta) = \frac{J_1(ka\sin\theta)}{ka\sin\theta} - kb\frac{b^2}{a^2}\cos\theta\sum_{m=0}^{M}\tau_m\Gamma\left(m+\frac{5}{2}\right)$$

$$\times \left(\frac{2}{kb\sin\theta}\right)^{m+3/2} J_{m+(3/2)}(kb\sin\theta). \qquad (92)$$

Similarly, the on-axis response may be obtained by combin-
ing the on-axis response in a finite open baffle given by Eq.
(68) with that in an infinite baffle given by Eq. (85), where
the latter is just unity

$$D(0) = \frac{1}{2}\left(1 - kb\frac{b^2}{a^2}\sum_{m=0}^{M}\tau_m\right). \qquad (93)$$

The normalized on-axis response is shown in Fig. 15 again
using Eq. (70). The normalized directivity function
$20\log_{10}(|D(\theta)|/|D(0)|)$ is plotted in Fig. 16 with $b=2a$ for
various values of ka and in Fig. 17 with $ka=\pi/2$ for vari-
ous values of b/a.

V. BABINET'S PRINCIPLE

Babinet's principle,[20] as developed by Bouwkamp[6] and
Wiener,[4] states that the diffraction pattern resulting from the
transmission of a plane wave through a hole in an infinite
resilient screen (i.e., with zero surface impedance) is equiva-
lent to that produced by the scattering of the same incident
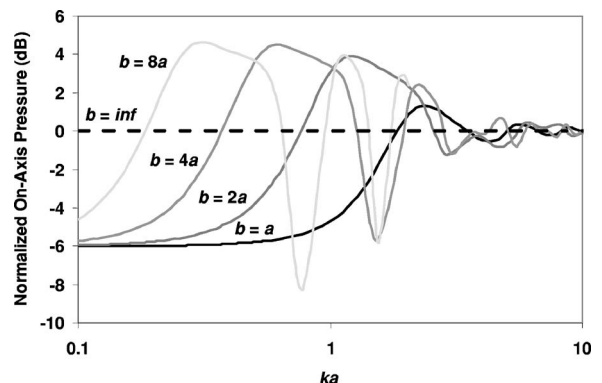
FIG. 15. Normalized far-field on-axis response of a disk in a finite closed-
back baffle with constant acceleration.
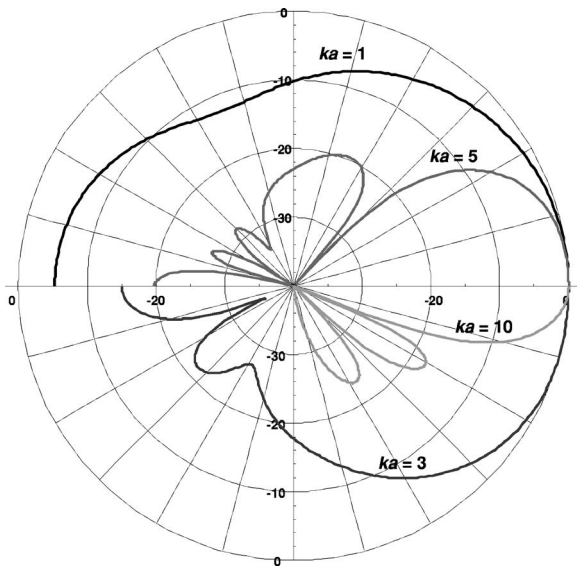
FIG. 16. Normalized directivity far-field function of a disk in a finite closed-back baffle with $b=2a$.

wave by the complementary rigid disk. Furthermore, the scattered wave is identical to that produced if the disk itself were oscillating in free space, providing the velocity of the disk is equal and opposite to the particle velocity of the incident wave at the surface of the disk or hole in the absence of any obstacle. This is illustrated in Fig. 18. For clarity, the diagram portrays the scattering of a sound wave at some very high frequency where there is minimal diffraction. However, the principle applies at all frequencies. The resulting sound field is given by

$$\widetilde{p}(\mathbf{r}) = \widetilde{p}_{\text{Inc}}(\mathbf{r}) + \widetilde{p}_{\text{Scat}}(\mathbf{r}), \tag{94}$$

where $\widetilde{p}_{\text{Inc}}(\mathbf{r})$ is the incident sound field in the absence of a hole or disk given by
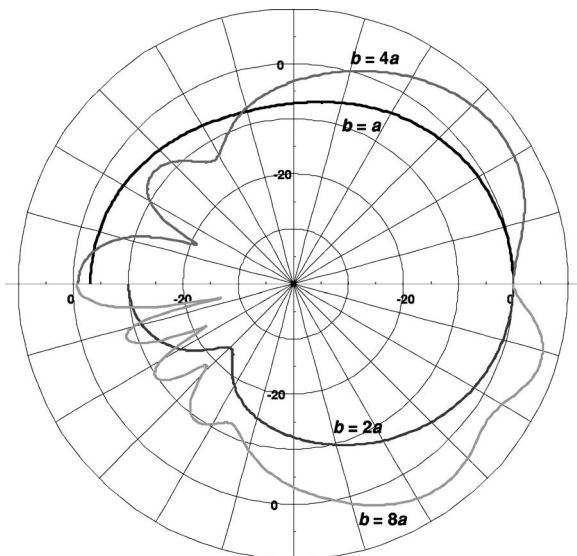
FIG. 17. Normalized directivity far-field function of a disk in a finite closed-back baffle with $ka=\pi/2$.
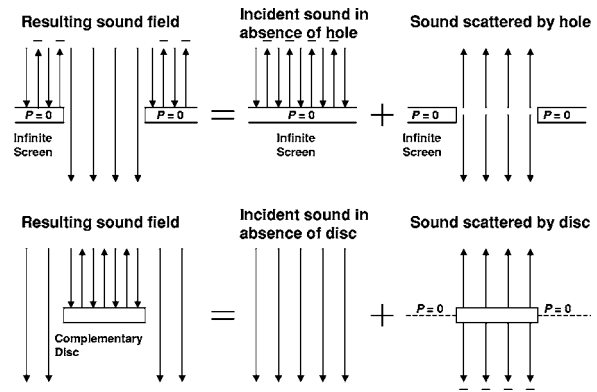
FIG. 18. Babinet's principle.

$$\widetilde{p}_{\text{Inc}}(z) = \begin{cases} -ik\rho c\widetilde{\Psi}(e^{ikz} - e^{-ikz}), & \text{bright side of screen} \\ 0, & \text{dark side of screen} \\ -ik\rho c\widetilde{\Psi}e^{ikz}, & \text{without disk(or screen)} \end{cases}$$

$$\tag{95}$$

and using Eq. (75) for $\widetilde{p}_{\text{Scat}}(w,z)$ or Eqs. (83) and (84) for $\widetilde{p}_{\text{Scat}}(r,\theta)$. Also, it can be stated that the force $\widetilde{F}_{\text{inc}}$ exerted upon the disk by an incident wave is given by

$$\widetilde{F}_{\text{inc}} = z_{\text{ar}}S\widetilde{u}_{\text{p}}, \tag{96}$$

where $\widetilde{u}_{\text{p}}$ is the undisturbed incident wave particle velocity at the surface of the disk and S is its surface area. The radiation impedance z_{ar} is given by Eqs. (55)–(57).

VI. APPROXIMATION FOR THE RIGID DISK IN FREE SPACE (RADIATION FROM AN UNBAFFLED RESILIENT DISK)

The following section contains a review of a historical approach to the problem of the disk in free space.[1] Prior to the work of Bouwkamp,[6] the Kirchhoff theory of diffraction had been used, which makes an assumption of equally distributed surface pressure in order to simplify the problem. In the case of diffraction, the pressure across the complementary hole in an infinite resilient screen is assumed to be the same as that of the incident wave, as if it were completely unaffected by the hole. In the case of radiation from a disk, if the pressure is constant, the velocity has to be allowed to vary across its radius so that it is no longer rigid. In fact, this produces a hypothetical sound source in which an equal driving force is applied directly to each air particle over the surface represented by the resilient "disk." Unlike a membrane, however, there is no boundary condition of zero displacement at the perimeter. Hence, this model creates some problems since it requires constant pressure to be maintained right up to the outer edge, where there is effectively an acoustic "short circuit" between the front and rear surfaces. Therefore, the reactive particle velocity at the perimeter is infinite, as is also the imaginary admittance.

The disk shown in Fig. 1 lies in the xy plane and a uniform driving force $\widetilde{F}$ is applied in the z direction. The area of a surface element is given by Eq. (2) and the boundary

conditions of Eq. (3), Eq. (4), and Eq. (6) are assumed with $b=a$. Let the surface pressure distribution on the front and rear faces be defined by

$$\tilde{p}_+(w_0) = -\tilde{p}_-(w_0) = -\frac{\tilde{F}}{2S}, \tag{97}$$

where the total surface area S on each face is given by $S = \pi a^2$. Inserting Eq. (97) together with Eq. (14) in Eq. (8) and integrating over the surface area of the disk yields

$$\tilde{p}(w,z) = \frac{a\tilde{F}}{2S_D} \int_0^\infty J_1(\mu a)J_0(\mu w)e^{-i\sigma z}d\mu, \tag{98}$$

where Eq. (22) has been used. The disk velocity $\tilde{u}_0(w)$ can be derived from the boundary condition of Eq. (6) as follows:

$$\tilde{u}_0(w) = \frac{i}{k\rho c} \frac{\partial}{\partial z} \tilde{p}(w,z)\big|_{z=0+}$$

$$= \frac{a\tilde{F}}{2k\rho c S_D} \int_0^\infty J_1(\mu a)J_0(\mu w)\sigma \, d\mu. \tag{99}$$

Integrating the velocity over the area of the disk provides the total volume velocity $\tilde{U}_0$ as follows:

$$\tilde{U}_0 = \int_0^{2\pi}\int_0^a \tilde{u}_0(w)w \, dw \, d\phi = \frac{\pi a^2 \tilde{F}}{k\rho c S_D} \int_0^\infty J_1^2(\mu a)\frac{\sigma}{\mu}d\mu, \tag{100}$$

where Eq. (22) has been used again. The acoustic radiation impedance z_{ar} can now be written as

$$z_{ar} = \frac{\tilde{F}}{S_D\tilde{U}_0} = \frac{2\rho c}{S_D Y_R}, \tag{101}$$

where Y_R is the normalized acoustic radiation *admittance*, which is expressed here as the sum of its real and imaginary parts

$$Y_R = G_R + iB_R, \tag{102}$$

where G_R is the normalized *conductance* given by

$$G_R = \frac{2}{k}\int_0^k \frac{\sqrt{k^2 - \mu^2}}{\mu} J_1^2(\mu a)d\mu, \tag{103}$$

and B_R is the normalized *susceptance* given by

$$B_R = \frac{2}{k}\int_k^\infty \frac{\sqrt{\mu^2 - k^2}}{\mu} J_1^2(\mu a)d\mu. \tag{104}$$

The solution[16] to Eq. (103) is

$$G_R = 1 + \frac{1}{ka}J_1(2ka) - 2J_0(2ka) - \pi[J_1(2ka)\mathbf{H}_0(2ka)$$

$$- J_0(2ka)\mathbf{H}_1(2ka)] \approx \frac{k^2a^2}{6}, \quad ka < 0.5. \tag{105}$$

Unfortunately, the integral in Eq. (104) does not converge and therefore cannot be solved due to the acoustic short circuit at the perimeter mentioned above. However, using the admittance of a disk in an infinite baffle from Sec. III and
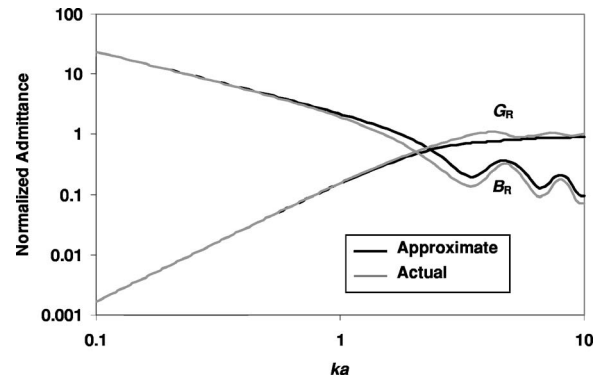
FIG. 19. Normalized radiation admittance of a disk in free space according to historical approximation.

doubling it yields a reasonably good approximation to the actual solution for a rigid disk as follows:

$$B_R = \frac{2ka\mathbf{H}_1(2ka)}{[ka - J_1(2ka))^2 + (\mathbf{H}_1(2ka)]^2} \approx \frac{3\pi}{4ka}, \quad ka < 0.5. \tag{106}$$

The real and imaginary admittances G_R and B_R are plotted in Fig. 19 along with the actual admittances derived from Sec. II for comparison. Also, the impedance may be written as

$$R_R = \frac{G_R}{G_R^2 + B_R^2} \approx \frac{8k^4a^4}{27\pi^2}, \quad ka < 0.5, \tag{107}$$

$$X_R = \frac{B_R}{G_R^2 + B_R^2} \approx \frac{4ka}{3\pi}, \quad ka < 0.5. \tag{108}$$

Again, using Eq. (64) together with Eq. (97) and inserting them into Eq. (63) leads to the following directivity function:

$$D(\theta) = \frac{2J_1(ka \sin\theta)}{ka \sin\theta}\cos\theta. \tag{109}$$

VII. CONCLUSIONS

Referring to Fig. 5, it can be seen that in the case of an unbaffled rigid disk radiating from both sides, the on-axis sound output falls at 6 dB/octave for small values of ka due to the decreasing path difference (as a proportion of wavelength λ) between the antiphase rear radiation and the front radiation, which it partially cancels.

At larger values of ka, the rear radiation from the disk moves in and out of phase with that from the front. However, the comb filter effect is fairly "smeared," the largest peak being 3 dB at $ka=\pi/\sqrt{2}$ or $\lambda=2\sqrt{2}a$. The reason for this is that rear radiation comprises many point sources spread all across the radius of the disk, each with a different path length to the front, so that at no particular frequency do they produce a combined source that is either directly in phase or out of phase with the output from the front.

By contrast, when including a circular baffle and then increasing its size, the actual radiating area decreases in proportion to the total so that it behaves more like a coherent point source at the center. Hence, in the case of $b=4a$, a deep null can be seen at $ka=\pi/2$ or $\lambda=4a$, which is the distance

from the center to the edge. Of course, a disk at the center of a circular baffle is the "worst case," and it would be interesting to compare these results with those of an offset disk in a circular, rectangular, or elliptical baffle, for example, in order to "smear" the path difference effect. It could be postulated that, if the size of the baffle were increased still further, the response would converge towards that of an infinite baffle, where the latter is just a ruler flat line.

At larger values of ka, it can be seen that the near-field pressure responses in finite and infinite baffles converge (see Figs. 4 and 9, respectively). It seems that, as the polar response narrows, obstacles that are placed either side of the sound source have less effect upon the sound field. Also, distinct interference patterns start to emerge due to the presence of a rigid (i.e., opaque) sound source, or obstacle in the case of a scattered incident wave.

The on-axis response of the disk in a closed baffle, shown in Fig. 13, holds few surprises. As would be expected from the averaged response of a disk in a finite baffle and an infinite baffle, the same ripples as in Fig. 5 appear, but are reduced in magnitude. It is fairly well known that a source in a small enclosure produces 6 dB less sound output at low to medium frequencies in free space than when placed against a large plane surface. At high frequencies, of course, the size of the baffle has little effect and the source behaves as though the baffle is infinite. From the curves of Fig. 13, some idea may be gained of how the frequency at which the 6-dB shift occurs is related to the size of the baffle. In the impedance plot of Fig. 14, Crane's results for $b=a$ and $b=8a$ are included for comparison. In the case of $b=a$, the two calculation methods differ by less than 10%. However, when b is increased to $8a$, the real impedance differs by up to 24%, although the imaginary impedance matches within 5%.

Applying the unbaffled curve of Fig. 5 to the sound transmitted through a circular aperture (according to Babinet's principle), it can be seen that the hole is a weak radiator for small ka, as would be expected when considering that as the radiation impedance approaches zero, it increasingly matches that of the rest of the reflective screen. Even what little power does pass through the hole tends to flow back towards the screen. The power transmission coefficient is the real unbaffled impedance of Fig. 3. As Bouwkamp[6] and Spence[7] pointed out, a hole in a rigid screen is a much more effective radiator of power at low frequencies, returning an asymptotic transmission coefficient (or real radiation admittance) of $8/\pi^2$. The latter is a pressure source like the unbaffled resilient disk described in Sec. VI, where the pressure is that of the undisturbed incident wave and could therefore be considered as a resilient disk in an infinite baffle.

It can be seen from Fig. 17 that the approximation to the radiation admittance of a rigid disk in free space is surprisingly close to the actual solution and indeed converges towards it at low frequencies, yielding rational asymptotic expressions for the admittance and impedance. The main difference is that the ripples are missing from the real part at higher frequencies. The real part, however, is the rigorous transmission coefficient of the hypothetical resilient disk, and shows that such a transducer would be a much more effective radiator at low frequencies than a rigid disk, for

which the transmission coefficient is the real impedance. However, the directivity function given by Eq. (109) is perhaps not such a useful approximation except at low frequencies ($ka \leqslant 3$). At high frequencies, it does not converge towards that of a disk in an infinite baffle, as does the actual directivity function of an unbaffled rigid disk.

Finally, when the surface velocity is expressed as a power series function of the radius, it is theoretically possible to accommodate any velocity distribution. This enables Eq. (17) for the velocity distribution to be applied to fluid-structure coupled problems, such as a circular membrane or plate, providing a suitable Green's function is used for the structure. In order to verify the method described in Sec. II, it has been found that Streng's results[14] can be reproduced, but only if the $m=0$ term of Eq. (7) is included. The reason originally stated for omitting this term[10] was that there is neither outward-radial nor axial particle displacement at the perimeter and therefore the radial pressure gradient there is expected to be zero. However, the radial particle displacement at the perimeter is due to the sum of the front and rear radial pressure gradients, which are equal and opposite and therefore cancel. The method has also been verified using FEM. It is hoped that the solutions provided in this paper serve as an illustration of how the symbolic handling capability of modern mathematical software tools may be used to simplify fundamental problems. For example, the conventional equations describing the radiation impedance of a rigid disk in free space or a finite open baffle were considered too complicated to reproduce in most textbooks, whereas Eqs. (47)–(51) and Eqs. (55)–(57) appear to be relatively compact.

ACKNOWLEDGMENT

[1] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), pp. 320, 365, 389, 390.

[2] L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1993), pp. 101–105, 118–128.

[3] H. F. Olson, *Acoustical Engineering* (Professional Audio Journals, Inc., Philadelphia, 1991), pp. 43–46, 92–99.

[4] F. M. Wiener, "On the relation between the sound fields radiated and diffracted by plane obstacles," J. Acoust. Soc. Am. **23**(6), 697–700 (1951).

[5] H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," Phys. Rev. **73**(4), 383–406 (1948).

[6] C. J. Bouwkamp, "Theoretical and numerical treatment of diffraction through a circular aperture," IEEE Trans. Antennas Propag. **AP18**(2), 152–176 (1970).

[7] R. D. Spence, "The diffraction of sound by circular disks and apertures," J. Acoust. Soc. Am. **20**(1), 380–386 (1948).

[8] J. Meixner and U. Fritze, "Das schallfeld in der nähe einer frei schwingenden kolbenmembran (The sound field in the vicinity of a freely oscillating piston diaphragm)," Z. Angew. Phys. **1**, 535–542 (1949).

[9] T. Nimura and Y. Watanabe, "Effect of a finite circular baffle board on acoustic radiation," J. Acoust. Soc. Am. **25**(1), 76–80 (1952).

[10] J. Pachner, "On the acoustical radiation of an emitter vibrating freely or in a wall of finite dimensions," J. Acoust. Soc. Am. **23**(2), 198–208 (1951).

[11] P. H. G. Crane, "Method for the calculation of the acoustic radiation impedance of unbaffled and partially baffled piston sources," J. Sound Vib. **5**(2), 257–277 (1967).

[12] T. J. Mellow, "On the mutual radiation characteristics of two rigid discs in open or closed finite circular baffles," *Joint Baltic–Nordic Acoustics Meet-*

ing 2004, 8–10 June 2004, Mariehamn, Åland, available online at the site ⟨http://www.acoustics.hut.fi/asf/bnam04/webprosari/onlineproc.html⟩

[13]J. H. Streng, "Calculation of the surface pressure on a vibrating circular stretched membrane in free space," J. Acoust. Soc. Am. **82**(2), 679–686 (1987).

[14]J. H. Streng, "Sound radiation from a circular stretched membrane in free space," J. Audio Eng. Soc. **37**(3), 107–118 (1989).

[15]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed., edited by A. Jeffrey (Academic, New York, 2000), pp. 671, Eq. (6.567.1), pp. 668, Eq. (6.561.5), pp. 658, Eq. (6.521.1), pp. 900, Eq. (8.402), pp. 915, Eq. (8.465), pp. 902, Eq. (8.411.1).

[16]S. Wolfram, *The Mathematica Book*, 5th ed. (Wolfram Media, Champaign, Illinois, 2003).

[17]J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. II, pp. 107.

[18]H. Stenzel, "Über die berechnung des schallfeldes einer kreisförmigen kolbenmembran (On the calculation of the sound field of a circular piston diaphragm)," Elektrische Nachrichten-Technik **12**, 16–30 (1935).

[19]L. V. King, "On the acoustic radiation field of the piezoelectric oscillator and the effect of viscosity on the transmission," Can. J. Res. **11**, 135–146 (1934).

[20]J. Babinet, "Mémoires d'optique météorologique (Memoirs on meteorological optics)," C. R. Acad. Sci. Paris, **4**, 638 (1837).

Radiative transfer equation for multiple diffraction

Emeline Reboul, Alain Le Bot,[a] and Joël Perret-Liaudet

*Laboratoire de Tribologie et Dynamique des Systèmes, École Centrale de Lyon, 36,
Avenue Guy de Collongue 69134 Ecully, France*

This paper aims to apply the radiative transfer method to acoustical diffraction by obstacles. Some fictitious sources are introduced at diffracting wedges and a transfer equation based on energy balance determines the diffracted powers. It leads to a set of linear equations on diffracted powers which can be solved in a finite number of steps. It is then possible to calculate the diffracted field anywhere. Some applications to diffraction by obstacles of various shapes are presented. Results of this method are compared with Geometrical Theory of Diffraction and BEM reference calculations. It is shown that this method is particularly efficient in case of multiple diffraction where the ray-tracing technique involves an infinite number of rays between a source and a receiver point. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2001467]

LIST OF SYMBOLS

γ_0 = solid angle of space;
Ω = acoustical domain;
Γ = surface of Ω;
Δ = set of diffracting points of Γ;
c = speed of sound;
m = attenuation factor;
W = energy density;
$\mathbf{I}$ = intensity vector;
G = energy density of direct field;
$\mathbf{H}, H$ = intensity of direct field, magnitude;
D_ω = energetic coefficient of diffraction;
ρ = magnitude of actual sources;
σ = magnitude of diffraction sources;
$\mathbf{u}, \theta, \alpha$ = emission direction, emission angles;
$\mathbf{v}, \varphi, \beta$ = incidence direction, incidence angles;
$\mathbf{s}$ = source point;
$\mathbf{r}$ = receiver point;
$\mathbf{p}, \mathbf{q}$ = diffraction point.

I. INTRODUCTION

The Geometrical Theory of Diffraction (GTD), stated by Keller[1] in the 1960s, leads to an elegant description of diffraction in terms of rays. Starting from a generalized Fermat's principle, GTD states the existence of diffracted rays by wedges and peaks, creeping rays, and many others in addition to the classical direct and reflected rays of geometrical optics. It is then possible to predict the field in the shadow zone of obstacles in a simple way, provided that all paths from the source to the receiver point are listed. The field at the receiver point is simply the sum of all the fields attached to individual rays.

GTD has been successfully applied in acoustics for the problem of single-edge diffraction[2,3] and all problems of diffraction around corners such as buildings or the top of the noise barriers. When the diffracting structure has more than one point of diffraction, waves can be diffracted at each point more than once. This is referred to as multiple diffraction, and can lead to an unlimited number of possible paths for diffracted rays. Pierce[4] gives an approximate expression for double-edge diffraction by a thick, three-sided barrier based on concepts inherent to the GTD. His formulation is particularly well suited for convex-shaped edges, small angles of diffraction, and when the width of the barrier exceeds one wavelength. It was used by Kurze[5] to assess the efficiency of wide barriers. The work of Medwin *et al.*[6] is dedicated to double diffraction in the time domain and is applicable to any multiple diffraction. The technique consists in introducing an infinite number of infinitesimal sources at the first edge, each of which spawns an infinite number of diffractions at the second edge. This method is applied by Ouis[7] to improve the evaluation of the insertion loss of a hard wedged barrier. The multiple diffraction, up to second order, between the top of the wedge barrier and its base is considered. In the same context, Jin *et al.*[8] predict the insertion loss of partially inclined noise barriers considering multiple diffraction occurring at convex as well as concave edges. Formulations for single to triple diffracted waves are constructed based on Kouyoumjian and Pathak's diffraction theory.[9] They show that including additional diffraction at edges improves GTD by removing many small oscillations in the frequency domain. In any case, all these methods take into account a finite number of diffraction order. GTD is particularly well suited in cases where the number of paths is limited or, at least, where all paths can be clearly identified. But, in all other cases, the determination of all source-receiver paths may lead to unsuspected difficulties.

There exists another way to present geometrical acoustics: the so-called "standard procedure" or factor view method in radiative transfer.[10] It is based on a transfer equation which states the equilibrium of energy exchange between two facing surfaces. All rays traveling between these two surfaces, including infinitely reflected ones, are taken into account in the form of an integral equation.

[a] Electronic mail: alain.le-bot@ec-lyon.fr

In acoustics, an application of this method leads to the so-called "radiosity method,"[11–14] which aims at the prediction of reverberation time beyond the validity of Sabine's formula. Applied to the determination of SPL in rooms,[15] vibrational levels in built-up structures[16] or radiation of sound,[17] the radiative transfer approach provides an alternative to both the ray-tracing technique in room acoustics and the statistical energy analysis in vibroacoustics. Related works based on an energy approach include for room acoustics[18] and for noise radiated by structures.[19]

This paper aims to generalize the radiosity method to account for diffraction, that is, to establish the transfer equation for diffraction sources. However, this study is limited to diffraction by wedges and peaks. Diffraction by smooth objects such as cylinders involves creeping waves that are not included in this work. For wedges and peaks, some fictitious diffraction sources are introduced at diffracting vertices. The total diffracted power of these sources is the sum of the diffracted powers of the individual rays. The transfer equation for diffraction sources is established. In some cases of multiple diffraction, i.e., where several diffraction sources interact, it reduces to a linear set of equations on power of diffraction sources. It is then possible to determine the diffracted powers and so, the energy field in the whole acoustical domain. This method allows the solving of multiple diffraction problems in a finite number of steps, to be compared with the infinite number of rays needed in the classical ray-tracing approach. Some results have been published[20] in a short note but limited to the particular two-dimensional case. In this paper, the development of the method is completely detailed including the three-dimensional case.

The outline of the present paper is as follows. In Sec. II all the theoretical material is presented, fictitious diffraction sources are introduced, and the transfer equation which determines their magnitude is derived from power balance. In Sec. III, the transfer equation is solved in the case of a plane wave incident on a three-dimensional wedge, and it is shown that it is consistent with the classical solution of GTD. The diffraction by double wedges is dealt with in Sec. IV. This example highlights the method for solving the transfer equation and the difference with classical ray-tracing algorithms of GTD. A more elaborate example is developed in Sec. V to assess the power of the method.

II. RADIATIVE TRANSFER EQUATION

Let us denote by G the energy density attached to the direct field and by $\mathbf{H}$ the intensity field. Expressions for direct fields depend only on the dimension of the wave but not on the kind of system. For instance, the decrease of energy of a circular wave is the same in vibrating plates and on liquid surfaces. The following expressions are valid for any isotropic and homogeneous medium:

$$G(\mathbf{s},\mathbf{r}) = \frac{1}{\gamma_0 c}\frac{e^{-ms}}{s^{n-1}}, \tag{1}$$

$$\mathbf{H}(\mathbf{s},\mathbf{r}) = \frac{1}{\gamma_0}\frac{e^{-ms}}{s^{n-1}}\mathbf{u}, \tag{2}$$

where $\gamma_0 = 2$ and $n=1$ for plane waves, $\gamma_0 = 2\pi$ and $n=2$ for cylindrical waves, and $\gamma_0 = 4\pi$ and $n=3$ for spherical waves. c is the group speed, $s = |\mathbf{s}-\mathbf{r}|$ the source-receiver distance, and m is the attenuation factor to account for the decrease of wave magnitude when traveling. $H=cG$ denotes the magnitude of intensity vector $\mathbf{H}$, and sometimes, for the sake of brievety, $G(\mathbf{s},\mathbf{r})$ and $H(\mathbf{s},\mathbf{r})$ will be denoted $G(s)$ and $H(s)$ with $s = |\mathbf{s}-\mathbf{r}|$.

In some cases, an obstacle located between the source and the receiver can stop the direct field. The receiver is not affected by the direct field, but can be affected by some reflected or diffracted fields. It is then convenient to introduce a visibility function $V(\mathbf{s},\mathbf{r})$ equal to 1 when $\mathbf{r}$ is visible from $\mathbf{s}$ and 0 otherwise. Final expressions for direct field are

$$G(\mathbf{s},\mathbf{r}) = V(\mathbf{s},\mathbf{r})\frac{1}{\gamma_0 c}\frac{e^{-ms}}{s^{n-1}}, \tag{3}$$

$$\mathbf{H}(\mathbf{s},\mathbf{r}) = V(\mathbf{s},\mathbf{r})\frac{1}{\gamma_0}\frac{e^{-ms}}{s^{n-1}}\mathbf{u}. \tag{4}$$

Thus, direct field is equal to zero in shadow zones of obtacles.

Now, consider the field in a domain Ω induced by some distributed actual sources $\mathbf{s}$ of power density ρ. The direct field is simply the superposition of all individual direct fields, that is, $\int \rho G d\Omega$ where $d\Omega$ is the Lebesgue measure in Ω. Furthermore, waves when traveling are reflected by the surface Γ of the domain and also diffracted by all wedges, peaks, and more generally all singularities of the surface Γ. The set of all diffracting lines is denoted Δ_1 while the set of all diffracting vertices is denoted Δ_0. Some fictitious sources are introduced at all these points where something occurs to incident waves. The power of these fictitious sources is denoted by σ, which may depend on the position of the source but also on the emission direction and time for time-varying problems. In any case, the field radiated by these sources is $\int \sigma G d\Gamma$ for reflection, $d\Gamma$ being the surface measure on Γ, $\int \sigma G d\Delta$ for diffraction by wedges, $d\Delta$ being the length measure on Δ_1 and $\Sigma \sigma G$ for diffraction by peaks, the sum running over the set Δ_0 of all diffracting peaks. Finally, the complete field at any point is obtained by adding the contributions of all types of sources. Therefore, the energy density is at any $\mathbf{r} \in \Omega$ and at time t,

$$W(\mathbf{r},t) = \int_\Omega \rho(\mathbf{s}, t - s/c) G(\mathbf{s},\mathbf{r}) d\Omega_\mathbf{s} + \int_\Gamma \sigma(\mathbf{p},\mathbf{u},t$$
$$- r/c) G(\mathbf{p},\mathbf{r}) d\Gamma_\mathbf{p} + \int_{\Delta_1} \sigma(\mathbf{p},\mathbf{u},t$$
$$- r/c) G(\mathbf{p},\mathbf{r}) d\Delta_\mathbf{p} + \sum_{\mathbf{p} \in \Delta_0} \sigma(\mathbf{p},\mathbf{u},t - r/c) G(\mathbf{p},\mathbf{r}),$$
$$\tag{5}$$

and the intensity is

$$\mathbf{I}(\mathbf{r},t) = \int_{\Omega} \rho(\mathbf{s},t-s/c)\mathbf{H}(\mathbf{s},\mathbf{r})d\Omega_{\mathbf{s}} + \int_{\Gamma} \sigma(\mathbf{p},\mathbf{u},t$$

$$-r/c)\mathbf{H}(\mathbf{p},\mathbf{r})d\Gamma_{\mathbf{p}} + \int_{\Delta_1} \sigma(\mathbf{p},\mathbf{u},t$$

$$-r/c)\mathbf{H}(\mathbf{p},\mathbf{r})d\Delta_{\mathbf{p}} + \sum_{\mathbf{p}\in\Delta_0} \sigma(\mathbf{p},\mathbf{u},t-r/c)\mathbf{H}(\mathbf{p},\mathbf{r}),$$

$$(6)$$

where $s=|\mathbf{r}-\mathbf{s}|$ and $r=|\mathbf{r}-\mathbf{p}|$. $\mathbf{u}$ is the unit vector from the source point $\mathbf{s}$ or $\mathbf{p}$ to the receiver point $\mathbf{r}$ and r/c (or s/c) is the time delay for the propagation of energy from the source point to the receiver point.

Let us turn to the determination of the unknown σ. On the surface Γ where reflection occurs, several reflection laws can be adopted. The case of diffuse reflection has been developed in Ref. 15 and leads to an integral equation on the potential σ which has been solved using a collocation method. Otherwise, a specular law of reflection leads to a functional equation on σ.[21] In all cases where it works, the method of image sources provides a solution of this functional equation.

Consider now the case of diffraction. $D_{\omega}(\mathbf{v},\mathbf{u})$ is an energetic diffraction coefficient depending on two variables, the incident direction $\mathbf{v}$ and the emission direction $\mathbf{u}$. The subscript ω is used to specify that this coefficient also depends on frequency. First, if σ belongs to Δ_0, the energetic coefficient of diffraction $D_{\omega}(\mathbf{v},\mathbf{u})$ is defined as the ratio of the emitted power $d\mathcal{P}_{\text{emit}}$ per unit solid angle du about $\mathbf{u}$ and the incident intensity I_{inc} stemming from $\mathbf{v}$,

$$D_{\omega}(\mathbf{v},\mathbf{u}) = \frac{1}{I_{\text{inc}}} \times \frac{d\mathcal{P}_{\text{emit}}}{du}. \qquad (7)$$

Second, when σ belongs to Δ_1, $D_{\omega}(\mathbf{v},\mathbf{u})$ is defined as the ratio of the emitted power $d\mathcal{P}_{\text{emit}}$ per unit solid angle du about $\mathbf{u}$ and per unit length $d\nu$ of the edge, and the incident intensity I_{inc} stemming from $\mathbf{v}$,

$$D_{\omega}(\mathbf{v},\mathbf{u}) = \frac{1}{I_{\text{inc}}} \times \frac{d\mathcal{P}_{\text{emit}}}{d\nu du}. \qquad (8)$$

When incident energy is distributed among all directions, the emitted power about $\mathbf{u}$ is the sum of contributions of incident intensities leading to the detailed energy balance,[22]

$$\frac{d\mathcal{P}_{\text{emit}}}{du} \text{ or } \frac{d\mathcal{P}_{\text{emit}}}{d\nu du} = \int D_{\omega}(\mathbf{v},\mathbf{u})I_{\text{inc}}(\mathbf{v},t)dv, \qquad (9)$$

where I_{inc} is the incident intensity impinging on the diffracting point $\mathbf{p}$ at time t and stemming from direct, reflected fields and the field diffracted by all other diffracting wedges.

On the other hand, at a distance ϵ of any diffraction point, the emitted intensity is σH, and therefore the flux crossing an infinitesimal solid angle is $d\mathcal{P}_{\text{emit}}=du\sigma H\epsilon^{n-1}$ $=du\sigma e^{-m\epsilon}/\gamma_0$ with the limit $du\sigma/\gamma_0$ for small ϵ. Substitution in Eq. (9) then leads to
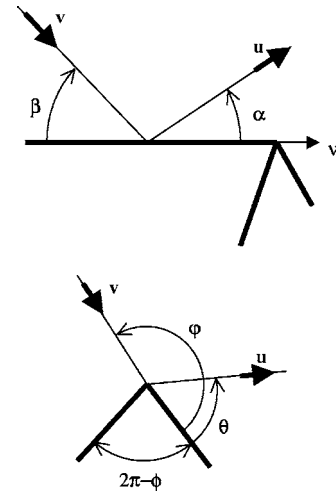
FIG. 1. Plane wave incident on a wedge with direction $\mathbf{v}$. The emission direction is denoted $\mathbf{u}$.

$$\frac{\sigma(\mathbf{p},\mathbf{u},t)}{\gamma_0} = \int_{\Omega} D_{\omega}(\mathbf{v},\mathbf{u})\rho(\mathbf{s},t-s/c)H(\mathbf{s},\mathbf{p})d\Omega_{\mathbf{s}}$$

$$+ \int_{\Gamma} D_{\omega}(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t-r/c)H(\mathbf{q},\mathbf{p})d\Gamma_{\mathbf{q}}$$

$$+ \int_{\Delta_1} D_{\omega}(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t-r/c)H(\mathbf{q},\mathbf{p})d\Delta_{\mathbf{q}}$$

$$+ \sum_{\mathbf{q}\in\Delta_0} D_{\omega}(\mathbf{v},\mathbf{u})\sigma(\mathbf{q},\mathbf{v},t-r/c)H(\mathbf{q},\mathbf{p}). \qquad (10)$$

This is a functional equation on the unknown σ for both Δ_0 and Δ_1. Once the potential σ is known, the field at any receiver point $\mathbf{r}$ is determined by applying Eqs. (5) and (6).

III. DIFFRACTION BY WEDGE

In this section, the problem of a steady-state plane wave diffracted by a wedge is solved using the method exposed in Sec. II. The aim is to check whether the energy solution agrees well with the classical GTD analysis in terms of rays. So, consider a plane wave of intensity I_{inc} impinging on a wedge with incidence $\mathbf{v}=\varphi,\beta$ as shown in Fig. 1. The edge of the wedge is along the ν axis. The outer angle of the wedge is denoted ϕ. For any emission direction $\mathbf{u}=\theta,\alpha$, the diffracted potential σ is given by Eq. (10) where the right-hand side reduces to the direct field term, i.e., the first integral,

$$\frac{\sigma(\nu,\theta,\alpha)}{4\pi} = D_{\omega}(\varphi,\beta;\theta,\alpha)I_{\text{inc}}. \qquad (11)$$

It is well known that all energy with incidence β is diffracted into a cone of axis the edge of the wedge and of half-angle β. This is the so-called Keller's cone. The bidirectional energetic coefficient of diffraction $D_{\omega}(\varphi,\beta;\theta,\alpha)$ then reduces to the particular form,

$$D_{\omega}(\varphi,\beta;\theta,\alpha) = D_{\omega}(\varphi,\theta)\delta(\alpha-\beta). \qquad (12)$$

It is found in the Appendix that $D_{\omega}(\varphi,\theta)=|d(\varphi,\theta)|^2$ is the energetic coefficient of diffraction at normal incidence,
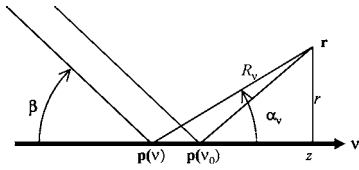
FIG. 2. The emission angle α_ν depends on the receiver point **r**. There is a single point $\mathbf{p}(\nu_0)$ where the emission angle equals the incidence angle β.

$d(\varphi, \theta)$ being the classical diffraction coefficient used in GTD. The energetic diffraction coefficient remains unchanged when incidence and emission directions are interchanged following the reciprocity principle,

$$D_\omega(\varphi, \theta) = D_\omega(\theta, \varphi). \tag{13}$$

Furthermore, the symmetry with respect to the medium plan of the wedge implies that the coefficient is invariant under the change of angle $\zeta \to \phi - \zeta$,

$$D_\omega(\varphi, \theta) = D_\omega(\phi - \varphi, \phi - \theta). \tag{14}$$

Now, introducing Eqs. (11) and (12) into the third integral of Eq. (5) leads to the diffracted energy at any point **r** $= r, \theta, z$ in cylindrical coordinates centered on the ν axis,

$$W_{\text{dif}} = \int_{-\infty}^{\infty} \sigma(\nu, \theta, \alpha_\nu) G(\mathbf{p}, \mathbf{r}) d\nu$$

$$= \frac{I_{\text{inc}}}{c} \int_{-\infty}^{\infty} D_\omega(\varphi, \theta) \delta(\alpha_\nu - \beta) \frac{d\nu}{R_\nu^2}, \tag{15}$$

where θ, α_ν are the emission angles at point **p** of coordinate ν on the edge toward **r** and R_ν is the distance between **p** and **r**. To evaluate this last integral, the following result of the theory of distributions is used;

$$\int g(x) \delta[f(x)] dx = \sum_i \frac{g(x_i)}{|f'(x_i)|}, \tag{16}$$

where the sum runs over all zeros of f included in the range of integration. It is found that the function $f(\nu) = \alpha_\nu - \beta = \arctan[r/(z-\nu)] - \beta$ has a single zero $\nu_0 = z - r/\tan\beta$ which corresponds to the unique point for which emission angle α_0 toward **r** equals the incidence angle β (Fig. 2). Then,

$$f'(\nu_0) = \frac{d\alpha_\nu}{d\nu} = \frac{1}{1 + \frac{r^2}{(z-\nu)^2}} \times \frac{d}{d\nu}\left(\frac{r}{z-\nu}\right) = \frac{r}{R_\nu^2}. \tag{17}$$

Finally, the diffracted energy is

$$W_{\text{dif}} = \frac{I_{\text{inc}}}{c} D_\omega(\varphi, \theta) \times \frac{1}{r}, \tag{18}$$

that is, the energy is inversely proportional to the wedge-receiver distance r. The decrease of the energy of the field like $1/r$ well agrees with the decrease of the magnitude of the field like $1/\sqrt{r}$ predicted by GTD.[2]

A numerical simulation is proposed to compare the results given by the radiative transfer method to those given by the GTD. The edge outer angle is $\phi = 7\pi/4$ rad. An incident plane wave hits the wedge and gives rise to a reflected wave, a diffracted wave, and a surface ray propagating along the
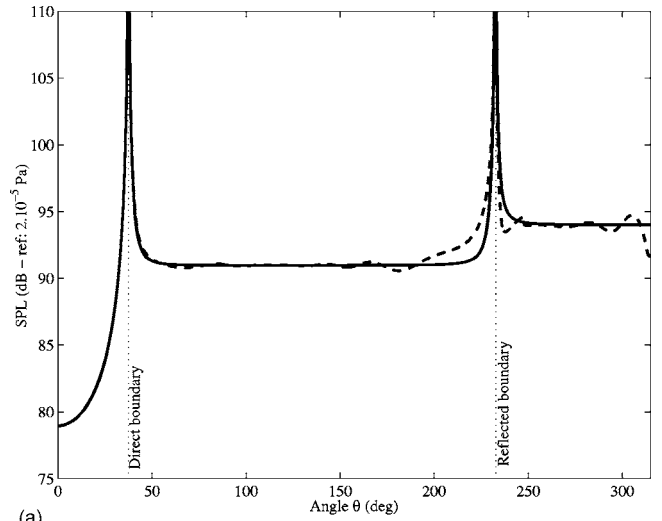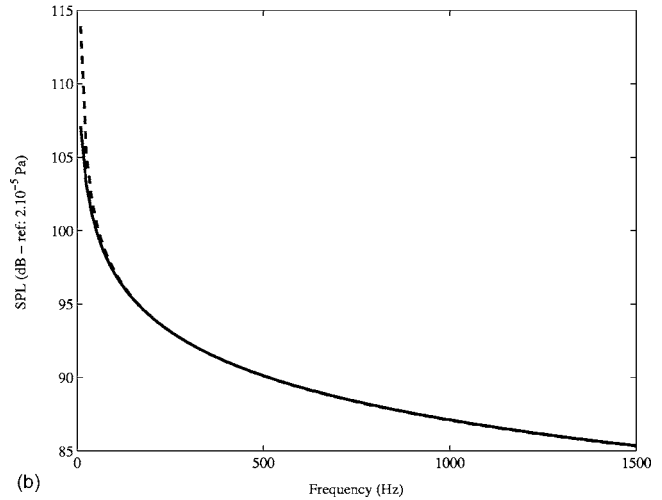
FIG. 3. Diffraction by a wedge of a plane wave with incidence $\varphi = 217.5°$. (a) Comparison of the SPL vs θ at $r = 1.5$ m obtained from the radiative equation (—) and GTD (---). The frequency is 1.5 kHz. The boundaries of direct field and diffracted field are, respectively, $\theta = \varphi - \pi$ and $\theta = 2\phi - \varphi - \pi$. (b) Comparison of the SPL vs frequency at $r = 1.5$ m and $\theta = 27.5°$ obtained from the radiative equation (—) and GTD (---). The receiver point is in the shadow area.

wedge. The receiver point is assumed to be far enough from the edge so that surface rays can be neglected. Two results are shown in Fig. 3. On the one hand, the behavior of the acoustical field around the wedge, at a constant distance $r = 1.5$ m from the edge, is represented in a plane perpendicular to the edge of the wedge. The frequency is 1.5 kHz. The incident direction is $\varphi = 217.5°$ from the right side of the wedge, $\beta = 40°$ (Fig. 1). In this problem, there are two boundary lines that, respectively, mark the transition from illumination to shadow for the incident field ($\theta = \varphi - \pi$) and for the reflected field ($\theta = 2\phi - \pi - \varphi$). GTD is known to fail near these geometrical boundaries, and so the radiative transfer method which uses the diffraction coefficient of the GTD also gives inaccurate results in these areas. Thus, both methods lead to similar results all around the wedge including these transition areas where they both lead to unphysical results. On the other hand, the same calculation is performed for a fixed position of the receiver point $r = 1.5$ m, $\theta = 27.5°$ in the shadow area of the wedge, but versus frequency. Both
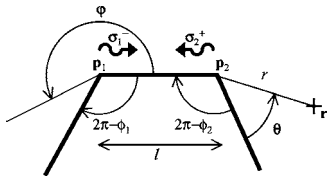
FIG. 4. Plane wave incident on a double wedge of width l with incidence angle φ. Two fictitious sources are located at $\mathbf{p}_1$ and $\mathbf{p}_2$.

methods are again in good agreement and show that the amplitude of the diffracted field decreases as the frequency increases.

IV. DIFFRACTION BY DOUBLE WEDGE

Let us now arrive to multiple diffraction. The simple case of double diffraction of a steady plane wave is solved by using GTD and Eq. (10). This example is chosen in order to illustrate similarities and differences of GTD and Eq. (10).

Consider a two-dimensional rigid three-sided barrier. An incident plane wave with intensity I_{inc} is impinging on the left-hand side, and one takes an interest in the evaluation of the energy density on the other side of the barrier. Let us introduce two fictitious diffraction sources at corners $\mathbf{p}_1$ and $\mathbf{p}_2$ of powers σ_1 and σ_2. The incidence angle φ and the emission θ are defined in Fig. 4. For the sake of simplicity, $\sigma_1(0)$ and $\sigma_2(\phi_2)$ are denoted σ_1^- and σ_2^+. Equations on σ_1^- and σ_2^+ are obtained by applying Eq. (10):

$$\frac{\sigma_1^-}{2\pi} = I_{inc}D_\omega(\varphi,0) + \sigma_2^+ H(l)D_\omega(0,0), \quad (19)$$

$$\frac{\sigma_2^+}{2\pi} = \sigma_1^- H(l)D_\omega(\phi_2,\phi_2), \quad (20)$$

where l is the width of the barrier, ϕ_i is the outer angle of corner $\mathbf{p}_i$. It has been assumed that the incident wave propagates upwards, i.e., $\varphi > \pi$. Otherwise a term for direct intensity must be added in the second equation. Applying again Eq. (10) leads to σ_2 at any emission angle θ,

$$\frac{\sigma_2(\theta)}{2\pi} = \sigma_1^- H(l)D_\omega(\phi_2,\theta). \quad (21)$$

By using the relationship $D_\omega(\phi_2,\phi_2)=D_\omega(0,0)$ from Eq. (14), the solution of Eqs. (19) and (20) is

$$\sigma_1^- = I_{inc}\frac{2\pi D_\omega(\varphi,0)}{1 - 4\pi^2 H^2(l)D_\omega^2(0,0)}. \quad (22)$$

And then,

$$\sigma_2(\theta) = I_{inc}\frac{4\pi^2 D_\omega(\varphi,0)H(l)D_\omega(\phi_2,\theta)}{1 - 4\pi^2 H^2(l)D_\omega^2(0,0)}. \quad (23)$$

The field at any point $\mathbf{r}$ on the right side of the barrier below the top, i.e., $\theta < \phi_2 - \pi$ is given by Eq. (5) with the only contribution of the diffracted field from $\mathbf{p}_2$:

$$W(\mathbf{r}) = \sigma_2(\theta)G(r), \quad (24)$$

where $r = |\mathbf{p}_2 - \mathbf{r}|$. Thus,

$$W(\mathbf{r}) = \frac{I_{inc}}{c}\frac{D_\omega(\varphi,0)D_\omega(\phi_2,\theta)}{1 - \dfrac{D_\omega^2(0,0)}{l^2}} \times \frac{1}{lr}, \quad (25)$$

by assuming that the attenuation factor is zero, that is $H(l) = 1/2\pi l$ and $G(r)=1/2\pi cr$. This expression for energy is valid under the condition that $D_\omega(0,0)/l < 1$.

Let us turn to the same problem by means of the GTD. The pressure p_2 at point $\mathbf{p}_2$ is determined by summing all contributions of rays arriving at this point and can be written as

$$p_2 = p_0 D_\omega(\varphi,0)\frac{e^{-ikl}}{\sqrt{l}}\sum_{n=0}^{\infty}\left[d(0,0)\frac{e^{-ikl}}{\sqrt{l}}\right]^{2n}, \quad (26)$$

where p_0 is the pressure of the incident plane wave at $\mathbf{p}_1$. The first term of the infinite sum ($n=0$) is a single diffracted wave, that is, the incident plane wave diffracted once at point $\mathbf{p}_1$ before reaching $\mathbf{p}_2$. The second term ($n=1$) is a triple diffracted wave, that is, the incident plane wave first diffracted at point $\mathbf{p}_1$ and then diffracted at point $\mathbf{p}_2$, being again diffracted at point $\mathbf{p}_1$ before reaching $\mathbf{p}_2$. Other terms are for a greater number of round trips between $\mathbf{p}_1$ and $\mathbf{p}_2$. The diffracted pressure at point $\mathbf{r}$ is

$$p(\mathbf{r}) = p_2 d(\phi_2,\theta)\frac{e^{-ikr}}{\sqrt{r}} \quad (27)$$

$$= p_0 d(\varphi,0)d(\phi_2,\theta)\frac{e^{-ikl}}{\sqrt{l}}\frac{e^{-ikr}}{\sqrt{r}}\sum_{n=0}^{\infty}\left[\left(d(0,0)\frac{e^{-ikl}}{\sqrt{l}}\right)^2\right]^n, \quad (28)$$

where $d(\phi_2,\theta)e^{-ikr}/\sqrt{r}$ is the cylindrical acoustical wave from the corner $\mathbf{p}_2$ to the observation point $\mathbf{r}$. It is instructive to determine the energy $W(\mathbf{r})=|p(\mathbf{r})|^2/2\rho c$ of this field. By squaring Eq. (28) and neglecting all cross-product terms, it yields

$$W(\mathbf{r}) = \frac{|p_0|^2}{2\rho c^2}|d(\varphi,0)|^2|d(\phi_2,\theta)|^2\frac{1}{lr}\sum_{n=0}^{\infty}\left[|d(0,0)|^2\frac{1}{l}\right]^{2n} \quad (29)$$

$$= \frac{I_{inc}}{c}\frac{D_\omega(\varphi,0)D_\omega(\phi_2,\theta)}{1 - \dfrac{D_\omega^2(0,0)}{l^2}} \times \frac{1}{lr}, \quad (30)$$

where $I_{inc}=|p_0|^2/2\rho c$ and $D_\omega=|d|^2$. The series converges when $D_\omega(0,0)/l < 1$. It is then apparent that GTD leads to the same energy field as predicted by Eq. (10) provided that all interferences between rays have been neglected. The set of Eqs. (5), (6), and (10) can thus be considered as an extension of geometrical acoustics in its strict sense, that is, with no phase attached to rays, which includes diffraction effects as done in GTD.

V. DIFFRACTION BY RECTANGLE

Consider a steady-state cylindrical wave of power ρ_0 diffracted by a rectangular obstacle of dimensions $a \times b$ as shown in Fig. 5. The incident wave is reflected on the front wall and diffracted by the four corners denoted by $\mathbf{p}_i$ with $i=1,\ldots,4$. Thus, the direct field is $\rho_0 G(\mathbf{s},\mathbf{r})$ at any point $\mathbf{r}$

Reboul, Le Bot, and Perret-Liaudet: Transfer equation for diffraction
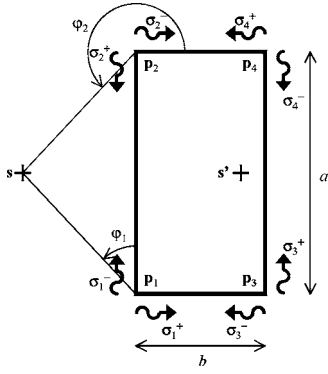
FIG. 5. Diffraction of a cylindrical wave emanating from **s** by a rectangle $a \times b$. Four fictitious sources of diffraction $\mathbf{p}_i$ are localized at the corners with powers $\sigma_i^{\pm}$ in grazing directions. The image source for reflected rays on the left side is **s'**.

and the reflected field is $\rho_0 G(\mathbf{s}', \mathbf{r})$ where **s'** is the image source given by reflection of **s** on the front wall.

At each diffracting corner is put a fictitious source $\sigma_i(\theta)$ whose particular values at angles $\theta=0$ and $\theta=3\pi/2$ are denoted σ_i^- and σ_i^+. Equations on $\sigma_i^{\pm}$ are obtained by applying Eq. (10). No reflected ray can be diffracted and therefore the second integral of Eq. (10) vanishes. Finally, $\sigma_i^{\pm}$ is only related to direct intensity $\rho_0 H(s)$ for points 1 and 2 and other values of $\sigma_j^{\mp}$,

$$\frac{\sigma_1^+}{2\pi} = \sigma_2^+ D_\omega\left(0, \frac{3\pi}{2}\right) H(a) + \sigma_3^- D_\omega\left(\frac{3\pi}{2}, \frac{3\pi}{2}\right) H(b)$$
$$+ \rho_0 D_\omega\left(\varphi_1, \frac{3\pi}{2}\right) H(s_1), \tag{31}$$

$$\frac{\sigma_1^-}{2\pi} = \sigma_2^+ D_\omega(0,0) H(a) + \sigma_3^- D_\omega\left(\frac{3\pi}{2}, 0\right) H(b)$$
$$+ \rho_0 D_\omega(\varphi_1, 0) H(s_1), \tag{32}$$

$$\frac{\sigma_2^+}{2\pi} = \sigma_1^- D_\omega\left(\frac{3\pi}{2}, \frac{3\pi}{2}\right) H(a) + \sigma_4^+ D_\omega\left(0, \frac{3\pi}{2}\right) H(b)$$
$$+ \rho_0 D_\omega\left(\varphi_2, \frac{3\pi}{2}\right) H(s_2), \tag{33}$$

$$\frac{\sigma_2^-}{2\pi} = \sigma_1^- D_\omega\left(\frac{3\pi}{2}, 0\right) H(a) + \sigma_4^+ D_\omega(0,0) H(b)$$
$$+ \rho_0 D_\omega(\varphi_2, 0) H(s_2), \tag{34}$$

$$\frac{\sigma_3^+}{2\pi} = \sigma_1^+ D_\omega\left(0, \frac{3\pi}{2}\right) H(b) + \sigma_4^- D_\omega\left(\frac{3\pi}{2}, \frac{3\pi}{2}\right) H(a), \tag{35}$$

$$\frac{\sigma_3^-}{2\pi} = \sigma_1^+ D_\omega(0,0) H(b) + \sigma_4^- D_\omega\left(\frac{3\pi}{2}, 0\right) H(a), \tag{36}$$

$$\frac{\sigma_4^+}{2\pi} = \sigma_2^- D_\omega\left(\frac{3\pi}{2}, \frac{3\pi}{2}\right) H(b) + \sigma_3^+ D_\omega\left(0, \frac{3\pi}{2}\right) H(a), \tag{37}$$

$$\frac{\sigma_4^-}{2\pi} = \sigma_3^+ D_\omega(0,0) H(a) + \sigma_2^- D_\omega\left(\frac{3\pi}{2}, 0\right) H(b), \tag{38}$$

where s_1, s_2, φ_1, and φ_2 are defined in Fig. 5. This is a set of linear equations on the eight unknowns $\sigma_i^{\pm}$. Furthermore, the potential σ_i in all other directions θ is also given by Eq. (10),

$$\frac{\sigma_1(\theta)}{2\pi} = \sigma_2^+ D_\omega(0, \theta) H(a) + \sigma_3^- D_\omega\left(\frac{3\pi}{2}, \theta\right) H(b)$$
$$+ \rho_0 D_\omega(\varphi_1, \theta) H(s_1), \tag{39}$$

$$\frac{\sigma_2(\theta)}{2\pi} = \sigma_1^- D_\omega\left(\frac{3\pi}{2}, \theta\right) H(a) + \sigma_4^+ D_\omega(0, \theta) H(b)$$
$$+ \rho_0 D_\omega(\varphi_2, \theta) H(s_2), \tag{40}$$

$$\frac{\sigma_3(\theta)}{2\pi} = \sigma_1^+ D_\omega(0, \theta) H(b) + \sigma_4^- D_\omega\left(\frac{3\pi}{2}, \theta\right) H(a), \tag{41}$$

$$\frac{\sigma_4(\theta)}{2\pi} = \sigma_2^- D_\omega\left(\frac{3\pi}{2}, \theta\right) H(b) + \sigma_3^+ D_\omega(0, \omega) H(a). \tag{42}$$

Thus, the functions $\sigma_i(\theta)$ are well defined by the particular values $\sigma_i^{\pm}$.

Finally, the field at any point **r** is given by adding the contributions of the direct field, the reflected field and the four diffracted fields. It yields

$$W(\mathbf{r}) = \rho_0 G(\mathbf{s}, \mathbf{r}) + \rho_0 G(\mathbf{s}', \mathbf{r}) + \sum_{i=1}^{4} \sigma_i(\theta_i) G(\mathbf{p}_i, \mathbf{r}), \tag{43}$$

$$\mathbf{I}(\mathbf{r}) = \rho_0 \mathbf{H}(\mathbf{s}, \mathbf{r}) + \rho_0 \mathbf{H}(\mathbf{s}', \mathbf{r}) + \sum_{i=1}^{4} \sigma_i(\theta_i) \mathbf{H}(\mathbf{p}_i, \mathbf{r}), \tag{44}$$

where θ_i is the emission angle at point $\mathbf{p}_i$ toward **r**. Remember that only some of the six fields can reach the point **r** depending on its position, the others being stopped by the obstacle.

Two numerical simulations are proposed in Figs. 6 and 7. The size of the rectangle is 10 m × 2 m and the source point is located at a distance of 4 m on the left of the rectangle and on the middle axis. In Fig. 6, the frequency range of the sound source is the third octave band centered on 250 Hz, and the wavelength of sound is about 1.4 m, which is not very different from the smaller dimension of the diffracting object. In Fig. 7, the frequency range of the sound source is the third octave band centered on 5000 Hz, and the wavelength of sound is about 0.07 m, which is very small with regards to the size of the diffracting object. For both cases, two results are provided, on the one hand the SPL from Eq. (43) is plotted at the centered frequency and on the other hand, the BEM is employed. For BEM studies, calculations are performed every 25 Hz for the third-octave band centered on 5000 Hz and every 5 Hz for the third-octave band centered on 250 Hz, and the final result is the quadratic average of these, or, in other words, is the rms value.

For both cases, the transfer equation leads to infinite results in the vicinity of the boundaries between illuminated and shadow zones. These boundaries are visible on the SPL
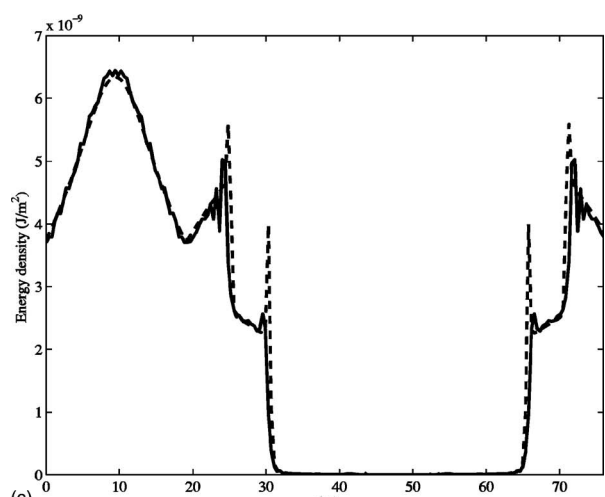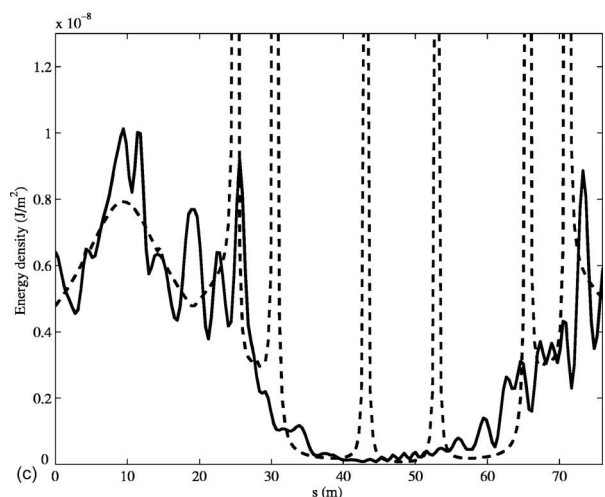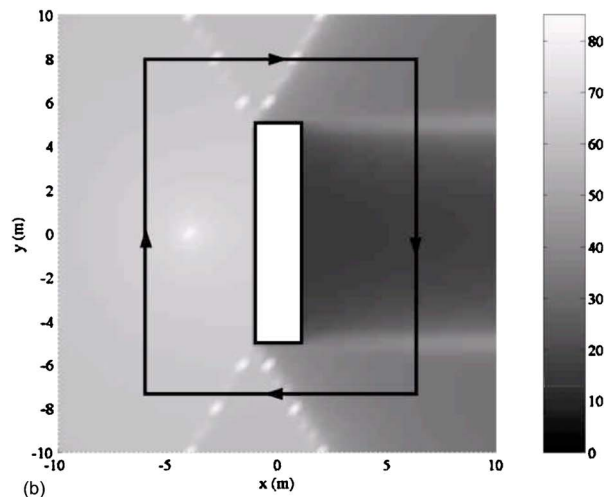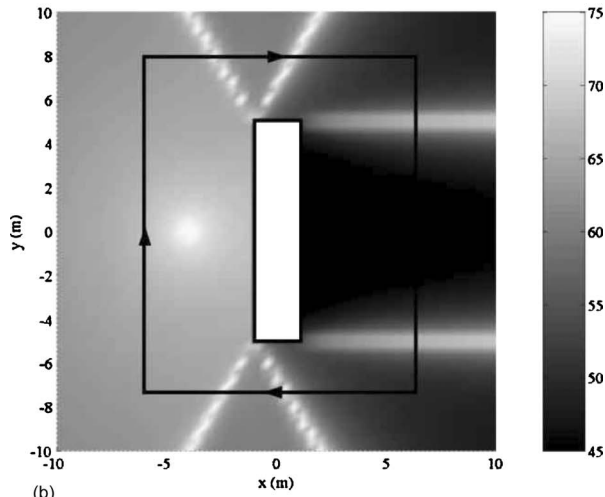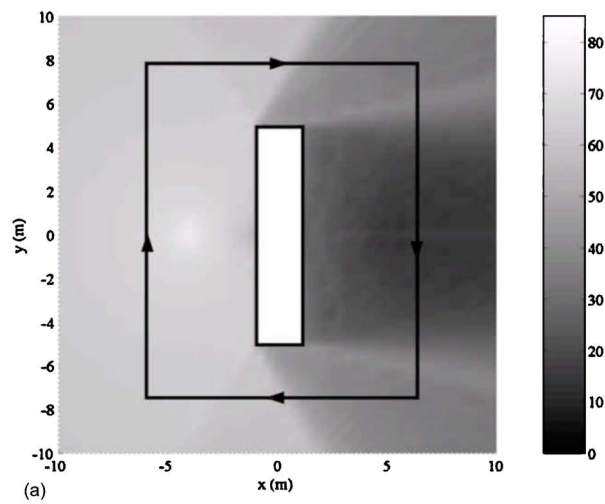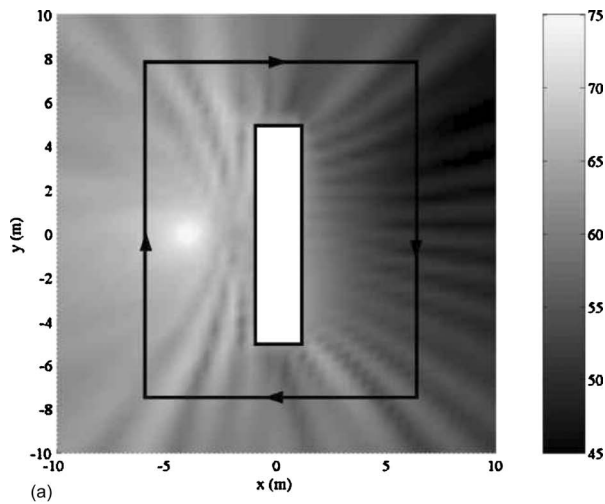
FIG. 6. Diffraction of a cylindrical wave by a rectangular obstacle (10 m × 2 m). SPL in dB around the obstacle predicted by (a) the boundary element method and (b) the transfer equation. (c) Energy density along the contour line beginning at the left bottom corner, with the boundary element method (—) and with the transfer equation (---). The frequency is 250 Hz.

FIG. 7. Diffraction of a cylindrical wave by a rectangular obstacle (10 m × 2 m). SPL in dB around the obstacle predicted by (a) the boundary element method and (b) the transfer equation. (c) Energy density along the contour line beginning at the left bottom corner, with the boundary element method (—) and with the transfer equation (---). The frequency is 5000 Hz.

map in Fig. 6(b) and in Fig. 7(b) and where they appear as white areas. Oblique boundaries correspond to the limit of influence of direct and reflected fields while others correspond to the limit of influence of fields diffracted by the opposite corner. It is well known that in the vicinity of these

boundaries the acoustical field cannot be developed in terms of rays. This is the same limitation as encountered in GTD since the diffraction coefficient implemented in this calculation is the one of GTD.

The main difference between the result given by the BEM and the result from the transfer equation is that the

fluctuations due to the existence of some remaining interferences are noticeable on the BEM result, while they do not appear on the result from the transfer equation which is based on the assumption that direct, reflected, and diffracted waves are uncorrelated. These interferences have a major effect when the wavelength of sound is close to the size of the diffracting object as shown in Fig. 6(a), and in this case, the result given by the transfer equation only represents the average behavior of the acoustical field around the rectangle [Fig. 6(c)]. However, when the frequency increases, or more generally, when the ratio between the wavelength and the size of the diffracting object decreases, the effect of interferences vanishes and, thus, both methods well agree as shown in Figs. 7(a)–7(c).

VI. DISCUSSION

One of the main advantages of the radiative transfer method for the prediction of sound fields is that computation times are significantly reduced compared to some other methods such as the BEM. The calculations absolutely do not depend on the wavelength of sound if parameters such as atmospheric or material absorption coefficients do not depend on frequency. Otherwise, the variations of these parameters define the frequency step to use in order to apply the method. Thus, in the case of diffraction, the energetic diffraction coefficient depends on frequency: the magnitude of the diffraction coefficient introduced in GTD decreases like $1/\sqrt{k}$ so the energetic diffraction coefficient decreases like $1/k$, where k is the wave number. The allowed frequency bandwidth for the study depends on the frequency variations of the energetic diffraction coefficient: the frequency step is chosen so that the diffraction coefficient does not vary much on the bandwith. Thus, the change in the nature of the acoustical field will be well described by the radiative transfer method, except for interference effects. However, as any ray approach, this method cannot give accurate results for low frequencies where the wavelength is of the same order as the typical length of the problem. In such cases, wave-based approaches should be employed. For instance, the typical lengths for the example of Sec. III (Figs. 1 and 3) are the source-edge and edge-receiver distances. Ray theory then applies provided that these distances are larger than one wavelength beyond the near-field influence. In the example of Sec. IV (Fig. 4), the distance between the two diffracting lines must also be larger than one wavelength whereas in the example of Sec. V (Figs. 5 and 7) length and width of the rectangle must respect the same criteria. The method would fail for arbitrarily small diffracting objects (Fig. 6).

VII. CONCLUSION

In this paper, it has been proposed to apply the radiative transfer approach to diffraction. The functional equation (10) gives the power of fictitious sources of diffraction. The interest of such an equation is to determine the total diffracted power in any direction. All well-known results on diffraction by wedges in the frame of GTD are recovered. Equation (10) leads to a set of linear equations. By solving this system, it is possible to determine the total diffracted power by corners, without following the rays as is usually done with the ray-tracing technique.

The method has two main limitations. First, some unphysical results are obtained in the vicinity of boundaries between illuminated and shadow zones. This is the same limitation as in GTD for same reason. Second, interference effects cannot be described by the present method because energies are added. This is an additional limitation which does not occur in GTD.

Compared with BEM this method is CPU-time saving. BEM is limited in practice to low frequencies. In the high frequency range, the number of fictitious sources introduced on the mesh is rapidly prohibitive in regard to current computers. On the other hand, the number of fictitious sources required in the transfer equation (10) does not depend on the frequency, which leads to a constant CPU time at any frequency. This provides an alternative method in the field of high frequency diffraction.

Furthermore, this approach is better suited than GTD to the case of multiple diffraction. All orders of diffraction are simultaneously taken into account and thus, in practice, results are more accurate than using GTD where a limited number of rays are used. Besides, it does not require the knowledge of all ray paths between the source and the receiver point which can be tedious to determine in practice.

APPENDIX

This Appendix concerns the evaluation of the energetic diffraction coefficient D_ω. Consider an incident ray in the direction $\mathbf{v} = \varphi, \beta$ impinging on a wedge at point $\mathbf{q}$. The diffracted field at any point $\mathbf{r}$ in the direction $\mathbf{u} = \theta, \alpha$ is given by

$$p_{\text{dif}}(\mathbf{r}) = p_{\text{inc}}(\mathbf{q}) d_\alpha(\varphi, \theta) \frac{e^{-ikR}}{\sqrt{R}}, \tag{A1}$$

where k is the wave number, $p_{\text{inc}}(\mathbf{q})$ is the incident field at the diffraction point $\mathbf{q}$, and $R = |\mathbf{q} - \mathbf{r}|$ is the distance measured along the diffracted ray. The diffraction coefficient d_α depends on φ and θ the incidence and emission directions defined to the right side of the wedge (Fig. 1) and also on the outer angle of the wedge ϕ by means of the wedge index $\nu = \pi / \phi$. According to Keller's law of diffraction[1] the angle of diffraction α equals the angle of incidence β so that the diffracted rays emanating from $\mathbf{q}$ belong to a cone whose axis is the diffracting edge. An expression for d_α is derived using the method of steepest descent,[2,4]

$$d_\alpha(\varphi,\theta) = \frac{e^{-i\pi/4}\nu\sin(\nu\pi)}{\sqrt{2\pi k\sin\alpha}}\left[\frac{1}{\cos(\nu\pi) - \cos[\nu(\theta+\varphi)]}\right.$$

$$\left. + \frac{1}{\cos(\nu\pi) - \cos[\nu(\theta-\varphi)]}\right]. \quad (A2)$$

d_α becomes singular on incidence or reflection boundaries. This coefficient depends on the frequency through the wave number k.

The diffracted intensity $I_{\text{dif}}(\mathbf{r}) = |p_{\text{dif}}|^2/2\rho c$ where c is the sound speed, ρ is the fluid density, is from Eq. (A1)

$$I_{\text{diff}}(\mathbf{r}) = \frac{I_{\text{inc}}(\mathbf{q})}{R}|d_\alpha(\varphi,\theta)|^2, \quad (A3)$$

with $I_{\text{inc}}(\mathbf{q}) = |p_{\text{inc}}(\mathbf{q})|^2/2\rho c$ the incident intensity. By denoting $r = R\sin\alpha$ the edge-receiver distance, Eq. (A3) is written by means of the diffraction coefficient $d(\varphi,\theta)$ at normal incidence $\alpha = \pi/2$,

$$I_{\text{dif}}(\mathbf{r}) = \frac{I_{\text{inc}}(\mathbf{q})}{r}|d(\varphi,\theta)|^2. \quad (A4)$$

By comparing with Eq. (18) and noting that $I_{\text{dif}} = cW_{\text{dif}}$, it is readily found that $D_\omega(\varphi,\theta) = |d(\varphi,\theta)|^2$. This is the energetic diffraction coefficient at normal incidence.

[1]J. B. Keller, "Geometrical theory of diffraction," J. Opt. Soc. Am. **62**, 116–130 (1962).

[2]A. D. Pierce, *Acoustics—An Introduction to Its Physical Principles and Applications*, 2nd printing (Acoustical Society of America, American Institute of Physics, New York, 1991).

[3]W. J. Hadden and A. D. Pierce, "Sound diffraction around screens and wedges for arbitrary point source locations," J. Acoust. Soc. Am. **71**, 1290 (1982).

[4]A. D. Pierce, "Diffraction around corners and over barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).

[5]U. J. Kurze, "Noise reduction by barriers," J. Acoust. Soc. Am. **55**, 504–518 (1974).

[6]H. Medwin, E. Childs, and G. M. Jebsen, "Impulse studies of double diffraction: A discrete Huyghens interpretation," J. Acoust. Soc. Am. **72**, 1005–1013 (1982).

[7]D. Ouis, "Noise attenuation by a hard wedge-shaped barrier," J. Sound Vib. **262**, 347–364 (2003).

[8]B. J. Jin, H. S. Kim, H. J. Kang, and J. S. Kim, "Sound diffraction by a partially inclined noise barrier," Appl. Acoust. **62**, 1107–1121 (2001).

[9]R. G. Kouyoumjian and P. H. Pathak, "A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface," Proc. IEEE **62**, 1448–1461 (1974).

[10]M. F. Modest, *Radiative Heat Transfer* (McGraw-Hill, New York, 1993).

[11]H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, London, 1991).

[12]M. Carroll and C. Chien, "Decay of reverberent sound in a spherical enclosure," J. Acoust. Soc. Am. **62**, 1442–1446 (1977).

[13]R. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," J. Sound Vib. **92**, 203–226 (1984).

[14]H. Kuttruff, "Energetic sound propagation in rooms," Acust. Acta Acust. **83**, 622–628 (1997).

[15]A. Le Bot and A. Bocquillet, "Comparison of an integral equation on energy and the ray-tracing technique for room acoustics," J. Acoust. Soc. Am. **108**, 1732–1740 (2000).

[16]A. Le Bot, "Energy transfer for high frequencies in built-up structures," J. Sound Vib. **250**, 247–275 (2002).

[17]V. Cotoni and A. Le Bot, "Radiation of plane structures at high frequency using an energy method," Int. J. Acoust. Vib. **6**, 209–214 (2001).

[18]L. P. Franzoni, D. B. Bliss, and J. W. House, "An acoustic boundary element method based on energy and intensity variables for prediction of high frequency broadband sound fields," J. Acoust. Soc. Am. **110**, 3071–3080 (2001).

[19]A. Wang, N. Vlahopoulos, and K. Wu, "Development of an energy boundary element formulation for computing high-frequency sound radiation from incoherent intensity boundary conditions," J. Sound Vib. **278**, 413–436 (2004).

[20]E. Reboul, A. Le Bot, and J. Perret-Liaudet, "Introduction of acoustical diffraction in the radiative transfer method," C. R. Mec. **332**, 505–511 (2004).

[21]A. Le Bot, "A functional equation for the specular reflection of rays," J. Acoust. Soc. Am. **112**, 1276–1287 (2002).

[22]W. Joyce, "Sabine's reverberation time and ergodic auditoriums," J. Acoust. Soc. Am. **58**, 643–655 (1975).

Compact nonporous windscreen for infrasonic measurements

Qamar A. Shams,[a] Allan J. Zuckerwar, and Bradley S. Sealey

NASA Langley Research Center, Mail Stop 238, Hampton, Virginia 23681

Infrasonic windscreens, designed for service at frequencies below 20 Hz, were fabricated from a variety of materials having a low acoustic impedance, and tested against four specifications (the first three in a small wind tunnel): (1) wind-generated noise reduction ("insertion loss") at a free-stream wind speed of 9.3 m/s, (2) transmission of low-frequency sound from a known source (subwoofer), (3) spectrum of sound generated from trailing vortices (aeolian tones), and (4) water absorption (to determine suitability for all-weather service). The operating principle is based on the high penetrating capability of infrasound through solid barriers. Windscreen materials included three woods (pine, cedar, and balsa), closed-cell polyurethane foam, and Space Shuttle tile material. The windscreen inside diameter ranged from 0.0254 to 0.1016 m (1 to 4 in.), and wall thickness from 0.003175 to 0.01905 m $\left(\frac{1}{8} \text{ to } \frac{3}{4} \text{ in.}\right)$. A windscreen made of closed-cell polyurethane foam revealed a wind noise reduction of 10–20 dB from 0.7 to 25 Hz, transmission coefficient near unity from 10 to 20 Hz, and spectral peaks beyond 20 Hz due to vortex-generated sound. Following a description of past methods, the principle of operation, and the experimental method, experimental data are presented for a variety of windscreens. [DOI: 10.1121/1.1992707]

I. INTRODUCTION

A fundamental difficulty in the detection of outdoor infrasound is the so-called "wind noise" problem. Because of ever-present wind-generated turbulence, the atmosphere is inherently noisy at frequencies of interest (typically between 0.1 and 20 Hz). Accordingly, effective wind screening is vital to the success of outdoor infrasonic measurements. Past methods of screening a microphone from the wind include a piped array, a barrier, or an open-mesh (e.g., cloth or foam) enclosure.

The piped array, as introduced by Daniels,[1] consists of long (603.5 m) lengths of pipe with flow resistors to provide areal sampling of pressure signals using a single acoustic sensor. The pipe is tapered to eliminate acoustic internal reflections. A filter with 100 inlets is found to reduce atmospheric noise approximately 20 dB at wind speeds up to 13 m/s. This method improves the reception of signals but is inherently directional. Burridge[2] devised a simpler device having an untapered pipe and concluded that one can obtain equally good results with a pipe of uniform bore. Consequently an array of several perforated pipes, called a "spider," fanning out radially from a microphone situated in an enclosed housing is needed to achieve omnidirectionality. Alcoverro[3] proposed a filter consisting of several clusters, or rosettes, of low-impedance inlets. Each inlet is connected to a summing manifold by a solid pipe which connects each secondary summing manifold to one primary summing manifold or directly to the sensor.

Hedlin and Raspet[4] used the concept of Lizska and Bernoulli scaling between wind speed and pressure. They used a 50% porous barrier with a 2-m height and a 5.5-m base. A fine mesh wire of 1 mm square was used on a side to reduce the wind flow further. They observed infrasonic noise reduction of 20–25 dB above the reduced corner frequency and 4 dB below the corner frequency (−3 dB). One main advantage of the wind barrier over the rosette spatial filter is that the signal is not integrated after propagation through a system of narrow pipes.

The piped array has the shortcomings of being expensive and requiring typically an area equivalent to an athletic field. It is not of practical use in a three-dimensional array (in which one or more microphones are elevated, for example on a tower). The wind barrier, likewise of a sizeable and noncompact configuration, displays "large reductions only when the scale size of the turbulence is smaller than the height of the barrier."

Several compact windscreens have been proposed, among which a few candidates have been considered for infrasonic service.

The windscreen of Bruel[5] consists of a wire framework covered with a cloth-type material of low air resistance, the microphone being located at the center of the structure. His observation was that if the selected cloth is of an open mesh, the sound waves will penetrate without any appreciable loss in energy, but the wind resistance of the material will be sufficient to dissipate any direct air-flow around the microphone. These windscreens were designed for the audio frequency range and are not effective for infrasound.

Anderson[6] offered a different design. His windscreen comprises an inner cylindrical screen, providing an enclosure for a microphone, and an outer cylindrical screen assembly substantially concentric with, and spaced from, the first screen, whereby a shunt airflow path is provided around the inner screen. Anderson claimed this windscreen to be highly effective and to generate no appreciable wind noises of itself within the frequency band of sound to be detected. The diameter of his windscreen depends upon the desired fre-

[a] Electronic mail: qamar.a.shams@nasa.gov

quency range. For a frequency range between 130 and 330 Hz, he used an inner windscreen of 0.1524 m (6 in.) and an outer windscreen of 0.381 m (15 in.) in diameter. No other spectral information is included.

Ballard[7] developed a windscreen for reducing or eliminating the effect of wind flow on signals received by a transducer in the low sonic and infrasonic range. He used a concentric arrangement of the multiple windscreens, which constitutes a wind noise reduction system. His smallest concentric screen diameters are in the range of 0.381–0.6096 m (15–24 in.) while the largest are about 2.4389–3.6576 m (96–144 in.). Ballard's patent shows unscaled time histories but no spectra.

Morgan and Raspet[8] compared the noise reduction characteristics of various spherical open-cell foam windscreens of different size and porosity. During their first test, they compared spherical screens with streamlined screens designed to reduce wake shedding. They determined that the performance of streamlined screens tested worse than the spherical screens. In their second test, they compared the wind noise reduction characteristics of different size and porosity windscreens, and concluded noise reduction to be relatively insensitive to diameter and pore size. Although data were taken in third-octave bands, only a wideband spectrum in the range 1.6 Hz to 25 kHz is reported.

The lack of adequate spectral information precludes the possibility of evaluating prior compact windscreens, made of cloth or open-cell foam, for infrasonic service.

An effective infrasonic windscreen must fulfill four basic requirements: (1) reduction of noise generated by ambient wind near the microphone, (2) transmission of infrasound propagating to the microphone position, (3) all-weather operation, and (4) avoidance of sounds within the infrasonic passband (<20 Hz) due to vortex shedding (aeolian tones).

The windscreen described here is based on an entirely different principle, namely that infrasound at a sufficiently long wavelength will penetrate any barrier of practical thickness. Thus a windscreen having solid, nonporous walls will block convected pressure fluctuations from the wind, while transmitting acoustic waves within the infrasonic passband. The transmission coefficient depends strongly upon the ratio of the acoustic impedance of the windscreen material to that of air. Several materials having a low impedance ratio (in parentheses), namely closed-cell polyurethane foam[9] (222), Space Shuttle tile material[10] (332), and balsa (323) have proved suitable in that they permit high transmission of infrasound with walls of practical thickness.

The purpose of this study is to test the windscreens experimentally against the requirements listed above; comparison with theory is beyond the scope of this paper.

II. PRINCIPLE OF OPERATION

The principle of the windscreen is based on the great penetrating capability of infrasound through matter. Consider the classical problem of sound transmission through a solid wall with air on either side, as illustrated in Fig. 1. The sound power transmission coefficient from medium 1 to medium 3 is
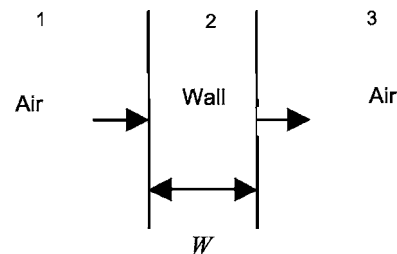
FIG. 1. The three-media problem: transmission of sound in air 1 through a solid wall 2 of thickness W into air 3 on the other side.

$$|T| = \frac{1}{\cos^2 x + \frac{1}{4}(Z_2/Z_1 + Z_1/Z_2)^2 \sin^2 x}, \quad (1)$$

where

$$x = \frac{2\pi f W}{c_2} \quad (2)$$

and f is the frequency, W is the wall thickness, Z_1 is the acoustic impedance of air, Z_2 is the acoustic impedance of the wall material, and c_2 is the speed of sound in the wall.[11] Equations (1) and (2) reveal that a high transmission coefficient is favored by a low frequency, low impedance ratio, and small wall thickness. The transmission pass band is taken here, for design purposes, to span the infrasonic region, namely at frequencies below 20 Hz. The impedance ratio depends upon the selection of suitable materials. The wall thickness is limited by considerations of structural stability. For illustration, the transmission coefficient for a material with a low impedance ratio (balsa, 323) and with a high impedance ratio (steel, 1.13×10^5) is shown in Fig. 2. A cutoff frequency f_c is defined as the location of the half-power point. For a balsa windscreen having a wall thickness of 0.0127 m $\left(\frac{1}{2}\text{ in.}\right)$ the cutoff frequency is 20 Hz, well suited for infrasonic service and structurally stable. For a steel windscreen having the same wall thickness, the cutoff frequency is 0.075 Hz, confining the transmission to the far infrasound.

Table I lists several candidate materials for infrasonic service. Windscreens made of white pine and cedar, requiring a wall thickness no greater than 0.003175 m $\left(\frac{1}{8}\text{ in.}\right)$ to
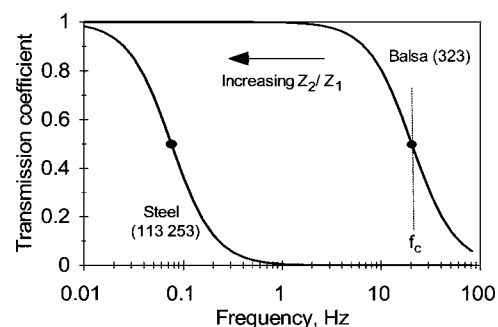
FIG. 2. Transmission coefficient of sound through a solid wall (balsa or steel) of thickness $W=0.0127$ m $\left(\frac{1}{2}\text{ in.}\right)$ versus frequency. The figures in parentheses are the specific acoustic impedance relative to air. The cutoff frequency f_c is the frequency at the half-power point. Balsa reveals excellent transmission at infrasonic frequencies.

Shams *et al.*: Compact infrasonic windscreen

TABLE I. Properties of candidate materials for infrasonic windscreen service.

Material	Density ρ (kg/m^3)	Sound speed c (m/s)	Impedance ratio[a] Z_2/Z_1
White pine	380	5147	4713
Cedar	310	4218	3151
Balsa[b]	102.5	1308	323
Space shuttle tile[b,c]	144	956	332
Polyurethane foam[b,d]	128	719	222

[a]Specific acoustic impedance of air: $Z_1 = 415$ Ns/m^3.
[b]In-house measurements.
[c]See Ref. 10.
[d]See Ref. 9.

FIG. 3. Top: The family of tested windscreens. Bottom: View from the bottom of windscreens having an internal diameter of 0.0762 m (3 in.), internal depth of 0.23 m (9 in.), and wall thicknesses of 0.01905, 0.0127, and 0.00635 m ($\frac{3}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ in.), respectively. In service a windscreen is fitted with a bottom cap to accommodate the microphone.

achieve a cutoff frequency of 20 Hz, were found to develop cracks in the sidewalls and proved impractical for field deployment. The remaining candidate materials—balsa, closed-cell polyurethane foam, and Space Shuttle tile material—were found to achieve the desired cutoff frequency of 20 Hz, with a structurally sound 0.0127-m ($\frac{1}{2}$-in.) wall, and thus proved favorable for field application.

Windscreens were fabricated in the shape of a hollow cylinder, capped permanently at one end and open at the other end for installation on the test fixture described below (see Fig. 3). The internal diameter ranged from 0.0254 to 0.1016 m (1 to 4 in.), the wall thickness from 0.003175 to 0.01905 m ($\frac{1}{8}$ to $\frac{3}{4}$ in.), and the interior height was fixed at 0.23 m (9 in.). For subsequent field service, the open end was fitted with a cap containing a hole for installation of a microphone.

III. INFRASONIC WINDSCREEN TEST SETUP

An existing standard on windscreen testing is explicitly intended for the audio range of frequencies (0.125–10 kHz).[12] Consequently, to test the infrasonic windscreen performance, a small wind tunnel was constructed consisting of a fan, test section, test fixture, anemometer, microphone, subwoofer, and data acquisition system, as illustrated in Fig. 4.

The fan (Orion model OA254AP-11-1TB) delivered 0.27 m^3/s (574 ft^3/min) airflow with a fan speed of 3500 rpm and rated noise level of 55 dB. The fan forced the air to flow down a converging duct, 1.28 m (48 in.) long $\times 0.0323$ m^2 (6×6 in.^{2}) at the exit. The air velocity was 11.6 m/s at the exit and fell to 9.3 m/s near the windscreen, as the air expanded into the microphone test section. The test section dimensions were 0.9144 m long $\times 0.533$ m $\times 0.533$ m (36 in. long $\times 21$ in. $\times 21$ in.). The microphone (and preamplifier) protruded through a pinewood fixture, which contained a boss to fit the open end of the windscreen. The windscreens that were tested were fitted snugly to the fixture assembly and were firmly held in place with two retaining cables, thus insuring that the interior microphone was tightly sealed. The fixture was mounted on a metal plate that was bolted to the test section. The windscreen test microphone was an infrasonic pressure-field microphone (B&K type 4193) with a frequency response of 0.07–20 Hz (-2 dB), and a rain cover (B&K type UA 0393) was installed

on the microphone to simulate field test conditions. The data acquisition system consisted of a PC-based multi-analyzer system (B&K Pulse type 3560 with a type 3110 input module). Its frequency response was determined by a filter setting, in this case 0.7–100 Hz (wind noise) or 0.7–200 Hz (transmission).

To measure the wind noise generated by the windscreen, the fan was turned on to generate airflows known to cause wind-induced windscreen noise, which was detected by the microphone inside the windscreen. The airflow rate was held constant at 9.3 m/s (21 mph), as measured with a hot wire
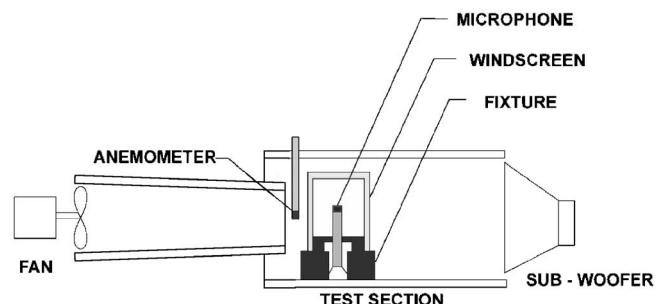
FIG. 4. Wind tunnel used for wind noise reduction (insertion loss) and sound transmission tests. Not to scale.
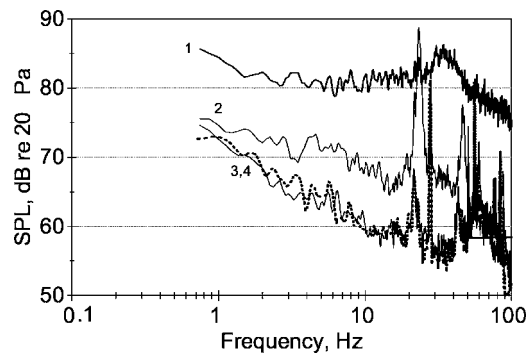
FIG. 5. Sound pressure level of wind noise measured by the microphone as a function of windscreen material. The windscreens had an internal diameter of 0.0762 m (3 in.) and wall thickness of 0.0127 m ($\frac{1}{2}$ in.). The wind speed was 9.3 m/s. Legend: (1) no windscreen, (2) Space Shuttle tile material, (3) balsa (dashed), and (4) polyurethane foam.

FIG. 6. Sound pressure level of wind noise measured by the microphone as a function of internal diameter (ID). The windscreen material was polyurethane foam and the wall thickness was 0.0127 m (0.5 in.). Legend: (1) no windscreen, (2) ID=0.0254 m (1 in., dashed), (3) 0.0508 m (2 in.), (4) 0.0762 m (3 in., dashed), and (5) 0.1016 m (4 in.).

anemometer probe near the windscreen test fixture. The anemometer was connected to a PC-based Labview program, which measured, calculated, and displayed the airflow value. The multi-analyzer system recorded the average of 20 sets of data from 0.7 to 100 Hz taken with a frequency resolution of 0.125 Hz. The noise reduction was determined by comparing the microphone signals with and without the windscreen.

The subwoofer (Janis model M1) used to perform the infrasound transmission test operated in the range of 10–200 Hz. To determine the windscreen transmission (still air) of low frequency sound, a function generator (HP Model 3314A) was used to generate a logarithmic frequency sweep. The signal was amplified with a power amplifier (B&K Type 2713) and then applied to the subwoofer. The frequencies tested were 10 to 200 Hz over two sweep intervals of 120 s each, while the microphone response was recorded by the multi-analyzer system. Here the frequency resolution was 0.5 Hz. The transmission coefficient was determined by comparing the microphone signals with and without the windscreen, as was the case with the wind noise reduction test.

IV. TEST RESULTS

A. Wind-generated noise reduction

Tests were conducted to evaluate the reduction of wind-generated noise in the infrasonic region and to compare windscreen materials, internal diameters, and wall thicknesses.

Figure 5 shows the sound pressure level for windscreens made of polyurethane foam, balsa, and Space Shuttle tile material, all having an internal diameter of 0.0762 m (3 in.) and wall thickness of 0.0127 m ($\frac{1}{2}$ in.). The upper plot (1) is the sound pressure level without the windscreen. The foam and balsa windscreens (3, 4) reveal a comparable noise reduction of about 20 dB at 10 Hz and 10 dB at 0.7 Hz. The Space Shuttle tile material (2) reveals a somewhat lesser wind-noise reduction. Several peaks due to aeolian tones can be seen, but all lie outside the infrasonic region. These tones are expected to appear at a Strouhal frequency of approximately 0.2×wind speed/outer diameter.
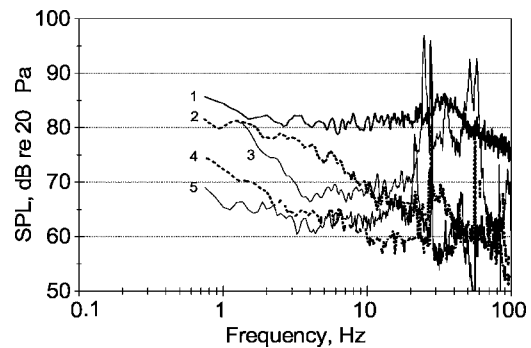
Figure 6 shows the dependence of sound pressure level upon the internal diameter of foam windscreens with internal diameters of 0.0254, 0.0508, 0.0762, and 0.1016 m (1, 2, 3, and 4 in.), respectively, and wall thickness of 0.0127 m ($\frac{1}{2}$ in.). The data reveal a general trend of increasing wind-noise reduction with internal diameter, but in the interval 10–20 Hz, the 0.0762-m windscreen shows the greatest wind-noise reduction.

Figure 7 shows the dependence upon wall thickness. The polyurethane windscreen internal diameter remained at 0.0762 m (3 in.), while the wall thickness had values of 0.00635, 0.0127, and 0.01905 m ($\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$ in.), respectively. The result is anomalous in that the wind-noise reduction appears to reach a maximum at a wall thickness of 0.0127 m.

The large tones are attributed to aeolian tones resulting from vortices shed from the cylinders. These are seen to lie beyond the infrasonic region.

B. Transmission of infrasound through the windscreen walls

The tests were conducted with the subwoofer to compare the same windscreens as used in the wind noise reduction tests. The transmission coefficient is defined as the
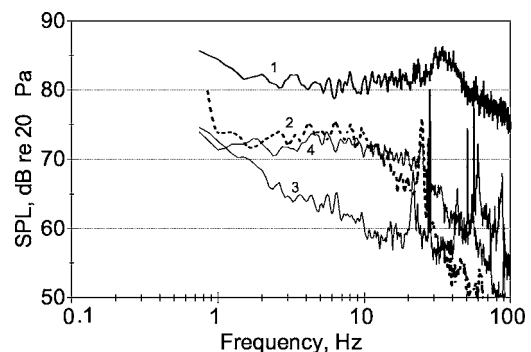
FIG. 7. Sound pressure level of wind noise measured by the microphone as a function of wall thickness W. The windscreen material was polyurethane foam and the internal diameter was 0.0762 m (3 in.). Legend: (1) no windscreen, (2) W=0.01905 m ($\frac{3}{4}$ in., dashed), (3) 0.0127 m ($\frac{1}{2}$ in.), and (4) 0.00635 m ($\frac{1}{4}$ in.).
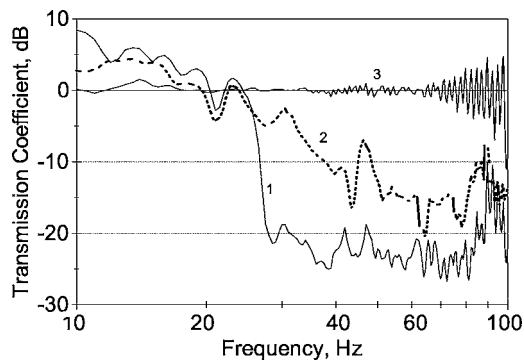
FIG. 8. Transmission coefficient as a function of windscreen material. The windscreen had an internal diameter of 0.0762 m (3 in.) and wall thickness of 0.0127 m ($\frac{1}{2}$ in.). Legend: (1) polyurethane foam, (2) balsa (dashed), and (3) Space Shuttle tile material.

FIG. 10. Transmission coefficient as a function of wall thickness W. The windscreen material was polyurethane foam and the internal diameter was 0.0762 m (3 in.). Legend: (1) $W=0.01905$ m ($\frac{3}{4}$ in.), (2) 0.0127 m ($\frac{1}{2}$ in.), and (3) 0.00635 m ($\frac{1}{4}$ in., dashed).

sound power ratio at the microphone with and without the windscreen. Data are shown for the frequency range 10–100 Hz.

Figure 8 shows the transmission coefficient for the windscreens made of polyurethane foam, balsa, and Space Shuttle tile material, all having an internal diameter of 0.0762 m (3 in.) and wall thickness of 0.0127 m ($\frac{1}{2}$ in.). The foam and balsa windscreens reveal a comparable transmission coefficient in the infrasonic region, below 20 Hz, but the foam has a higher attenuation in the audio region. Below 20 Hz both windscreens show an unexplained gain around 6 dB. The transmission coefficient of the Space Shuttle tile material remains at unity (0 dB) over the entire range of test frequencies.[13]

Figure 9 shows the dependence of the transmission coefficient upon internal diameter. Except for the windscreen with the 0.0254-m ID (1 in.), the windscreens all show the unexplained gain at frequencies below 20 Hz. The foam and balsa windscreens show comparable infrasonic transmission coefficients, the foam being slightly better.

Figure 10 shows the dependence upon wall thickness. A thickness of 0.0127 m ($\frac{1}{2}$ in.) appears to be optimal.

C. Water retention

Water retention is a prime specification for the suitability of a material for outdoor service. Herewith, the polyurethane

FIG. 9. Transmission coefficient as a function of internal diameter (ID). The windscreen material was polyurethane foam and the wall thickness was 0.0127 m (0.5 in.). Legend: (1) ID=0.0254 m (1 in.), (2) 0.0508 m (2 in.), (3) 0.0762 m (3 in., dashed), and (4) 0.1016 m (4 in.).

windscreen, having the 0.0762-m diameter×0.0127-m wall thickness, was held in water for 10 min, removed, had its excess water shaken off, and its weight monitored as the water evaporated. The weight measurements revealed an exponentially decaying water content with time. From an extrapolation back to zero time, it was determined that the windscreen retained 2.54% water by weight—a specification indicating excellent suitability for outdoor service.

V. CONCLUSIONS

A compact and structurally sound windscreen has been developed for microphones deployed in the field for the detection of outdoor infrasound. The windscreen principle is based on the high penetrating capability of infrasound through barriers, which is dependent upon the acoustic impedance ratio (wall-to-air) and wall thickness. Windscreen designs were found to satisfy four requirements for infrasonic service (20 Hz and below): high reduction of wind-generated noise (insertion loss), high transmission of infrasound below the cutoff frequency, absence of aeolian tones within the infrasonic passband, and low water retention as needed for all-weather service.

Most woods, like pine and cedar studied here, have an impedance ratio near 3000 or higher, in which case transmissions below 20 Hz require walls so thin as to cause structural stability problems. Three materials, namely balsa, Space Shuttle tile material, and closed-cell polyurethane foam, all revealing an impedance ratio in the 200–300 range, were found experimentally to meet the above acoustic requirements with windscreen walls of practical, structurally sound thicknesses. A windscreen made of closed-cell polyurethane foam, having an internal diameter, internal height, and wall thickness of $0.0762 \times 0.2286 \times 0.0127$ m $\left(3 \times 9 \times \frac{1}{2} \text{ in.}\right)$ respectively, was found experimentally to achieve the best specifications, namely,

wind-noise reduction:	10 dB at 0.7 Hz and >20 dB at 20 Hz,
infrasonic transmissions:	≥0 dB below 20 Hz,
aeolian tone frequencies:	>20 Hz,
water absorption:	2.54%.

Balsa and Space Shuttle tile material revealed comparable acoustical performance for the same geometry, but the

use of balsa would require a protective film for outdoor service and that of Space Shuttle tile material would be constrained by its high cost.

The selection of specific instruments for testing does not imply endorsement by the National Aeronautics and Space Administration.

[1]F. B. Daniels, "Noise reducing line microphone for frequencies below 1 Hz," J. Acoust. Soc. Am. **31**, 529–531 (1959).

[2]R. Burridge, "The Acoustics of Pipe Arrays," Geophys. J. R. Astron. Soc. **26**, 53–69 (1971).

[3]B. Alcoverro, "Proposition d'un systeme de filtrage acoustique pour une station infrason IMS," CEA-DASE Scientific Report, No. 241 (1998).

[4]M. A. H. Hedlin and R. Raspet, "Infrasonic wind-noise reduction by barriers and spatial filters," J. Acoust. Soc. Am. **114**, 2379–1386 (2003).

[5]P. V. Bruel, "Aerodynamically Induced Noise of Microphones and Windscreens," B&K Technical Review No. 2- 1960, Copenhagen, Denmark.

[6]L. J. Anderson, "Windscreen for Microphones," U.S. Patent 2,520,706 (1950).

[7]H. N. Ballard, "Multiple Wind Screen Noise Attenuation System," U.S. Patent 3,550,720 (1970).

[8]S. Morgan and R. Raspet, "Investigation of the Mechanisms of Low-Frequency Wind Noise Generation Outdoors," J. Acoust. Soc. Am. **92**, 1180–1183 (1992).

[9]LAST-A-FOAM FR-3708, manufactured by General Plastics, Manufacturing Company, Tacoma, WA 98409. This formation is a closed cell rigid polyurethane foam, having a density of 8 lb/ft^3 (128 kg/m^3) per manufacturer's specification.

[10]Low-temperature reusable surface insulation (LRSI). See K. M. Joels, *Space Shuttle Operator's Manual, Revised Edition* (Ballantine Books, New York, 1988).

[11]L. E. Kinsler and A. R. Frey, *Fundamentals of Acoustics* (Wiley, New York, 1962), p. 136.

[12]ANSI S1.17–2004/Part 1, "Microphone Windscreens—Part 1: Measurements and Specification of Insertion Loss in Still or Slightly Moving Air" (Acoustical Society of America, Melville, NY, 2004).

[13]This is attributed to the high porosity of the Space Shuttle tile material. See T. W. Clyne, I. O. Golosnoy, J. C. Tan, and A. E. Markaki, "Porous Materials for Thermal Management under Extreme Conditions," Philos. Trans. R. Soc. London, Ser. A (to be published).

Focal depth shifting of a time reversal mirror in a range-independent waveguide

S. C. Walker,[a] Philippe Roux, and W. A. Kuperman

Marine Physical Laboratory of the Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093-0238

A time-reversal mirror refocuses back at the original probe source position. A goal has been to refocus at different positions without model based calculations. A method to refocus at different ranges has already been developed using frequency shifting. Here we present a technique to refocus at different depths than the original probe source in a shallow ocean range-independent waveguide. The requirement is to collect data from various ranges at a single depth, as from a moving broadband radiator, over a distance sufficient to construct the relevant frequency-wave-number (f-k) structure of the waveguide. With this information, it is then possible to focus at arbitrary depth at any of the ranges that the probe source data were taken. Experimental results confirm the theory. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940447]

I. INTRODUCTION

Acoustic time-reversal (TR) focusing has been demonstrated to produce temporally and spatially focused acoustic signals in a static ocean environment. TR focusing consists of recording the pressure field from a distant probe source over a portion of the water column, reversing the signal in time and propagating the time-reversed signal back through the medium resulting in a pressure field that is temporally and spatially focused at the probe source location. The spatiotemporal focusing is a consequence of the time-reversal invariance of the linear lossless wave equation describing acoustic propagation in the ocean environment.[1] In waveguide acoustic environments that are range independent, the acoustic field propagates as dispersive normal modes.[2] In shallow ocean waveguide applications, TR is often implemented using a vertical line array (VLA) of acoustic transducers covering some or all of the water column. The VLA is often referred to as a time-reversal mirror (TRM).[1]

A TRM is limited by the requirement that a probe source initially broadcast from the desired focal spot. A method for shifting the TR focus in range by altering the frequency characteristics of the pressure field has been experimentally demonstrated.[3] As well, a method for shifting the TR focus in both depth and range in an iso-velocity waveguide in the high bandwidth limit has been proposed.[4] The ability to depth shift a TR focus might find applications to MIMO communications.[5,6] For example, the data rate can be increased by sending different information to different depths simultaneously. This paper introduces a method for shifting the TR focus in depth from the initial probe source depth in the finite bandwidth modal propagation regime. The method is an extension of the frequency-wave-number (f-k) mode extraction method introduced in Ref. 7. Although other methods for extracting the depth dependence of the propa-

gating modes exist (such as the control feedback method introduced in Ref. 8, the amient noise eigen-value decomposition methods discussed in Refs. 9 and 10, and the moving source singular-value decomposition method proposed in Ref. 11), we find the f-k analysis of Ref. 7 particularly useful for depth shifting.

Similar to the method of Ref. 12 generalized to broadband signals in the far-field regime, the f-k structure of the waveguide is constructed by Fourier transforming, in both range and time, the acoustic pressure fields measured over a TRM from sources at many ranges all at the same depth, z'. Due to the dispersion of the waveguide, it is possible to isolate the modes according to their discrete wave numbers in the f-k domain and extract them individually. In addition, their wave-number components may be modulated independently. Noting that the relative intensities of the modal components of the f-k structure are a function of the source depth and the depth dependence of the modes, the extracted modes can be used to remodulate the isolated modal components individually resulting in a depth shifting of the f-k structure. Transforming the depth shifted f-k structure back to the time-range domain yields the desired depth shifted time-domain TRM pressure fields for each initial source range. The depth shifted pressure fields can be time-reversed to produce TR foci at any of the initial ranges. The process is presented schematically in Fig. 1. The dimensions of Fig. 1 describe our experiment at ultrasonic frequencies ($\sim$600 kHz, $D/\lambda \approx 10$, $\Delta r/\lambda \approx 250$), where λ is the carrier wavelength of the acoustic field, D is the water depth, and Δr is the range over which data is recorded. The process is equally applicable to spatial scales characteristic of the shallow ocean in the regime $\Delta r/\lambda \approx (D/\lambda)^2$.[13] Choosing a practical limit on the range aperture, $\Delta r \lesssim 5$ km, limits the applicability of this technique to regimes involving 5–40 propagating modes, $5 \lesssim D/\lambda \lesssim 40$, where the lower limit results from the minimal TR focus resolution for which depth shifting is useful.

---

[a]Electronic mail: shane@physics.ucsd.edu
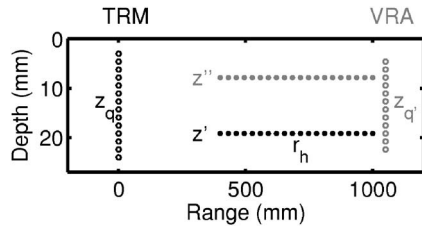
FIG. 1. Experimental setup. A time-reversal mirror (TRM) composed of Q transducers at depths z_q (black open circles) located at range $r=0$ individually recorded pressure fields from acoustic point sources (black closed circles) at ranges $r=r_h$ and depth z'. The resulting depth shifted TR foci at depth z'' at selected ranges (gray closed circles) were measured with a vertical receive array (VRA–gray open circles) with elements at depths $z_{q'}$.

Though this technique may be adapted to weakly range-dependent (adiabatic) bathymetries, we have chosen to limit our analysis to range-independent environments in order to facilitate the description of the underlying physics. Broadening the scope of the discussion to range-dependent environments would overly complicate the analysis.

The theory of depth shifting by mode remodulation is presented in Sec. II. It is developed and presented in conjunction with experimental results in Sec. III. The experimental results serve to illustrate the theory. Complications associated with the depth shifting procedure are dealt with in Sec. IV. A discussion and summary is given in Sec. V.

II. THEORY OF MODE REMODULATION DEPTH SHIFTING

According to normal mode theory, the pressure field at angular frequency component, ω, measured over a vertically oriented TRM (with elements at discrete depths z_q $\{q=1,2,\ldots,Q\}$) located at $r=0$ due to a point source at (r,z') in a range-independent waveguide can be expanded into a superposition of depth-dependent normal modes,

$$\boldsymbol{P}(z',r,\omega) = \sum_{m=1}^{M} \boldsymbol{\phi}_m(\omega)\phi_m(z',\omega)\psi_m(r,\omega), \qquad (1)$$

where

$$\psi_m(r,\omega) \equiv \frac{i\widetilde{S}(\omega)}{4\rho_0}H_0^{(1)}(k_m(\omega)r). \qquad (2)$$

$\widetilde{S}(\omega)$ is the complex frequency component of the source signal, $k_m(\omega)$ is the mode m propagating wave number, and ρ_0 is the medium density. Throughout the paper boldface symbols represent vectors over the depths of the TRA elements. Thus $\boldsymbol{\phi}_m(\omega) \equiv \{\phi_m(z_q,\omega)\}$ and $\boldsymbol{P}(z',r,\omega) \equiv \{P_q(z',r,\omega)\}$ are Q component vectors representing the projections over the TRM of the mode m depth dependence and the pressure field, respectively. M is the total number of propagating modes supported by the waveguide.

Each normal mode arrives at the TRM with an associated amplitude and phase given by $\phi_m(z',\omega)\psi_m(r,\omega)$. The depth dependence of the mode amplitudes, $\phi_m(z',\omega)$, has been isolated in Eq. (1) in order to emphasize the dependence of the mode amplitudes on the source depth. The pres-

sure fields from point sources at different depths at the same range differ only by the relative amplitudes of the modes and are related through

$$\boldsymbol{P}(z'',r,\omega) = \sum_{m=1}^{M} \boldsymbol{\phi}_m(\omega)\phi_m(z'',\omega)\psi_m(r,\omega)$$

$$= \sum_{m=1}^{M} \boldsymbol{\phi}_m(\omega)\left(\frac{\phi_m(z'',\omega)}{\phi_m(z'',\omega)}\right)\phi_m(z',\omega)\psi_m(r,\omega).$$

$$(3)$$

Noting that a TRM produces an acoustic focus back at the source depth, depth shifting can be achieved by remodulating the relative weighting of the modes according to the term in parentheses. This remodulation strategy requires that the mode amplitudes at depths z' and z'' be known and that the modes be isolated, or uncoupled, so that they may be individually remodulated.

There are two basic strategies for isolating the modes so that they may be independently modulated. The first strategy capitalizes on the orthonormality of the modes over a full spanning array, $(1/\rho_0)\boldsymbol{\phi}_n \cdot \boldsymbol{\phi}_m = \delta_{nm}$. Here, the dot notation (·) indicates an inner product. The modal amplitudes can be isolated by projecting the individual modal depth dependencies over the measured array data,

$$\frac{1}{\rho_0}\boldsymbol{\phi}_n \cdot \boldsymbol{P}(z',r,\omega) = \phi_n(z',\omega)\psi_n(r,\omega). \qquad (4)$$

In this sense, the projection operation acts as a filter. The isolated modal amplitudes can be remodulated and the modal depth-dependence reintroduced. The depth shifted pressure field is synthesized from the depth shifted modal components. It is important to emphasize that projection requires an accurate knowledge of the modes over the entire water column and is only effective in full-spanning applications. Projection will be revisited in Sec. IV A. In the second strategy, mode isolation is achieved by directly localizing the wave-number characteristics of the modes.[7] By performing a wave-number analysis, it is possible to isolate the modes according to their discrete propagating wave numbers. Though this requires the accumulation of pressure fields from many ranges, it has the advantages that it can be applied over a limited aperture TRM and that it results in depth shifting over many ranges (see Fig. 1). As a result, the depth shifting procedure developed throughout the remainder of the paper is based on the latter strategy.

III. LABORATORY DEMONSTRATION OF DEPTH SHIFTING

Depth shifting is now developed in the context of wave-number isolation based on the mode extraction method of Ref. 7. Experimental results are presented and used to illustrate the development of the wave-number isolation depth shifting technique. The development is carried out in three phases: mode extraction, depth shifting, and time-reversal. For the sake of clarity, a discussion of complicating factors that arise during the development is delayed until Sec. IV.

A. Mode extraction from experimental data

The first step of the depth shifting method requires the extraction of the acoustic modes from measured data alone. Applying the mode extraction method of Ref. 7, the modes are extracted from the *f-k* structure of the waveguide according to their discrete propagating wave numbers. The *f-k* structure results from applying Fourier transforms over both time and range to accumulated pressure field data from many ranges,

$$P(z',k,\omega) = \mathcal{F}_r\{\mathcal{F}_t\{P(z',r,t)\}\}$$

$$= \sum_{m=1}^{M} \boldsymbol{\phi}_m(\omega)\phi_m(z',\omega)\psi_m(k,\omega). \qquad (5)$$

Here $\mathcal{F}_t\{\}$ and $\mathcal{F}_r\{\}$ indicate the Fourier transforms over time and range, respectively. The ranges need not be evenly spaced.[14] The modal wave-number components of the *f-k* structure, $\psi_m(k,\omega)$, are peaked about $k(\omega)=k_m(\omega)$. For example, for the case of uniform range sampling at interval dr over a range aperture Δr centered at range r_c in the far-field limit $(\lim_{k_m r\rightarrow\infty} H_0^{(1)}(k_m r)\sim e^{-ik_m r})$ of a range-independent waveguide [see Eq. (2)], the modal components of the *f-k* structure can be approximated by the classical array diffraction result,

$$\psi_m(k,\omega) \approx \frac{i\widetilde{S}(\omega)e^{-ik_m(\omega)r_c}}{4\rho_0\sqrt{k_m(\omega)r_c}}\left(\frac{\sin\left((k_m(\omega)-k(\omega))\dfrac{\Delta r}{2}\right)}{\sin\left((k_m(\omega)-k(\omega))\dfrac{dr}{2}\right)}\right).$$

$$(6)$$

In deriving Eq. (6) we neglected the effects of cylindrical spreading in applying the range-wave-number transform. The fast Fourier transform in range results in a signal-to-noise ratio (SNR) gain that is proportional to the number of range samples, $\Delta r/dr$ (dr is the nominal sampling interval).

The first phase of the experiment involved the accumulation of pressure field data over range. The experiment was performed in a laboratory waveguide using an ultrasonic transducer (400 kHz bandwidth at a carrier frequency of 600 kHz). The waveguide consisted of approximately 27 mm ($\approx 10.5\ \lambda$) of water over a homogeneous sand bottom. Together, the far-field requirement and the finite length of the waveguide constrained the operational range limits to between 400 and 1000 mm ($\approx 160-400\ \lambda$).

The experimental configuration is illustrated in Fig. 1. A TRM (open black circles) located at $r=0$ individually recorded pressure fields from the acoustic source at many ranges (closed black circles) all at the same depth, z'. The TRM consisted of $Q=31$ evenly spaced hydrophone elements spanning most of the water column. Denoting the TRM elements by the index q, the time-domain pressure field measured at element q due to a source at depth z' and range r' is represented by the notation $P_q(z',r',t)$. Panels (a) and (c) of Fig. 2 show typical measured time-domain pressure fields (the remaining panels of Fig. 2 are discussed later).
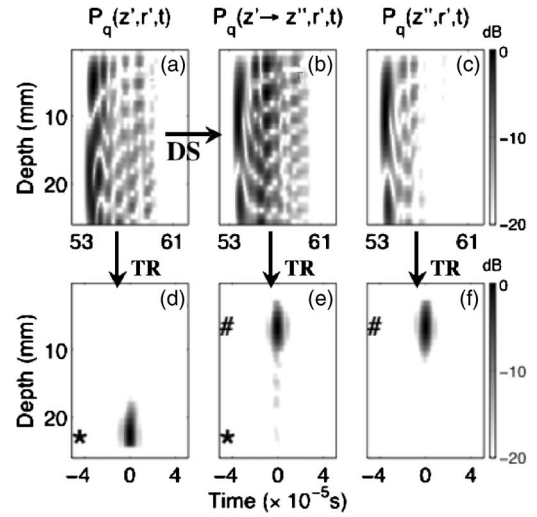
FIG. 2. Experimental depth shifting (DS) result at a single range. (a) Time domain pressure field, $P_q(z',r',t)$, recorded on the TRM due to an acoustic source at (r',z') produced (d) TR focus at (r',z'). (b) Depth shifted time domain TRM pressure field, $P_q(z'\rightarrow z'',r',t)$, and (e) resulting TR focus at (r',z''). (c) TRM pressure field, $P_q(z'',r',t)$, due to acoustic source at (r',z'') and (f) its TR focus at (r',z''). The source depth, z', and desired focal depth, z'', are indicated by the symbols (∗) and (♯) respectively. $z'\approx 25$ mm, $z''\approx 7$ mm.

Figure 3(a) shows the depth averaged modulus of the *f-k* structure, $1/Q\Sigma_{q=1}^{Q}|\hat{P}_q(z',k,\omega)|$, which results from Fourier transforming the measured data over range. The hat notation, $\hat{\boldsymbol{P}}\equiv\{\hat{P}_q\}$, indicates that a frequency-wave-number rotation transformation has been applied to the time-range domain pressure field prior to applying the Fourier transformations. This initial rotation transformation, introduced in Ref. 15, facilitates mode extraction. As discussed in Ref. 7, the rotation transforms the location of the modal wave numbers in the *f-k* domain, $k_m(\omega)\rightarrow k_1(\omega)-k_m(\omega)$. In addition to exaggerating the intermodal intervals, this choice of rotation mitigates aliasing of the high order modes. As a result, the range sampling interval can be increased, $dr=2\pi/|k_w-k_b|$, where k_w and k_b are the wave numbers at the carrier frequency in the water and bottom, respectively. The inverse transformation is well defined and easily applied to the depth shifted pressure field discussed in Sec. III B.

Most of the modes (appearing as dark curved regions)
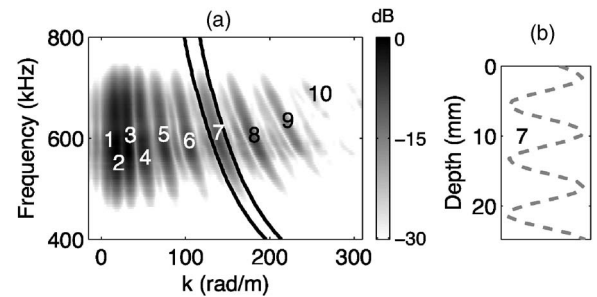
FIG. 3. (a) Depth averaged modulus of the rotated *f-k* structure of the experimental waveguide. Propagating modal wave numbers appear as dark gray curved regions. The modal wave numbers, which are enumerated for clarity, appear at their rotated values, $k_m(\omega)\rightarrow k_1(\omega)-k_m(\omega)$. Solid black lines indicate the mode 7 mask region. (b) Mode 7 depth dependence extracted from masked region.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Walker, Roux, and Kuperman: Depth shifting   1343

are localized and distinct from one another in wave number. Following the mode extraction method detailed in Ref. 7, the well-localized modes were masked and extracted individually. A mask defines a region in the *f-k* domain containing only one mode that is projected over all TRM elements. The only requirement on the masking procedure is that each mask contain a significant contribution from only one mode. Otherwise the mask shape is arbitrary. The mode 7 mask is plotted in Fig. 3(a) as an example. Figure 3(b) shows the resulting extracted mode 7. Modes 3–10 were successfully extracted in this manner. Because the extraction process involves frequency averaging, the extracted modes are the modes averaged over their respective masks. To denote this averaging, the set of extracted modes are written with an overline notation, $\bar{\phi}_m(z)$. Due to the frequency averaging associated with the mode extraction process, it is necessary to limit the frequencies spanned by each mask to bandwidths where the modes are weakly frequency dependent. The major advantage of this method is that the modes can be extracted over a partial spanning array (in this case the TRM) using data alone.

Modes 1 and 2 were not well extracted. Due to the limited range aperture of the experiment, the low order modes were not well localized and could not be separated and individually masked. In addition to limiting mode extraction, poor localization resulting from insufficient range aperture also complicates depth shifting. Poorly localized modes cannot be modulated independently. Due to SNR considerations, the range aperture limitation is likely to be encountered in at sea applications of the depth shifting (DS) method. The issues of determining poorly extracted modes and remodulating poorly localized modes are dealt with in Secs. IV A and IV B.

B. Depth shifting experimental data

In addition to facilitating mode extraction, the mode masking technique makes it possible to independently remodulate the modes with only a knowledge of the source depth. The source depth need not be known *a priori*. It can be determined by applying mode matching techniques to the set of extracted modes.[16–18] Once the source depth, z', is determined, the set of extracted modes are used to calculate the modal remodulation coefficients, $\bar{\phi}_m(z'')/\bar{\phi}_m(z')$[the parentheses term of Eq. (3)]. Remodulation is accomplished simply by multiplying each mask by its corresponding remodulation coefficient. This is valid since the method has been restricted to weakly frequency dependent regimes. Intermodal *f-k* structure not included in any masks is irrelevant to far-field modal propagation and is discarded (set to zero). As well, it is possible to discard high order modes for which the TRM depth sampling is inadequate. Figure 4 shows the experimental masked *f-k* structure [compare to Fig. 3(a)]. Modes 3–10 were remodulated by directly multiplying each mask by its corresponding remodulation coefficient. The independent remodulation of modes 1 and 2 was achieved by the method described in Sec. IV A.

Since the array data now represent an ensemble of acoustic fields excited at depth z'' rather than z', applying the
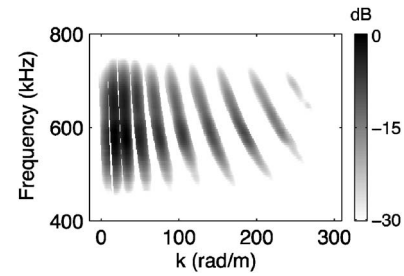
FIG. 4. Depth averaged modulus of the masked rotated *f-k* structure of the experimental waveguide. Dark regions indicate the relevant masked modal regions. White regions indicate intermodal modal structure exterior to masks that has been set to zero.

inverse *f-k* rotation transformation and inverse Fourier transforming the depth shifted *f-k* structure back to the time and range domains yields depth shifted time domain pressure fields for all original data ranges,

$$\boldsymbol{P}(z' \rightarrow z'', r, t) = \widetilde{\mathcal{F}}_r\{\widetilde{\mathcal{F}}_t\{\boldsymbol{P}(z' \rightarrow z'', k, \omega)\}\}. \quad (7)$$

$\widetilde{\mathcal{F}}_t\{\}$ and $\widetilde{\mathcal{F}}_r\{\}$ indicate the respective inverse Fourier transformations. Panel (b) of Fig. 2 shows an experimental depth shifted time domain pressure field from a selected range. The depth shifted pressure field resembles the pressure field measured from a source at the desired depth, z'', at the selected range shown in panel (c), particularly with respect to the early arrivals (low order modes). Close examination of panels (b) and (c) reveals that there is more energy in the high order modes relative to the low order modes in the depth shifted pressure field than in the measured pressure field. This is likely due to errors associated with the remodulation procedure. As a result, the depth shifted TR pressure more intense sidelobes in depth than the expected TR result [panel (c)].

C. TR focusing experimental data

Time-reversing the depth shifted pressure fields produces TR foci at any of the initial source ranges. This result was experimentally verified. Depth shifted TR foci were experimentally observed at all initial source ranges (represented schematically in Fig. 1 by the closed gray circles). The foci were measured at selected ranges with a limited aperture vertical receive array (VRA). Figure 2(e) shows the experimental TR focus produced by the depth shifted pressure field shown in panel (b). The limited aperture of the VRA resulted in observed clipping at the shallowest and deepest depths. The data source depth and desired focal depth are indicated by (∗) and (♯), respectively. As a control, this result is compared to the optimal TR foci [panels (d) and (f)] produced by the measured pressure fields from the source and desired depths shown in panels (a) and (c), respectively. Figure 5 shows experimental depth shifted TR foci at various depths at a given selected range. Figure 6 shows experimental depth shifted TR foci at various ranges at a given desired focal depth.
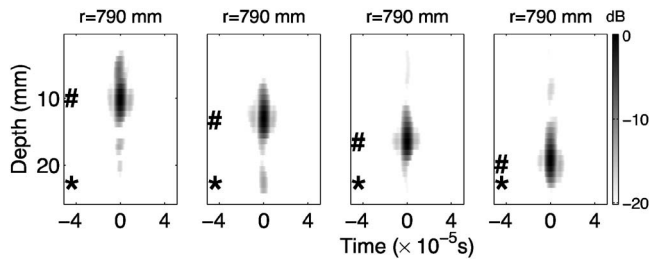
FIG. 5. Experimental depth shifting results at multiple depths. TR foci produced at several depths at the same range. The source depth, z', and desired focal depth, z'', are indicated by the symbols ($*$) and ($\sharp$) respectively.

IV. OVERCOMING SOME COMPLICATIONS

Thus far the DS procedure has been developed in the context of ideal conditions; namely, that all modes are well localized and maskable in the f-k domain. In practice, this is often not the case. In the following we discuss methods for dealing with poorly localized modes.

A. Remodulating coupled modes

As mentioned in Sec. III A, wave-number resolution may limit the localization of the low order modes which complicates the masking procedure. The range aperture, Δr, sampled by the data determines the wave-number resolution,

$$dk = \frac{2\pi}{\Delta r}. \tag{8}$$

Adjacent modes (in the f-k domain) for which the intermodal spacing is of the order of the wave-number sampling interval, dk, cannot be masked, extracted, or independently remodulated. Because intermodal spacing decreases with mode order, the lowest order modes are the most prone to the effects of wave-number resolution limitations.

The simplest strategy for dealing with modes that cannot be separated is to eliminate them from the depth shifted f-k structure entirely. Neglecting the low order modal contributions leads to sidelobes in the time-reversed field. The experimental TR field resulting from a depth shifted pressure field from which modes 1 and 2 were eliminated is shown in Fig. 7. Comparing this to a case where the depth shifted pressure field included properly remodulated contributions from modes 1 and 2 [Fig. 2(e)], we see that neglecting the low order modes leads to poor depth shifting results.

In the special case where the depth dependence of the poorly localized modes is known over most of the water
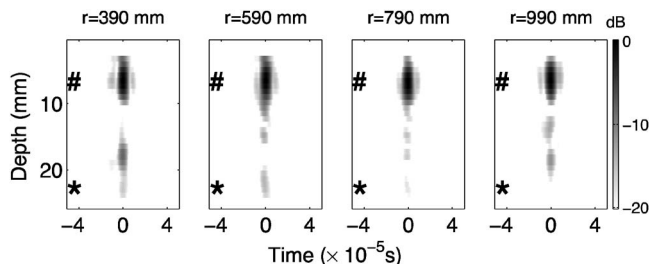
FIG. 6. Experimental results at a multiple ranges. TR foci produced at several ranges. The source depth, z', and desired focal depth, z'', are indicated by the symbols ($*$) and ($\sharp$) respectively.
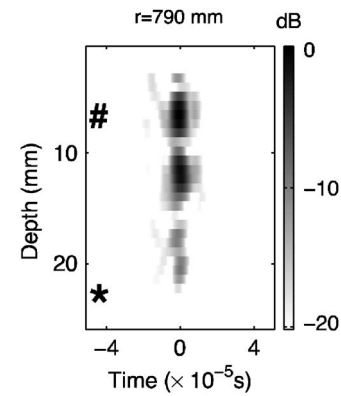
FIG. 7. Experimental depth shifting result. Time-reversal of depth shifted pressure field from which poorly localized, unmaskable modes 1 and 2 were removed. The source depth, z', and desired focal depth, z'', are indicated by the symbols ($*$) and ($\sharp$) respectively.

column (by most here we mean the portion of the space over which the modes can be considered orthogonal to one another), modal wave-number isolation and remodulation can be achieved by the projection procedure mentioned in Sec. II. However, recalling that the mode extraction procedure of Ref. 7 is also limited by the wave-number resolution, the depth dependence of poorly localized modes cannot be determined by this technique. Later on we outline an optimization procedure for determining the depth dependence of poorly localized modes from a knowledge of the depth dependence of the known, well-localized modes. Here we continue with the projection remodulation procedure under the assumption that the depth dependence of the poorly localized modes is known over most of the water column.

Applying the projection of a given mode to f-k domain rotated pressure field, $\boldsymbol{P}(z',k,\omega)$, yields

$$\Psi_\alpha(k,\omega) \equiv \frac{1}{\rho_0}\boldsymbol{\Phi}_\alpha \cdot \boldsymbol{P}(z',k,\omega) \approx \phi_\alpha(z',\omega)\psi_\alpha(k,\omega), \tag{9}$$

where $\boldsymbol{\Phi}_\alpha$ represents an optimized poorly localized mode projected over the TRM. The wave-number components are then remodulated independently, $(\Phi_\alpha(z'')/\Phi_\alpha(z'))\Psi_\alpha(k,\omega)$. The depth dependence of the corresponding modes, $\boldsymbol{\Phi}_\alpha$, is reintroduced and the components summed with the masked regions of the depth shifted f-k structure. Figure 8 shows a schematic of the process applied to experimental data. Panel (a) shows a magnification of the unlocalized low order modes (see Fig. 3). Panels (c) and (e) show the modulus of the isolated mode 1 and 2 wave-number components, $|\Psi_1(k,\omega)|$ and $|\Psi_2(k,\omega)|$, respectively, resulting from the projection procedure. The mode 1 and 2 depth dependencies used to make the projections appear in panels (b) and (d). The array diffraction behavior predicted by Eq. (6) is apparent for each mode. The projection would not have succeeded in separating modes 1 and 2 had the TRM spanned only a small portion of the water column.

Remembering that mode extraction only works in the case of well-localized, maskable modes, we point out that modes 1 and 2 shown in panels (b) and (d) did not result from the original extraction technique. Rather they resulted from an optimization algorithm.
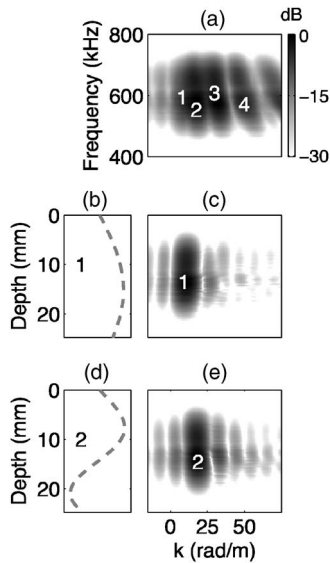
FIG. 8. Rotated frequency-wave-number (f-k) structure resulting from projection of optimized modes. (a) Magnification of Fig. 3 showing low order modes. (b) Optimized mode 1. (c) Projected mode 1 f-k structure. (d) Optimized mode 2. (e) Projected mode 2 f-k structure.

We developed a genetic algorithm,[19] based on the orthogonality conditions relating the modes, to construct a small subset of the lowest order modes from a larger subset of known higher order modes. Based on the appearance of the f-k structure (Fig. 3) we assumed that modes 4–10 were well extracted, modes 1 and 2 were poorly extracted, and mode 3 was questionable. The optimization cost functions for modes 1–3 were defined as

$$F_\alpha \equiv |\mathbf{\Phi}_\alpha \cdot \mathbf{\Phi}_\alpha - c_{\alpha\alpha}| + \sum_{m=4}^{10} |\mathbf{\Phi}_\alpha \cdot \bar{\phi}_m - c_{\alpha m}|. \qquad (10)$$

The $\mathbf{\Phi}_\alpha$ ($\alpha = 1, 2, 3$) represent the optimized modes. $c_{\alpha m}$ and $c_{\alpha\alpha}$ are scalars introduced to account for the fact that the optimization was not performed over the entire space. Because the TRM spanned most of the water column and did not sample the bottom, the set of known modes are not expected to be strictly orthogonal to one another, nor are they expected to be strictly orthogonal to the optimized modes. These scalars were calculated using a crude environmental model based on minimal knowledge of the environment, namely an estimate of the bottom sound speed, bottom density, and the sound speed profile. The resulting optimized modes (solid black curves) are plotted along with the modes derived from the masking extraction process (gray dashed curves) in Fig. 9. That the optimized mode 3 so closely resembles the extracted result tells us that the mode 3 mask was valid. Therefore, only modes 1 and 2 were remodulated by projection.

B. Complications due to weakly excited modes

Depth shifting achieves optimal results when the source excites all propagating modes. Modes whose excitations are too low to appear distinctively in the f-k structure cannot be masked nor extracted. Hence, these modes cannot be included in the depth shifted pressure fields which in turn ad-
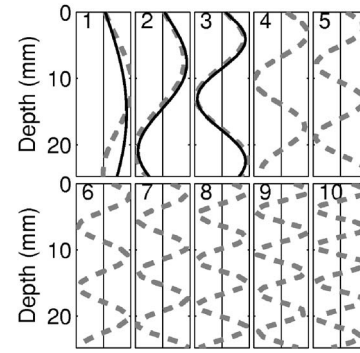
FIG. 9. Experimental mode extraction results. Modes resulting from mode masking extraction technique (gray dashed lines) are plotted along with optimized modes (dark solid lines). Modes 4–10 were used to optimize for modes 1–3.

versely effects the resulting TR focus. Therefore, the efficacy of the DS algorithm depends on the source depth. Optimal TR focusing results when the source is not near any of the modal nodes. In a Pekeris waveguide, for example, the bottom is an ideal depth for the source.

As well, care must be taken in the case of maskable modes whose excitations are small but measurable. Because the source depth excitation appears in the denominator of the mode remodulation coefficient, $\bar{\phi}_m(z'')/\bar{\phi}_m(z')$, small errors in either the denominator or numerator may result in large errors in remodulation. Remodulation errors resulting in overexcitation are avoided by defining a threshold, β_{max}, for the remodulation coefficients. All coefficients above the threshold are set to zero removing the given mode from the depth shifted f-k structure. Underexcitation is preferable to overexcitation.

V. SUMMARY AND DISCUSSION

Depth shifting of a TR focus over a large portion of a range independent waveguide using broadband sources from many ranges has been experimentally demonstrated at ultrasonic frequencies in a laboratory setting that scales to typical shallow water scenarios. Data from many ranges were used to perform a frequency-wave-number (f-k) analysis of the waveguide so that the modes could be isolated according to their characteristic wave numbers. The isolated modes were extracted and remodulated independently resulting in a depth shifting of the waveguide f-k structure. The depth shifted f-k structure was used to produce depth shifted TR foci over the extent of the original range aperture. In cases where the range aperture is large enough to resolve the characteristic wave numbers of all the modes, it is possible to achieve optimal depth shifted TR focusing with a partial spanning TRM. In cases, such as our experiment, where the range aperture is not adequate to resolve all the modes, optimal depth shifted TR focusing can only be achieved with a TRM that spans most of the water column. We achieved good experimental results using a 90%-spanning TRM.

[1]W. A. Kuperman, W. S. Hodgkiss, and H. C. Song, "Phase-conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[2]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computa-*

tional Ocean Acoustics (AIP, New York, 1994).

[3]H. C. Song, W. A. Kuperman, and W. S. Hodgkiss, "A time-reversal mirror with variable range focusing," J. Acoust. Soc. Am. **103**, 3224–3240 (1998).

[4]S. Conti, P. Roux, and M. Fink, "Depth and range shifting of a focal spot in an acoustic waveguide," Appl. Phys. Lett. **80**, 3647–3649 (2002).

[5]D. Kilfoyle, J. Preisig, and A. B. Baggeroer, "Spatial modulation over partially coherent multiple-input/multiple-output channels," IEEE Trans. Signal Process. **51**, 794–804 (2003).

[6]H. C. Song, P. Roux, W. S. Hodgkiss, W. A. Kuperman, T. Akal, and M. Stevenson, "Multiple-input/multiple-output coherent time reversal communications in a shallow water acoustic channel," IEEE J. Ocean. Eng. (in press).

[7]S. C. Walker, P. Roux, and W. A. Kuperman, "Data-based mode extraction with a partial water column spanning array," J. Acoust. Soc. Am. **118**, 1518–1525 (2005).

[8]J. R. Buck, J. C. Preisig, M. Johnson, and J. Catipovic, "Single-mode excitation in the shallow-water acoustic channel using feedback control," IEEE J. Ocean. Eng. **22**, 281–291 (1997).

[9]S. N. Wolf, D. K. Cooper, and B. J. Orchard, "Environmentally adaptive signal processing in shallow water," Oceans '93, Engineering in Harmony with Ocean Proceedings, Vol. **1**, pp. 99–104.

[10]P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, "Matched field processing with data-derived modes," J. Acoust. Soc. Am. **109**, 1355–1366

(2001).

[11]T. B. Neilsen and E. K. Westwood, "Extraction of acoustic normal mode depth functions using vertical line data," J. Acoust. Soc. Am. **111**, 748–756 (2002).

[12]G. V. Frisk and J. F. Lynch, "Shallow water waveguide characterization using the Hankel transform," J. Acoust. Soc. Am. **76**, 205–216 (1984).

[13]G. A. Grachev, "Theory of acoustic field invariants in layered waveguides," Acoust. Phys. **39**, 748–756 (1993).

[14]L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast Fourier transform," SIAM Rev. **46**, 443–454 (2004).

[15]B. Nicolas, J. Mars, and J.-L. Lacoume, "Geoacoustical parameters estimation with impulsive boat-noise sources," IEEE J. Ocean. Eng. **28**, 494–501 (2003).

[16]E. C. Shang, "Source depth estimation in waveguides," J. Acoust. Soc. Am. **77**, 1413–1418 (1985).

[17]T. C. Yang, "A method of range and depth estimation by modal decomposition," J. Acoust. Soc. Am. **82**, 1736–1745 (1987).

[18]G. R. Wilson, R. A. Koch, and P. J. Vidmar, "Matched mode localization," J. Acoust. Soc. Am. **84**, 310–320 (1988).

[19]P. Gerstoft, SAGA Users Guide 4.0, an Inversion Software Package, An updated version of SAGA Users Guide 2.0, an Inversion Software Package, SACLANT Undersea Research Centre, SM-333, La Spezia, Italy, 1997.

A scattering-chamber approach for solving finite rough surface scattering problems

John A. Fawcett

DRDC-Atlantic, P.O. Box 1012, Dartmouth, Nova Scotia, B2Y 3Z7, Canada

In this paper a new method is derived for the computation of scattering from a finite, rough free surface. The free surface is infinite in extent but only a portion of it is rough. In order to reduce the amount of numerical computation for such a problem, it is desirable to restrict the computations to the interval of roughness, even for remote sources and receivers. This can be easily done in the case that the rough portion of the surface is only directed into the surrounding fluid medium. In this case, the use of the appropriate half-space Green's function will restrict the integral equation to the interval of roughness only. However, for general deformations this Green's function cannot be used. The use of truncated integral equations utilizing the free space Green's function is discussed. An alternate approach is then described. A system of boundary conditions is derived for a finite curve containing the interval of roughness and a surrounding contour in the fluid half-space. The resulting equations are solved using the method of wave-field superposition. The derived method is also easily generalized to the case that the rough surface under consideration is the upper boundary of a waveguide. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1993130]

I. INTRODUCTION

The study of scattering from a rough surface is a problem of much interest in underwater acoustics, electromagnetics, and geophysics. The case of a rough upper pressure release (or free) surface (i.e., the ocean/air interface) is often studied to understand its effects on signal propagation, reverberation, scattering, etc., or, in general, its effect on sonar system performance, both passive and active (for example, Refs. 1–4). The rough upper surface may be spatially extended or, as is considered in this paper, the extent of the roughness of interest can be finite and small relative to, for example, the source and receiver distances. For computational efficiency, it is desirable to have a computational method which requires the solution for the unknown pressure field quantities in only a small region containing the interval of roughness. In addition, the depth of the acoustic transmitter may be small relative to its horizontal range to the area of surface roughness. In this case the acoustic energy incident upon the region of roughness will have a very small angle of incidence (as measured from the horizontal). This small-angle scenario is emphasized in this paper.

The scattering of sound from such rough surfaces has been examined by many authors, often using approximation techniques such as perturbation theory,[5] the Kirchhoff approximation,[6] the small-slope approximation,[7] or more numerical techniques such as parabolic equation methods,[8] finite difference methods,[9] or wedge-based methods.[10] The technique we consider in this paper is the boundary integral equation method (BIEM). This method has the advantage that it includes all wave propagation effects, including backscatter, but the unknown field quantities which must be determined are restricted to the upper pressure release surface. Various authors[11–18] in the areas of acoustics, electromagnet-ics, and geophysics have utilized boundary integral equation methods for rough free surface or range-dependent waveguide problems.

In order to consider a solution on a surface which has an infinite horizontal extent, some approaches invoke periodicity of the medium and the solution, and other authors use a truncated domain utilizing the free-space Green's function or its normal derivative. The solution to these periodically extended or truncated integral equation problems provide an approximate solution to the infinite surface scattering problem. For a fixed incident field and a fixed set of receivers, the length of the truncation interval (or the length of the interval which is periodically extended) is increased until a suitably converged solution is obtained. There are certainly many problems for which an accurate solution can be obtained in such a manner with a reasonable number of computation points. However, there are clearly scenarios where the required length of these computational intervals will be large. For example, the incident field on the roughness area from a distant, shallow point source will consist of the direct energy and a specular reflection from the flat upper surface (and the case of an incident plane wave at a shallow grazing angle is similar). The specular reflection point of a shallow-angle ray on the flat surface may still be distant from the roughness area. Similarly, for distant shallow receivers, some of the energy scattered from the roughness area (as well as the incident field) will reflect off the flat upper surface before reaching the receiver and for shallow-angle rays this reflection point will be distant from the roughness interval. For these cases, by sufficiently increasing the computational domain, the truncated solutions will yield an accurate approximation to the infinite surface solution. However, the required number of computational points may become very large. The advantage of the method described in this paper is that it will use only a moderate computational grid independent of the

source and receiver locations. The incident fields need not be tapered and, in fact, the method is also applicable to the case where the scattering surface is the upper surface of a bounded waveguide. In this case, a distant point source will produce energy incident upon the roughness interval at a wide variety of angles.

In the case that the roughness profile (deformation to the flat surface) points only into the fluid half-space, then by using the half-space Green's function which satisfies the pressure release boundary condition along $z=0$ (the flat portion of the infinite scattering surface is denoted as $z=0$) the resulting boundary integral equation can be restricted to just the interval of roughness.[17] The integral for the extrapolation of the surface solution to yield the pressure at a receiver is also limited to just the interval of surface roughness. For this type of roughness, this form of boundary integral equation accomplishes what is desired: a finite computational domain which allows for the exact numerical computation (assuming a sufficiently fine discretization) of the scattering from a finite interval of roughness surrounded by an infinite, flat scattering surface. The sources and receivers may be arbitrarily far from this finite region. Unfortunately, in the case that the roughness has both outward and inward deformations, then, as will be discussed, this particular boundary integral approach is no longer valid. The aim of this paper is to obtain an analogous approach for this more general case. We wish to find a method which utilizes a fixed, compact computational domain for the solution of the scattering problem and does not depend upon the location of the sources and receivers relative to the interval of roughness.

To derive the method of this paper, a set of boundary conditions for the pressure field and its normal derivative along a section of the scattering surface (including the interval of roughness) and a curve extending into the bounding fluid is considered. The curve along the surface and the curve in the fluid constitute a single curve. The interior of this curve is what we consider to be the "scattering chamber." The portion of the curve in the fluid relates the incident and scattered fields in the surrounding half-space to the field quantities within this chamber. The concept of using a bounding curve or surface to bring incident energy into and scattered energy out of a small region is not new and has been used previously in finite difference[19] methods (in this reference, the term "scattering chamber" is used). In Ref. 20, a finite element solution is coupled to the elastic field in the outside surrounding medium using a multipole expansion expression. In the present paper, all field quantities within the chamber and exterior to the chamber are expressed in terms of the pressure fields from point sources exterior (for the interior fields) and interior to the chamber (for the external fields). The coupling of the solutions across the curve in the water column and the satisfaction of the pressure release boundary condition along the top of the chamber allow for the amplitudes of the point sources to be determined. It is also straightforward to model the incident field and the scattered field from the chamber into the surrounding medium if, in fact, the surrounding medium is a bounded waveguide. Wave superposition methods or multi-center multiple expan-
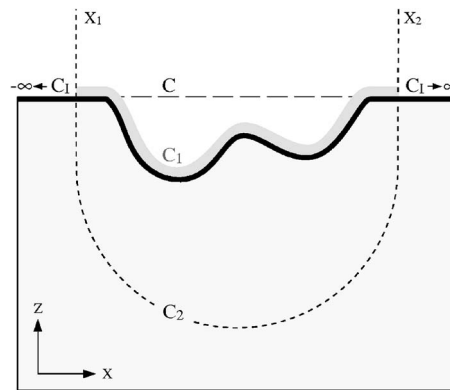
FIG. 1. A schematic of the geometry for the rough surface scattering problem.

sion methods have been used by many authors[21–26] to solve surface, volume, and target scattering problems.

In Sec. II (Theory), the boundary integral equations which can be used for the modeling of scattering from a rough surface are described and some of the approaches and issues when using truncated integral equations are discussed. The problem with using the half-space Green's function for general rough surfaces is described and the scattering chamber approach of this paper is derived. In Sec. III (Numerical Examples) the case of a roughness profile with only an inward deformation is considered. For this case, we can use the exact half-space Green's function approach of Ref. 17 as a benchmark solution. We compare the results of the benchmark solution with those from the scattering chamber approach for a variety of different types of incident fields: incident plane waves, tapered plane waves, and point sources and many different receiver ranges. We also show the computed results using truncated integral approaches (for a fixed computational interval). For the case of a roughness profile with both inward and outward deformations, we can no longer use the exact boundary integral approach to benchmark the scattering chamber approach and, in this case, rely on a truncated integral equation approach (making sure that this solution has converged for the receivers considered) to validate the method. Finally, we present a rough surface/ bounded, penetrable waveguide example using the approach of this paper.

Finally, it should be noted that the boundary integral equation and wave-field superposition methods described in this paper are only computationally feasible for regions of surface roughness of fairly limited size. For many realistic situations, methods such as the parabolic equation[8,19] method would need to be used. The methods described in this paper could be used to provide benchmark solutions for other computational approaches.

II. THEORY

In Fig. 1 a schematic of a rough pressure-release surface is shown. The flat surface is $z=0$. Four contours are shown: (1) C_I is the infinite contour along the upper surface (deformation and flat surface), (2) C_1 is a truncated version of C_I including any flat segments, (3) C is the line $z=0$ corresponding to the horizontal extent of C_1, and (4) C_2 is a con-

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

John A. Fawcett: Rough surface scattering   1349

tour such that $C_1 \cup C_2$ is a closed contour enclosing the roughness patch. The method of this paper will also apply for the case of a rigid surface with the appropriate change of boundary conditions and Green's function. The truncated integral equation approaches described in this paper will not utilize the curve C_2, but such an enclosing curve will be used by the wave-field superposition method. For the truncated integral equation approaches the curve C_1 will include a significant portion along the flat surface. For other approaches, the curve C_1 will consist only of the section of roughness (and, perhaps, a few neighboring points along the flat surface).

A. Boundary integral approaches

Considering the infinite surface described parametrically by $(x(s), z(s))$ (the contour C_I in Fig. 1), the Helmholtz equation relating the pressure field away from the surface, at a field point $\vec{x} = (x, z)$, to the values of the normal derivative of the pressure field, p_n, on the surface can be written

$$p(\vec{x}) = \int_{-\infty}^{\infty} G(\vec{x}; \vec{x}(s')) p_{n'}(\vec{x}(s')) ds' + p^{inc}(\vec{x}), \qquad (1)$$

where

$$G(\vec{x}; \vec{x}(s')) \equiv \frac{i}{4} H_0^1 (k\sqrt{(x - x(s'))^2 + (z - z(s'))^2}), \qquad (2)$$

with H_0^1 denoting the zeroth-order Hankel function of the first kind and $p^{inc}(\vec{x})$ denotes the incident field. The normal derivative points upwards in the positive z direction. The choice of the Hankel function used for the Green's function is dictated by the time convention used—in this paper, the form $\exp(-i\omega t)$ is assumed.

Following, for example, Thorsos,[6] we apply the normal derivative operator to Eq. (1) to obtain

$$p_n(\vec{x}) = \int_{-\infty}^{\infty} G_n(\vec{x}; \vec{x}(s')) p_{n'}(\vec{x}(s')) ds' + p_n^{inc}(\vec{x}). \qquad (3)$$

Taking the limit of Eq. (1) as the field point $\vec{x}$ approaches the surface $(x(s), z(s))$ yields

$$\frac{p_n(\vec{x}(s))}{2} = \int_{-\infty}^{\infty} G_n(\vec{x}(s); \vec{x}(s')) p_{n'}(\vec{x}(s')) ds' + p_n^{inc}(\vec{x}(s)), \qquad (4)$$

where special care must be taken with the limit of the integral as $\vec{x}$ approaches the surface, resulting in the factor of $\frac{1}{2}$ with the left-hand side of Eq. (4). In practice, Eqs. (1) and (4), which we will implement numerically, must be truncated at finite integral limits. These limits can be numerically determined by increasing them until the computed pressure fields at the desired set of receivers have converged (using some appropriate definition of convergence).

In Ref. 17 another approach was taken for computing the scattered fields in a waveguide with a rough free surface (with only inward deformations with respect to the flat upper surface); the waveguide Green's function was used in the integral equations. Because this function satisfies the appropriate boundary conditions on the undeformed portions of the boundaries it is easy to show that the boundary integral equation can be restricted to the area of deformations, with all other integral contributions being zero. For an upper pressure release surface, we write for the half-space Green's function,

$$\widetilde{G}(x, z; x', z') = \frac{i}{4} (H_0^1(kR) - H_0^1(k\widetilde{R})), \qquad (5)$$

where

$$R \equiv \sqrt{(x - x')^2 + (z - z')^2}, \quad \widetilde{R} \equiv \sqrt{(x - x')^2 + (z + z')^2}. \qquad (6)$$

The incident field, $\widetilde{p}_n^{inc}(\vec{x}(s))$, in this case is the wave field which would exist if the upper surface were flat. For the case that the horizontal range of the surface roughness is $[x_1, x_2]$ we can write Eq. (4) in the form

$$\frac{p_n(\vec{x}(s))}{2} = \int_{x_1}^{x_2} \widetilde{G}_n(\vec{x}(s), \vec{x}(s')) p_{n'}(\vec{x}(s')) ds' + \widetilde{p}_n^{inc}(\vec{x}(s)) \qquad (7)$$

and Eq. (1) as

$$p(\vec{x}) = \int_{x_1}^{x_2} \widetilde{G}(\vec{x}, \vec{x}(s')) p_{n'}(\vec{x}(s')) ds' + \widetilde{p}^{inc}(\vec{x}). \qquad (8)$$

An important consideration[17] in the numerical treatment of Eq. (7) is the behavior of $\widetilde{G}(\vec{x}, \vec{x}(s'))$ as $x' \to x$ and $z' \to z$ with the rough surface smoothly approaching the mean plane $z = 0$. In this case, the term $\widetilde{G}_n(\vec{x}, \vec{x}(s'))$ becomes increasingly singular due to the image source at $(x', -z')$. As is shown in Ref. 17, the integrand can be regularized by subtracting an analytic expression from $\widetilde{G}_n(\vec{x}, \vec{x}(s'))$ and extracting a corresponding integral contribution. This contribution when moved to the left-hand side of Eq. (7) causes this side of the equation to become approximately equal to $p_n(\vec{x}(s))$ for $z' \to z$ and $z \approx 0$ [recall that with the definition of the incident field in this BIEM that $\widetilde{p}_n^{inc}(\vec{x}(s))$ on the flat surface is equal to twice the value of the corresponding incident field for the BIEM using the free space Green's function].

The problem with this method of using the half-space Green's function is that it is only valid for deformations of the boundary extending into the acoustic half-space. For outward deformations, the function $\widetilde{G}(x, z; x', z')$ of Eq. (5) now has two singularities within portions of the half-space and, in fact, because of this it cannot be referred to as a Green's function (Ref. 27, p. 353). In particular, if one considers a simple outward deformation and the reflection of this deformation into the waveguide, the function $\widetilde{G}(x, z, ; x', z')$ has two singularities in this resulting region. This means that in the derivation of the Helmholtz integral equation, Eq. (1), from Green's identity[28] (for the two-dimensional half-space volume V and the upper bounding surface S),

$$\int_V (p(\vec{x}')\nabla'^2 G(\vec{x};\vec{x}') - G(\vec{x};\vec{x}')\nabla'^2 p(\vec{x}'))dV'$$

$$= \int_S (p(\vec{x}')G_{n'}(\vec{x},\vec{x}') - G(\vec{x};\vec{x}')p_{n'}(\vec{x}'))dS', \qquad (9)$$

there will be certain values of $\vec{x}$ for which the first term in the volume integral of Eq. (9) yields $p(\vec{x}) - p(\tilde{\vec{x}})$. This will occur when the image point of $\vec{x}$ with respect to $z=0$, $\tilde{\vec{x}}$, is also in the volume V. Thus for outward deformations, the Helmholtz integral equation resulting from Eq. (9) no longer has the form of Eq. (1). However, the method of Ref. 17 provides an useful method for benchmarking more general methods in the case of an inward deformation.

There are two basic steps in the computation of the pressure field at a field point $\vec{x}$ using a BIEM for general deformations of the upper surface. [For inward deformations, we can use Eqs. (7) and (8).] First there is the integral equation [e.g., Eq. (4)] to solve for the surface field $p_n(\vec{x}(s))$ and then the extrapolation of this field to $\vec{x}$ [e.g., using Eq. (1)]. However, even when the surface field has been computed from the solution of Eq. (4) for a general deformation, it is possible to use the half-space Green's function, using Eq. (8) to extrapolate the computed surface field [and it is only the values of $p_n(\vec{x})$ for $z \neq 0$ which are required]. This integral relation is not valid for receiver points whose negative image about $z=0$ lies within the rough half-space. However, this is often a small restriction on the receivers which can be considered. The advantage of using Eq. (8) is that it explicitly models the specular reflection of the incident field (i.e., the incident field as defined in this BIEM consists of the direct and upper flat surface reflection) and also any interactions of the field scattered by the roughness with the infinite flat surface. The integral relation of Eq. (1) can also be accurately used as long as the truncated interval is sufficiently large to accurately model the interactions of the incident and scattered fields with the infinite upper surface. When considering the computed solutions using a truncated BIEM, we will consider both methods of extrapolation.

In this paper, different types of incident fields will be considered. First, an untapered incident plane wave is considered. This is an incident field which is infinite in extent. Second, an incident tapered plane wave is considered. This type of incident field may be an appropriate model for some types of incident beams. Numerically, it has the advantage that the amount of incident energy outside of a finite domain can be considered negligible. In this case, the approximation differences which arise from the truncation of the incident field should be small for a numerical domain which is approximately the same size (or slightly bigger) as the effective extent of the tapered incident field. An example of an incident tapered plane wave constructed from a superposition or integral of plane waves[6] is described mathematically by

$$p^{inc}(x,z) = \frac{1}{\sqrt{\pi}\Delta\theta} \int_{-\pi/2}^{\pi/2} \exp\left(-\frac{(\theta' - \theta_{inc})^2}{(\Delta\theta)^2}\right)$$

$$\times \exp(ik(x\sin\theta' + z\cos\theta'))d\theta', \qquad (10)$$

where $\Delta\theta \equiv 2/(kg\cos\theta_{inc})$, g is a specified tapering dis-

tance and θ_{inc} is the specified angle of incidence. In Ref. 6 it is shown how an accurate analytical approximation to this integral can be derived. In this paper we will simply numerically evaluate the normal derivative of the expression of Eq. (10), thus avoiding any questions about the accuracy to which the approximate analytic expression for Eq. (10) satisfies the wave equation.[6,16] In the case that the half-space BIEM is used, the corresponding incident field $\tilde{p}^{inc}$ is given by

$$\tilde{p}^{inc}(x,z) = \frac{1}{\sqrt{\pi}\Delta\theta} \int_{-\pi/2}^{\pi/2} \exp\left(-\frac{(\theta' - \theta_{inc})^2}{(\Delta\theta)^2}\right)$$

$$\times \{\exp(ik(x\sin\theta' + z\cos\theta'))$$

$$- \exp(ik(x\sin\theta' - z\cos\theta'))\}d\theta'. \qquad (11)$$

Although tapering of the incident field can be very useful in the numerical modeling of scattering from a rough surface, there are some issues which need to be considered. First, in some theoretical investigations, it is the scattering response as a function of the plane-wave incident angle which is desired. The tapered incident field is effectively a spectral average over plane waves at different incident angles. In order to obtain a very small interval of angular (spectral) averaging, then the spatial extent of the incident field must be large. Conversely, if one wishes to have an incident field of very small spatial extent, then a large amount of angular spectral averaging results. For many situations, a good compromise between the spectral and spatial extent can be found. However, if the grazing angle of interest is very small, then there is the additional problem when one wishes the Gaussian weighting to be negligible for $\theta' = \pm\pi/2$ [i.e., the spectral values at the endpoints of the integral of Eq. (10) should be small]. In this case, it is necessary to make g suitably large.[16]

Another type of incident field which is untapered is that from a point source (although the amplitude of the field from the source does decrease with the square root of the range from the source). For a source distant from the scattering region and at a shallow depth, the field incident on the roughness interval is similar to that from an untapered plane wave at a shallow grazing angle. Finally, we will consider a case of a point source where there is also a bounding bottom seabed.

B. The scattering chamber approach

In Ref. 17 it was outlined in the Appendix how a system of integral equations could be set up for the case of an outward deformation. However, for the general case of both inward and outward deformations, this method becomes complicated. In this paper, we propose a simpler approach. In Fig. 1, a deformed surface is shown and an arc of a circle which contains the deformation area within it. Within the area defined by C_1 and C_2 we have the set of integral relations for points $\vec{x}(s)$ along the curves:

$$\int_{C_1} G(\vec{x}(s);\vec{x}(s_1'))p_{n'}(\vec{x}(s_1'))ds_1'$$

$$+\int_{C_2} G(\vec{x}(s);\vec{x}(s_2'))p_{n'}(\vec{x}(s_2'))ds_2'$$

$$-\int_{C_2} G_{n'}(\vec{x}(s);\vec{x}(s_2'))p(\vec{x}(s_2'))ds_2' = p(\vec{x}(s)) \qquad (12)$$

with $p(\vec{x}(s)) \equiv 0$ along C_1. We now consider an integral relation on the exterior of C_2 (i.e., consider the observation point just below the circular arc) and we use the half-space Green's function $\tilde{G}$ so that contributions from the flat portion of the upper surface are zero,

$$-\int_{C_2} \tilde{G}(\vec{x}(s);\vec{x}(s_2'))p_{n'}(\vec{x}(s_2'))ds_2'$$

$$+\int_{C_2} \tilde{G}_{n'}(\vec{x}(s);\vec{x}(s_2'))p(\vec{x}(s_2'))ds_2' + p^{inc}(\vec{x}(s)) = p(\vec{x}(s)).$$
$$(13)$$

The combined set of integral equations gives enough conditions to solve for p_n along C_1 and for p and p_n along C_2. In Eqs. (12) and (13) the normal is pointing out of the area enclosed by C_1 and C_2 and s_1' and s_2' are taken to increase in a clockwise direction (e.g., s_1' increases with increasing x along C_1). This system of equations is a Fredholm system of the first kind and hence it is not clear how numerically stable the system will be. In fact, we found numerically that for small incident angles, the numerical convergence as a function of the discretization size was poor. One can differentiate the equations with respect to the normal to derive a new set of equations. However, this leads to integral kernels with various singularities. Instead of using this integral approach, we now describe a wave-field superposition technique which can be used with the domain of Fig. 1.

The total pressure field within $\Omega \equiv C_1 \cup C_2$ (including the flat portions of C_1) is taken to be represented by a set of point sources located a fraction of a wavelength outside of Ω. This type of approach is used in a variety of other wave-field superposition applications.[21–26] We take a discretization of C_1 and C_2 and consider the point sources offset along the normal direction from the corresponding discretization points along C_1 and C_2 (some of these sources are shown as squares in Fig. 2). Let us denote the number of point sources offset from C_1 as N_1 and the number of point sources offset from C_2 as N_2, with $N_3 \equiv N_1 + N_2$. At a point $\vec{x}$ in Ω we can write for the total pressure field

$$p(\vec{x}) = \frac{i}{4}\sum_{q=1}^{N_3} a_q H_0^1(kR_q(\vec{x})), \qquad (14)$$

and for its normal derivative along a curve

$$p_n(\vec{x}) = \frac{-i}{4}\sum_{q=1}^{N_3} ka_q H_1^1(kR_q(\vec{x}))\frac{\partial}{\partial n}R_q(\vec{x}). \qquad (15)$$

Here, the normal derivative is computationally taken to have a positive z component (since we will be simply satisfying a
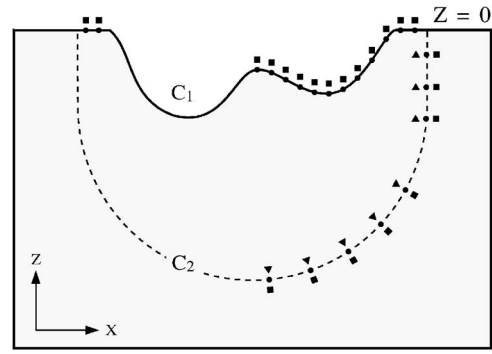
FIG. 2. A schematic of the geometry for the rough surface scattering problem showing some of the locations of the computational point sources for the wave-field superposition method. The outer ones (squares) are used to represent the total field within $C_1 \cup C_2 \equiv \Omega$ and are free space point sources. The sources just within C_2 represent the scattered field in the half-space outside of Ω. They also include the negative image about $z=0$ in their representation so that the scattered field automatically satisfies the pressure release condition along $z=0$.

continuity equation at a set of points, the polarity of the normal vector is not important as long as it is consistently defined for all field quantities). In Eqs. (14) and (15) we use free space point sources to represent the total field within Ω. The pressure release condition along the top surface of Ω will used as one of the boundary conditions to determine the coefficients of the point sources. Similarly for a point in the half-space outside of Ω we consider a set of N_2 point sources just inside C_2 to represent the scattered field (some of these sources are shown as triangles in Fig. 2), but in this case they will take the form that satisfies the pressure release condition along the surface $z=0$,

$$\tilde{p}(\vec{x}) = \frac{i}{4}\sum_{q=1}^{N_2} b_q[H_0^1(kR_q(\vec{x})) - H_0^1(k\tilde{R}_q(\vec{x}))] \qquad (16)$$

and

$$\tilde{p}_n(\vec{x}) = \frac{-i}{4}\sum_{q=1}^{N_2} kb_q\left[H_1^1(kR_q(\vec{x}))\frac{\partial}{\partial n}R_q(\vec{x})\right.$$

$$\left. - H_1^1(k\tilde{R}_q(\vec{x}))\frac{\partial}{\partial n}\tilde{R}_q(\vec{x})\right]. \qquad (17)$$

Here R_q is the range from the source point $\vec{x}_q$ and $\tilde{R}_q$ is the range from the image point above the upper flat surface. We can now set up systems of equations for a_q and b_q by imposing the various constraints at discrete values $\vec{x}_m$: the pressure release condition along C_1 and the continuity of the expressions [Eqs. (14)–(17)] for the pressure field and its normal derivative along C_2,

$$p(\vec{x}_m) = 0, \quad \vec{x}_m \in C_1, \qquad (18)$$

$$p(\vec{x}_m) = \tilde{p}(\vec{x}_m) + p^{inc}(\vec{x}_m), \quad \vec{x}_m \in C_2, \qquad (19)$$

$$p_n(\vec{x}_m) = \tilde{p}_n(\vec{x}_m) + p_n^{inc}(\vec{x}_m), \quad \vec{x}_m \in C_2. \qquad (20)$$

The first boundary condition yields N_1 equations, the second boundary condition, N_2 equations, and the third boundary condition, N_2 equations, for a total of $2N_2+N_1=N_3+N_2$

John A. Fawcett: Rough surface scattering

equations with $N_3 + N_2$ unknowns. One could attempt to satisfy the boundary conditions at more discrete points than unknowns, resulting in an overdetermined system of equations. For the case of a rigid boundary condition, Eq. (18) is replaced with the constraint $p_n(\vec{x}_m) = 0$ and the representations of Eqs. (16) and (17) are modified to satisfy the rigid boundary condition along $z = 0$.

In the case that there is a lower seabed boundary, then the expressions of Eqs. (16) and (17) are replaced by

$$\widetilde{p}(\vec{x}) = \sum_{q=1}^{N_2} b_q G^w(\vec{x}; \vec{x}_q) \tag{21}$$

and

$$\widetilde{p}_n(\vec{x}) = \sum_{q=1}^{N_2} b_q G_n^w(\vec{x}; \vec{x}_q), \tag{22}$$

where G^w denotes the waveguide Green's function (this is similar to the approach taken in Ref. 24 for a target in a perfect waveguide). Also the incident fields in Eqs. (19) and (20) are now those corresponding to the waveguide. With the representations of Eqs. (21) and (22), the solution exterior to C_2 satisfies the waveguide boundary conditions at both the top and seabed interfaces. We will consider a Pekeris waveguide, that is, a constant sound speed water column bounded above by a pressure release surface and below by a homogeneous seabed. The Green's function for such a waveguide can be expressed as a wave-number integral.[29] However, in this paper we will use the complex image representation for propagation in a waveguide[30] to express the Green's function as a set of point sources with complex amplitudes and positions. This has the advantage that the expression is computationally efficient and also allows one to simply replace the half-space Green's function in the computational code with a sum over the complex images. We can use this same expression to evaluate the field anywhere in the waveguide for ranges up to several waveguide depths; otherwise, one could use the wave-number integral or the discrete modal representation for ranges sufficiently far from the scattering region.

III. NUMERICAL EXAMPLES

In the numerical examples we will consider a variety of boundary integral approaches as well as the wave-field superposition method. First, we will implement truncated versions of Eq. (4) and Eq. (1) to solve and to extrapolate the solution; we will call this method "I." Second, we will use Eq. (4) to solve for the normal derivative along the truncated surface and then use Eq. (8) (i.e., using the half-space Green's function) to extrapolate the solution to the field points. This will be denoted as method "II." In this case, the incident field corresponds to the incident field of method I with the addition of the field reflected from a flat surface along $z = 0$. The wave superposition/scattering chamber (WSSC) method, using Eqs. (18)–(20), is also numerically implemented. Once the point source coefficients a_q and b_q in Eqs. (14)–(17) have been determined, then for a specified receiver position, the appropriate sum, Eq. (14) or (16) is used, depending on whether the receiver is interior or exte-
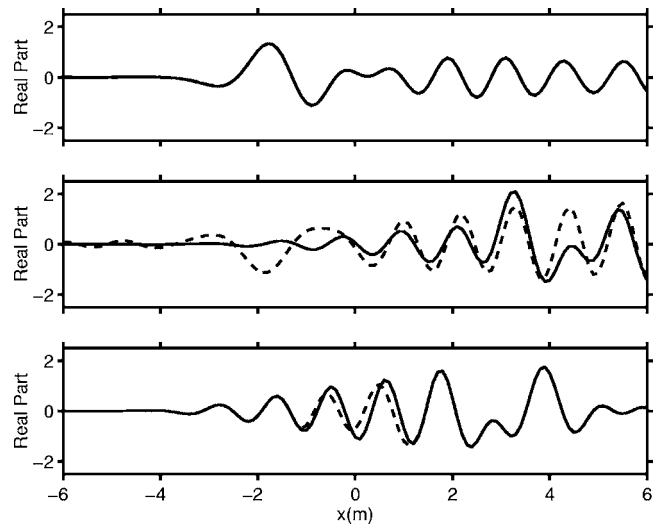
FIG. 3. The top panel shows the scattered pressure field (the difference between the total field and the field which would exist in the half-space for a perfectly flat top surface) as computed by method I (solid line) and the half-space method (dashed line) for an inward deformation for receivers along the line $z = -2.5$ m. The second panel shows the same results for the same deformation but outwards pointing. The third panel shows the results of using method I (solid line) and method II (dashed line) for a line of receivers at $z = -1.0$ m and the outward deformation. The incident field is a tapered ($g = 9$ m) plane wave incident at $60°$.

rior to the domain Ω. Finally, there is the method of using the half-space Green's function in both the integral equation and in the extrapolation [i.e., Eqs. (7) and (8)]. This will serve as the benchmark method in the case of inward deformations.

For the first set of computations we consider a deformation of the form,

$$z(x) = -0.5a\left(1 + \cos\left(\pi \frac{x}{2w}\right)\right), \quad -\frac{w}{2} < x < \frac{w}{2}. \tag{23}$$

For $a > 0$ the deformation is pointing into the fluid half-space and for $a < 0$ it is pointing outwards. The frequency for the computations is 1.5 kHz, the width of the deformation is $w = 10$ m, and the maximum height (negative direction) is 1.2 m. For the truncated BIEM the deformation area is discretized with 101 points and an additional 160 points are used on either side of the deformation (i.e., an extra 320 points) so that the total horizontal extent is 42 m with an equal Δx spacing of 0.1 m. The total system of equations has dimension 421×421. For the BIEM using the half-space Green's function [Eqs. (7) and (8)] there is no need to extend the computational domain outside of the deformation region and hence only 101 points are used. As a first numerical example we illustrate the problems which arise when using the half-space Green's function as the kernel of the BIEM. First, in Fig. 3 we show a comparison of the computed scattered pressure field for a line of receivers located along the line $z = -2.5$ m and a tapered plane wave incident at an angle of $60°$ using the standard truncated integral equation approach (method I—solid line) and using the half-space Green's function approach (dashed line). We use the definition that the scattered field is the difference between the total field and the field which would exist for the flat surface half-

space. The results show very good agreement as is expected since the deformation is completely inward pointing. We now repeat the numerical computations with the sign of the deformation reversed so that it is pointing strictly outward. As can be seen the agreement is very poor. This is because the use of the half-space Green's function is not valid in this case. Finally, for the outward deformation, we extrapolate the solution computed along the upper surface, using the truncated integral equation, to the receivers located along the line $z=-1.0$ m using the truncated integral, Eq. (1) (method I—solid line in Fig. 3) and using the half-space Green's function (method II—dashed line in Fig. 3). The solutions are in very good agreement except for the interval where the line of receivers is within the reflected image of the deformation. This region, where method II is inaccurate, is the interval predicted by theory. This example confirms that the half-space Green's function approach yields the correct solution for inward deformations and can serve as an accurate benchmark solution in this case. However, the approach breaks down for deformations which contain outward components. It also shows that the extrapolation using the half-space Green's function is accurate as along as the receivers are not within the region of the negative image of the deformation.

In the next series of computations we consider $a>0$ in Eq. (23) (i.e., inward pointing) and thus the half-space Green's function approach can serve as a benchmark in these cases. The computational parameters for the integral equations are the same as for the previous example. We stress that, as discussed previously, the computational results from method I would approximate the benchmark solution accurately if the computational domain was suitably extended. Here, we are instead fixing the size of the domain to what we consider a reasonable size and examining the resulting agreement with the benchmark solution provided by the half-space Green's function approach. For the WSSC method, we use 100 points along the rough surface and 20 points along the flat portion (10 on each side of the roughness area). The same number of points is used along the curve in the water. Thus, the resulting system of equations has a dimension of 360×360. By numerical experimentation it was determined that displacing the point sources a distance of 0.12 m from the computational curves yielded excellent results and this value is used for all the computations.

The scattered pressure field is computed for a line of receivers located at $z=-2.5$ m and for -20 m $<x<40$ m and 80 m $<x<100$ m (Fig. 4, top and bottom panels, respectively). An untapered plane wave is incident at an angle of 70°. In the top panel of Fig. 4 we show the resulting scattered pressure field (real part) computed using the benchmark method (solid line) and method I (dashed line). In the second panel we show the corresponding results for method II (dashed line) and in the third panel the results obtained using the WSSC approach (dashed line). As can be seen all three methods give good agreement with the benchmark solution in the region of the roughness (-5 m $<x<5$ m). The results of method I begin to diverge from the benchmark solution for values of x away from the center of the computation grid. In Fig. 4 we see that the results of method II and the scatter-
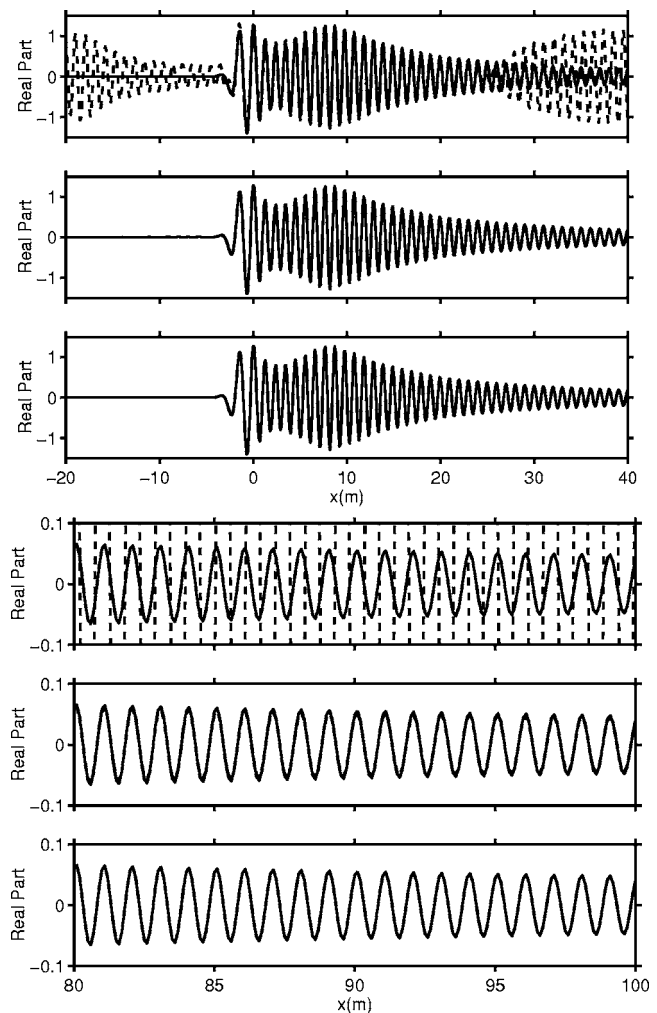
FIG. 4. The top three plots show the scattered pressure field computed at a line of receivers $z=-2.5$ m and -20 m $<x<40$ m for plane wave incident at 70°. The deformation is 1.2 m inwards, the frequency is 1500 Hz. The solution is computed (top) using the exact integral formulation (solid line) and the truncated formulation, method I (dashed line); (middle) using the exact integral formulation (solid) and method II extrapolation of the solution; and (bottom) the exact integral formulation (solid) and the WSSC solution. The same computations are then repeated for 80 m $<x<100$ m.

ing chamber approach (WSSC) both show excellent agreement with the benchmark solution for 80 m $<x<100$ m. The method II extrapolation uses only the computed values of $p_n(\vec{x})$ along the section of roughness and it explicitly accounts for the interactions of the incident and scattered fields with the flat portions of the upper surface. Method I shows poor agreement with the benchmark method for these longer ranges. As we will see with the next example, the main source of the disagreement is due to the abrupt truncation of the incident untapered plane wave.

In Fig. 5 the previous computations are repeated for a tapered ($g=9$ m) plane wave incident at an angle of incidence of 70°. In this case we find that once again the results of method II and the WSSC are in excellent agreement with the benchmark solution (scattered field). The results of method I are now in much better agreement with the benchmark solution, but one can see that for 80 m $\lesssim x \lesssim 100$ m the results are still not in good agreement. This is because, even with the tapered incident field, there is still an error in mod-
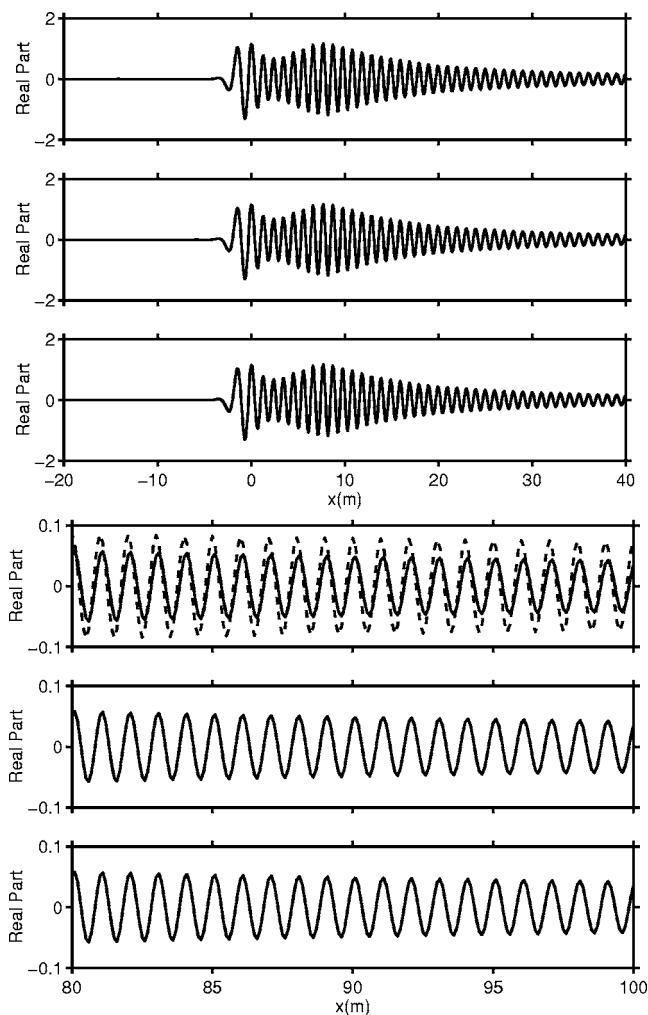
FIG. 5. The top three plots show the scattered pressure field computed at a line of receivers $z=-2.5$ m and $-20 < x < 40$ m for a tapered ($g=9$ m) plane wave incident at 70°. The deformation is 1.2 m inwards, the frequency is 1500 Hz. The solution is computed (top) using the exact integral formulation (solid line) and the truncated formulation, method I (dashed line); (middle) using the exact integral formulation (solid) and method II extrapolation of the solution; and (bottom) using the exact integral formulation (solid) and the WSSC solution. The same computations are then repeated for 80 m $< x <$ 100 m.

eling the interaction of the scattered field with the upper surface when the computational domain is not sufficiently large.

As the angle of incidence becomes closer to the horizontal, then the truncated solution begins to diverge from the exact solution even in the case that the extrapolation method of method II is used. In Fig. 6 we show the results (scattered field) of method II (dashed) and WSSC (dashed) compared with the benchmark solution (solid) for an untapered plane wave at an angle of incidence of 85°. For this example, the agreement of method II with the benchmark solution is poorest for -20 m $\leq x \leq 40$ m and is noticeable, but not as bad, for 80 m $\leq x \leq 100$ m. The WSSC solution is in excellent agreement with the benchmark solution everywhere. It is possible to consider a tapered incident field at this angle of incidence but, as discussed, in order to obtain an incident tapered wave which has a negligible spectral amplitude at $\theta' = \pi/2$ the spatial extent of this incident tapered plane wave needs to be large.
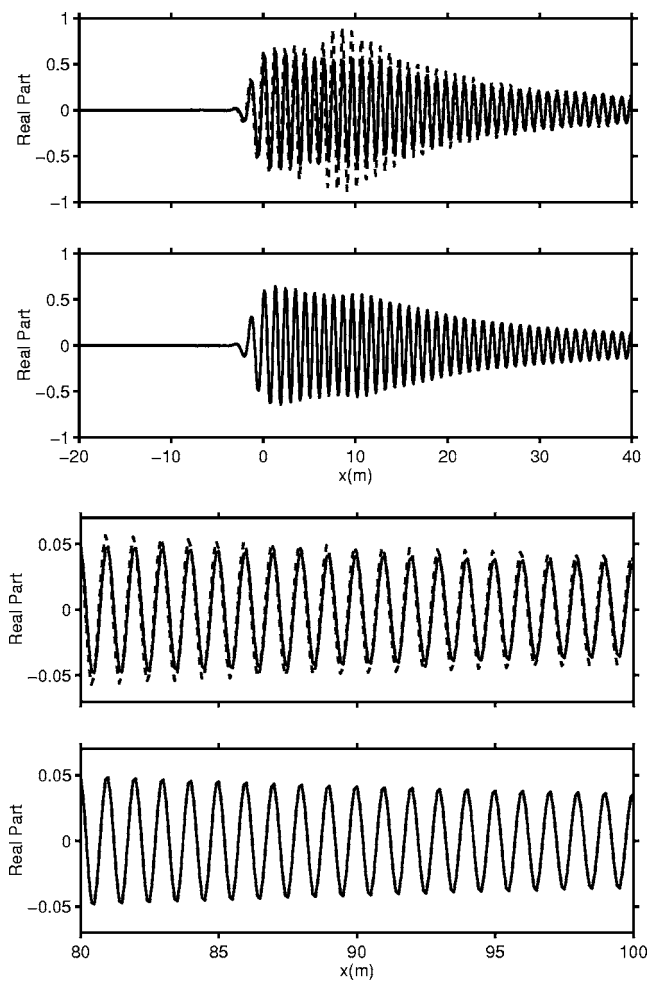
FIG. 6. The top two plots show the scattered pressure field computed at a line of receivers $z=-2.5$ m and -20 m $< x <$ 40 m for a plane wave incident at 85°. The deformation is 1.2 m inwards, the frequency is 1500 Hz. The solution is computed (top) using the exact integral formulation (solid line) and method II (dashed line) and (bottom) the exact integral formulation (solid) and the WSSC solution (dashed line). The same computations are then repeated for 80 m $< x <$ 100 m.

Finally, a distant point source at $x_s=-1000$ m and $z_s=-5$ m is considered. In this case the agreement of method II with the benchmark solution (scattered fields) is very poor (the levels are incorrect by about a factor of 3). However, as is seen in Fig. 7 the agreement of the benchmark solution with the WSSC is still excellent.

Thus far, with the exception of the first example, we have considered a rough surface which consists of only an inward deformation. In this case, we could use an exact BIEM approach as the benchmark solution. For the more general case of a rough surface with both inward and outward deformations we will need to rely upon a truncated solution, using either method I or II to extrapolate the integral equation solution, as the "benchmark." For the results shown in Fig. 8 we consider a plane wave incident at 70° with tapering ($g=9$ m) and use the truncated integral equation approach. We will use the half-space Green's function (method II) to extrapolate the solution to a line of receivers at $z=-2.5$ m. A sinusoidal deformation,

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

John A. Fawcett: Rough surface scattering   1355
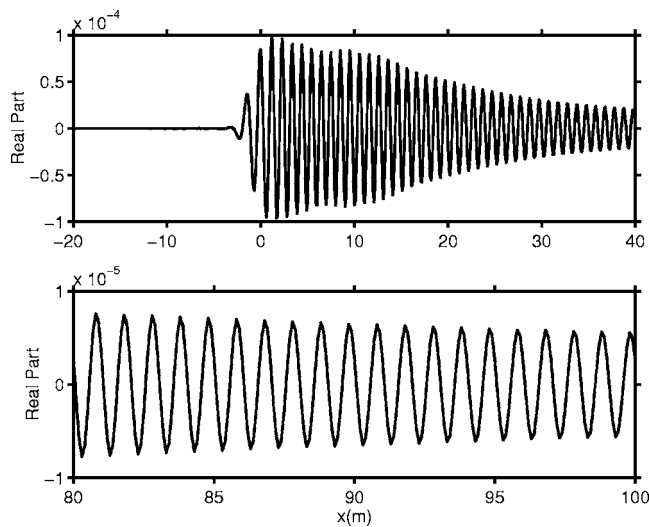
FIG. 7. The benchmark scattered pressure field solution (solid line) and the WSSC results (dashed line) for a point source located at $x_s=-1000$ m and $z_s=-5$ m and a line of receivers at $z=-2.5$ m.

$$z(x) = a \sin\left(\frac{2\pi x}{w}\right), \quad -5 \text{ m} \leq x \leq 5 \text{ m}, \quad (24)$$

is considered where $a=1.2$ m and $w=10$ m. As can be seen from Fig. 8, the agreement between the benchmark solution (solid line) provided by method II and the WSSC approach is excellent (scattered fields are shown). This gives us confidence that the WSSC approach is accurate for general deformations of the upper surface.

In Fig. 9 we compute a grid of values for the sinusoidal deformation of Eq. (24) and a point source located at $x=-15$ m and $z=-0.5$ m; this corresponds approximately to an angle of 88° with respect to the center of the scattering surface. In the first panel we have subtracted off the function,

$$\frac{i}{4}[H_0^1(kR) - H_0^1(k\tilde{R})]. \quad (25)$$

This perhaps does not make much physical sense for points within the outward deformation, but does emphasize the dif-
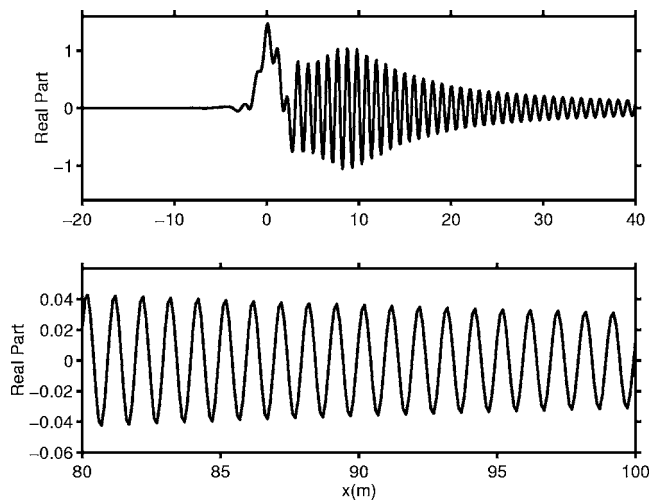
FIG. 8. The benchmark scattered pressure field solution: method II for a tapered plane wave incident at 70° ($g=9$ m) (solid line) for the sinusoidal deformation of the upper surface and the WSSC results (dashed line) for a line of receivers at $z=-2.5$ m.
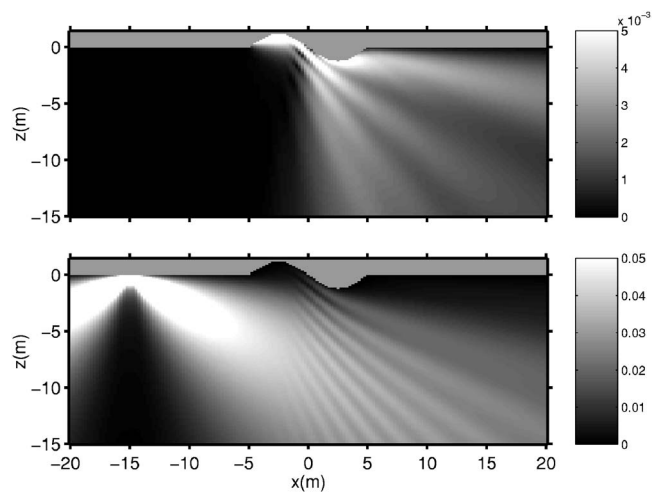
FIG. 9. The upper panel shows the scattered pressure field (total–flat surface solution) and the bottom panel shows total pressure field for a point source located at $x=-15$ m and $z=-0.5$ m. The frequency is 1500 Hz.

ference between the field in the presence of the scattering surface and the field which would exist for the perfectly flat surface. In the second panel the amplitude of the total field is shown. The interference effects between the scattered and incident fields is evident in the half-space below and to the right of the scattering region.

We now repeat the half-space computations but for the source in a 15-m-deep waveguide. The bottom has a sound speed $c_p=1700$ m/s and $\rho=1.5$ g/cm³. There is an attenuation of 0.5 dB/λ. A set of four complex images is used to represent the reflected field from the seabed and this set is used to construct a sequence of images for the waveguide.[30] We construct the set to account for up to an including two seabed reflections. This set is then used in the computational code to replace the two-term half-space Green's function. For the bounded waveguide, energy will be incident upon the scattering region from a large distribution of angles. The scattered field (incident subtracted from total) and the total fields are shown in Fig. 10. Because of the interactions of the incident and scattered fields with seabed boundary, the resulting fields have a more complicated spatial structure than for the unbounded half-space. It is important to emphasize the ease with which the unbounded half-space scattering method was modified to include the effects of the waveguide propagation, simply by using the appropriate Green's function for the exterior of the scattering chamber. In this case, by using a complex image approximation it was very easy to simply add an extra loop to the half-space code to include the extra complex image positions.

IV. SUMMARY

In this paper we have described some of the issues associated with the modeling of scattering from a finite interval of surface roughness. In particular, for a distant point source or for an untapered plane wave with a shallow grazing angle (and, possibly, distant receiver points from the area of roughness) the extent of the computational domain will need to be large in order for a truncated integral equation approach to yield an accurate approximation. A tapered incident field can
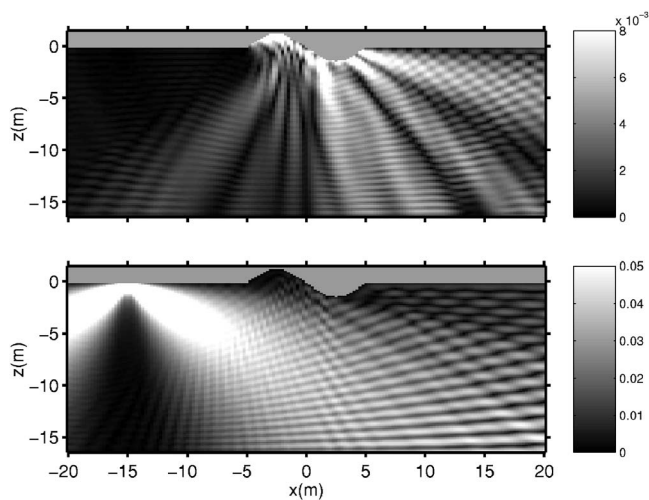
FIG. 10. The upper panel shows the scattered pressure field (total-flat surface solution) and the bottom panel shows total pressure field for a point source located at $x=-15$ m and $z=-0.5$ m. The frequency is 1500 Hz. There is a bottom seabed at a depth of 15 m with $c_p=1700$ m/s, $\rho=1.5$ g/cm^3, and attenuation of 0.5 dB/λ.

be used in some situations to mitigate some of the truncation issues. Also, we showed that using the half-space Green's function to extrapolate the integral equation solution to receiver points could be advantageous. However, there remain cases (for example, a distant point source and distant receivers) where a truncated integral equation approach, even using the half-space Green's function for extrapolation, will not yield an accurate approximation unless a very large computational domain is used. The use of the half-space Greens' function in the BIEM allows one to restrict the integral equation to just the area of roughness and avoid these large computational domains. However, this approach is only valid if the deformation of the upper surface is only inward. For the more general case, we proposed a wave superposition/scattering chamber approach. This was shown to be accurate for all cases considered and has the advantage that the size of the computational domain does not depend upon the locations of the source or receivers. As well, we showed that it is easily extendable to the case of a waveguide with a flat bottom seabed.

[1] S. T. McDaniel, "Sea surface reverberation: A review," J. Acoust. Soc. Am. **94**, 1905–1922 (1993).
[2] L. J. Ziomek, "Generalized Kirchhoff approach to the ocean surface-scatter communication channel. Part I. Transfer function of the ocean surface," J. Acoust. Soc. Am. **71**, 116–126 (1982).
[3] A. N. Venetsanopoulos and F. B. Tuteur, "Stochastic filter modeling for the sea-surface scattering channel," J. Acoust. Soc. Am. **49**, 1100–1107 (1971).
[4] D. Dowling and D. Jackson, "Coherence of acoustic scattering from a dynamic rough surface," J. Acoust. Soc. Am. **93**, 3149–3157 (1993).
[5] E. Y. Harper and F. M. Labianca, "Scattering of sound from a point source by a rough surface progressing over an isolvelocity ocean," J. Acoust. Soc. Am. **58**, 349–364 (1975).
[6] 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).
[7] E. Thorsos and S. Broschat, "An investigation of the small slope approximation for scattering from rough surfaces," J. Acoust. Soc. Am. **97**, 2082–2093 (1995).
[8] D. A. Miles, R. N. Hewitt, M. K. Donnelly, and T. Clarke, "Forward scattering of pulses from a rough sea surface by Fourier synthesis of parabolic equation solutions," J. Acoust. Soc. Am. **114**, 1266–1280 (2003).
[9] R. A. Stephen, "Modeling sea surface scattering by the time-domain finite-difference method," J. Acoust. Soc. Am. **100**, 2070–2078 (1996).
[10] R. S. Keiffer and J. C. Novarini, "A time domain rough surface scattering model based on wedge diffraction: Application to low-frequency backscattering from two-dimensional sea surfaces," J. Acoust. Soc. Am. **107**, 27–39 (2000).
[11] R. L. Holford, "Scattering of sound waves at a periodic, pressure-release surface: An exact solution," J. Acoust. Soc. Am. **70**, 1116–1128 (1981).
[12] D. F. McCammon and S. T. McDaniel, "Application of a new theoretical treatment to an old problem: Sinusoidal pressure release boundary scattering," J. Acoust. Soc. Am. **78**, 149–156 (1985).
[13] S. Gaffet and M. Bouchon, "Source location and valley shape effects on the P-SV near displacement field using a boundary integral equation—discrete wavenumber representation method," Geophys. J. Int. **106**, 341–355 (1991).
[14] H. Takenaka, B. L. N. Kennett, and H. Fujiwara, "Effect of 2-D topography on the 3-D seismic wavefield using a 2.5-D discrete wavenumber-boundary integral equation method," Geophys. J. Int. **124**, 741–755 (1996).
[15] F. J. Sanchez-Sesma and F. Luzon, "Seismic response of three-dimensional alluvial valleys for incident P, S, and Rayleigh waves," Bull. Seismol. Soc. Am. **85**, 269–284 (1995).
[16] R. T. Marchand and G. S. Brown, "On the Use of Finite Surfaces in the Numerical Prediction of Rough Surface Scattering," IEEE Trans. Antennas Propag. **47**, 600–604 (1999).
[17] T. W. Dawson and J. A. Fawcett, "A boundary integral equation method for acoustic scattering in a waveguide with nonplanar surfaces," J. Acoust. Soc. Am. **87**, 1110–1125 (1990).
[18] P. Gerstoft and H. Schmidt, "A boundary element approach to ocean seismoacoustic facet reverberation," J. Acoust. Soc. Am. **89**, 1629–1642 (1991).
[19] R. A. Stephen and S. A. Swift, "Modeling seafloor geoacoustic interaction with a numerical scattering chamber," J. Acoust. Soc. Am. **96**, 973–990 (1994).
[20] M. G. Imhoff, "Scattering of acoustic and elastic waves using hybrid multiple multipole expansions—Finite element technique," J. Acoust. Soc. Am. **100**, 1325–1338 (1996).
[21] M. Dravinski and T. K. Mossessian, "Scattering of plane harmonic P, SV and Rayleigh waves by dipping layers of arbitrary shape," Bull. Seismol. Soc. Am. **77**, 212–235 (1987).
[22] R. D. Miller, E. T. Moyer Jr., H. Huang, and H. Uberall, "A comparison between the boundary element method and the wave superposition approach for the analysis of the scattered fields from rigid bodies and elastic shells," J. Acoust. Soc. Am. **89**, 2185–2196 (1991).
[23] L. Song, G. H. Koopmann, and J. B. Fahnline, "Numerical errors associated with the method of superposition for computing acoustic fields," J. Acoust. Soc. Am. **89**, 2625–2633 (1991).
[24] A. Sarkissian, "Multiple scattering effects when scattering from a target in a bounded medium," J. Acoust. Soc. Am. **96**, 3137–3144 (1994).
[25] J. E. Murphy, G. Li, S. Chin-Bing, and D. King, "Multifilament source model for short-range underwater acoustic problems involving penetrable ocean bottoms," J. Acoust. Soc. Am. **99**, 845–850 (1996).
[26] M. G. Imhof, "Multiple multipole expansions for acoustic scattering," J. Acoust. Soc. Am. **97**, 754–763 (1995).
[27] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol. I* (Interscience, New York, 1953).
[28] M. C. Junger and D. Feit, *Sound, Structures and Their Interaction* (Acoustical Society of America, New York, 1993).
[29] G. Frisk, *Ocean and Seabed Acoustics: A Theory of Wave Propagation* (PTR Prentice-Hall, New York, 1994).
[30] J. Fawcett, "A method of images for a penetrable acoustic waveguide," J. Acoust. Soc. Am. **113**, 194–204 (2003).

Target phase: An extra dimension for fish and plankton target identification

Richard Barr[a)] and Roger F. Coombs[b)]

National Institute of Water and Atmospheric Research, Private Bag 14-901, Kilbirnie, Wellington, New Zealand

The acoustic signal backscattered from a fish in water, $\mathcal{L}_{bs}(f)$, at a frequency, f, differs from the incident signal in both magnitude, $|\mathcal{L}_{bs}(f)|$, and phase, $\arg[\mathcal{L}_{bs}(f)]$, and it has been common practice for many years to use the backscatter magnitude from individual fish as an aid to species identification. However, very little use has been made of the phase of the backscattered signal relative to that of the incident acoustic pulse. If the gross phase changes due to propagation through water are compensated for, the residual phase signature is found to contain useful target-specific information. The phase signature can be characterized by estimating the rate at which the echo phase changes, relative to the transmitted pulse, during the echo from a fish. Clear groups are produced when single fish targets from *in situ* data are plotted in complex target space (target strength versus target rate of change of phase) and this is explored here by computing the acoustic backscatter, in complex target space, from a series of simple large and small model fish targets, both with and without gas-filled swimbladders, using a Monte Carlo technique. It is shown that all the features found in the *in situ* data can be explained in terms of the size and attitude of the fish. The modeling was carried out at the frequency of 38 kHz, which pertained to the *in situ* data, and the specific results only apply to this frequency. However, the complex target approach is generally applicable to fish target strength analysis independently of frequency. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1979503]

I. INTRODUCTION

Historically, the acoustic identification of fish has tended to follow two different paths. One system of identification, which is performed at great range (>100 m), makes use of standard fishing industry echogram displays. The shape of fish *marks* on the echogram, their size, and distance from the sea bed or hills, etc. have all been used as indicators of specific fish types (LeFeuvre *et al.*, 2000; Rose and Legget, 1988; Scalabrin *et al.*, 1996). The advantage of this approach is that it can be used in *real time* while fishing. The disadvantage is that few features of fish shoals exhibit high species discrimination (Scalabrin *et al.*, 1996). The second approach is to locate a single fish on the edge of large schools and, by measuring their target strength *in situ*, use the magnitude of the target strength as an indication of the species. This technique necessitates deploying a transducer reasonably close to the fish (<100 m) so that single targets can be resolved (e.g., Do and Coombs, 1989). A useful extension to this technique is to perform the measurements at a number of different frequencies simultaneously and weight the responses on the different frequencies in order to optimize the species selectivity of the summed response (Kloser *et al.*, 2002).

The target strength, measured at a frequency f, is commonly defined (e.g., Medwin and Clay, 1998) in terms of the modulus of the complex acoustical backscattering length, $\mathcal{L}_{bs}(f)$,

$$\text{TS}(f) = 20 \log[|\mathcal{L}_{bs}(f)|/1 \text{ m}](\text{dB}). \tag{1}$$

Previous target identification studies using target strength have only used the modulus of the backscatter data as in Eq. (1). However, the backscattering length, $\mathcal{L}_{bs}(f)$, is a complex quantity and one can measure not only the amplitude of the backscattered pulse from a target compared to the incident pulse, $|\mathcal{L}_{bs}(f)|$, but also the phase, $\arg[\mathcal{L}_{bs}(f)]$. How this phase changes with elapsed time, or range, if we assume propagation in sea water, gives us a new dimension in fish target studies.

Experimentally, phase information has previously been used to determine the position of a target in the transducer beam (e.g., Ehrenberg, 1979) and to help discriminate between single and multiple targets (Soule *et al.*, 1995; 1996; 1997). In these cases, however, phase differences were measured between the different elements of a four-quadrant transducer. In this paper the phase of the echo from a target is determined by reference to the transmitted pulse. If the gross phase changes due to propagation through water are removed, the residual phase signature is found to contain useful target specific information. Barr *et al.* (2000) created a descriptor, the target rate of change of phase, which displayed such target specific characteristics largely independent of range. They defined the target rate of change of phase as the rate at which echo phase changed, relative to the trans-

mitted pulse, during the echo from the fish. Barr *et al.* (2000) also selected the maximum echo amplitude, and the rate of phase change, evaluated at maximum amplitude, to define a target in complex target space. Using the rate of change of phase as the new descriptor and referencing the scale to the time of the maximum echo amplitude effectively removes range effects and gives a parameter whose properties are dominated by the physical characteristics of the target. They then used this complex target strength information to discriminate, *in situ*, between different fish groups in which the magnitude of the target strength was the same but the phase response different.

Barr *et al.* (2000) based target identification largely on associations between trawl catches and expected acoustic responses from conventional target strength studies. The expected phase response of orange roughy, in particular, was based on some preliminary target modeling work, and this paper greatly extends this with the aim of increasing understanding of the observed acoustic responses by modeling the scattering from simple, but realistic, physical models of the fish and other scatterers found in trawls and other sampling devices.

II. DATA

The data were collected using NIWA's Computerised Echo Sounder Technology (Barr *et al.* 2000; Coombs *et al.*, 2003). In this system, signals received by the transducer are amplified and then digitized at 100 kHz and fed to a digital signal processor (DSP56002), where they are complex demodulated, filtered, and decimated to give a final sample rate of 10 kHz. The output voltage, A, from the receiver for any fish is, from Eq. (1), proportional to $|\mathcal{L}_{bs}(f)|$, where f is 38 kHz. We have characterized this using the maximum amplitude, A_{max}, of a received pulse. In our past target strength studies (e.g., Coombs and Barr, 2004) we have estimated A_{max} by locating the highest-amplitude sample in a pulse envelope and fitting a parabola to it and the two points to either side (three points in all). The target strength, in decibels, is derived from the square of the amplitude with calibrations referenced to a tungsten carbide sphere. Similarly, the signal phase also changes across a received pulse as a consequence of the range to the target and, more importantly, the acoustical properties of the target, $\arg[\mathcal{L}_{bs}(f)]$. As for amplitude, we fit phase change, ϕ, to a quadratic relationship with time, t, where

$$\phi = a + bt + ct^2. \tag{2}$$

Hence the rate of change of phase with time is

$$\frac{d\phi}{dt} = b + 2ct. \tag{3}$$

For a sequence of samples ($\Delta t = 100 \ \mu s$) at $t = -2, -1, 0, +1, +2$, where $t = 0$ is the time of the maximum amplitude, A_{max}, our rate of change of phase descriptor, ϕ'_{max}, is simply b from Eq. (3). Together, A_{max} and ϕ'_{max} describe a target in complex target space. To facilitate a comparison with the earlier work, we use the velocity of sound in seawater to convert the unit of rate of change of phase

from degrees per second to degrees per meter.

In Barr *et al.* (2000) we presented two plots of target strength versus the rate of change of phase derived from many thousands of *in situ* single targets. The recordings were made at short range (<100 m) and in deep water (≈ 800 m) on the Chatham Rise, about 1000 Km to the east of mainland New Zealand. The sounding frequency for all of these data was 38 kHz and the pulse length 12 cycles (0.32 ms). To cover more of the pulse, in estimating ϕ'_{max}, we used two samples to either side of the peak (five points in all). The predominant large fishes in the trawl catches from Barr *et al.* (2000) were orange roughy (*Hoplostethus atlanticus*) and black oreo (*Allocytus niger*). Small fishes such as myctophids (*Myctophum spp*) were often caught in the trawl meshes as were many other types of small organisms such as salps.

The two plots are reproduced here as the upper and lower panels of Fig. 1, which show clear groupings of the fish targets. The upper panel is from an oreo survey and has three clear target groups. The lower panel, from an orange roughy survey, shows four groups, although these are perhaps not quite so clear. Target group **a**, in the upper panel, was ascribed to black oreo, group **b** to myctophids, and group **c** was thought to be euphausids or other small planktonic targets. The data for this figure were collected at a location where trawl catches were predominantly of black oreo. The targets in the lower panel of Fig. 1 came from an area where catches were mainly of orange roughy and the arc-shaped feature **a** was ascribed to this species. The feature **b** was again attributed to myctophids and **d** to euphausids or other small sea creatures. Peak **c** was also provisionally assigned to orange roughy as a possible result of a failure in the single target detection algorithm. As noted in the Introduction, our aim is to reproduce similar target responses to those of Fig. 1 using simple fish models, and to determine how basic parameters of these models affect the target response. In the following discussions the plots of target strength versus the rate of change of the phase are referred to as plots in complex target space.

III. MODELING

A. Anatomically based acoustic backscatter models

We have characterized the trawl catches using black oreos, orange roughy, and myctophids and these three species are shown in Fig. 2 (from Paul, 1986). The bodies of the fish have been modeled as general spheroids and the swimbladders as prolate spheroids. The parameters of the model fish and the propagation media, applicable to a depth of 1000 m, are given in Table I. Unless otherwise stated, model parameters are all taken from Table I. The swimbladders of the myctophid and black oreo are assumed to be gas filled, while the orange roughy is assumed to have a swimbladder filled with wax esters (Phleger and Grigor, 1990). The data on the propagation medium, the orange roughy, and the black oreo are taken from Barr (2001) and on the myctophid from Marshall (1960). These models are intended to capture the essential acoustic scattering features of fish representing large and small species with gas-filled swimbladders and a large spe-

FIG. 2. Drawings of the three fish modeled in this study taken from Paul (1986). Note that the typical adult oreo and roughy are ≈35 cm in length, whereas adult myctophids are only ≈7.5 cm.
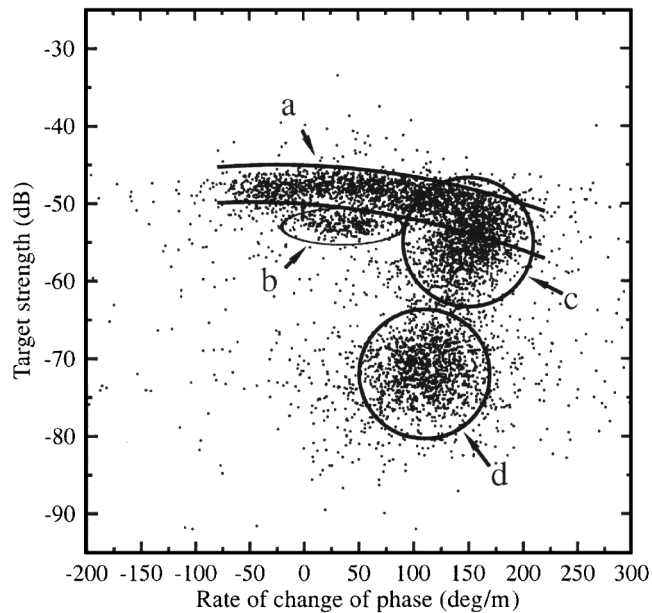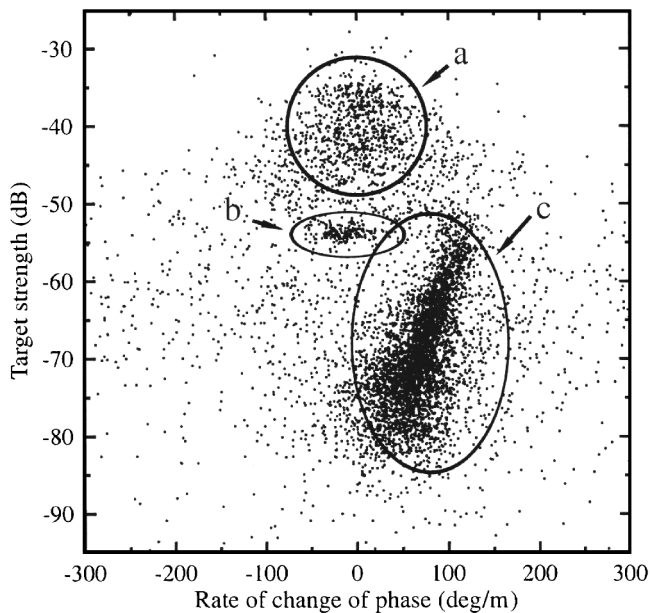
FIG. 1. Plots of acoustic target strength versus the target rate of change of phase (complex target space) (from Barr *et al.*, 2000). Upper panel: Data recorded during an oreo survey. Lower panel: Data recorded during an orange roughy survey.

cies with a wax-filled swimbladder; they are not meant to accurately model all the biological features of the fish. Myctophids in particular have very complicated swimbladder morphology that tends to shift from being entirely gas filled in juveniles to increasingly wax ester filled with age (Butler and Pearcy, 1972; Neighbors and Nafpaktitis, 1982). Morphology also varies between different species, and Butler and Pearcy (1972) grouped the swimbladders of eight myctophid species into three types: gas-filled, fat invested, and atrophied or reduced, although in all of these species, small specimens had thin-walled gas filled swimbladders.

We assume that all the dimensions of our fishes scale linearly and that the body proportions remain constant. For oreo and roughy modeling this seems reasonable since the

data in Fig. 1 were derived from adult fishes whose lengths had typical standard deviations of only 10%. Thus, scaling linearity is only required to hold good over a limited length range. The possible effects of the model limitations are discussed later.

The Kirchhoff method was used to evaluate the target strength of the spheroids (Medwin and Clay, 1998; Foote and Francis, 2002). The results have been found to be accurate when the acoustic wavelength is small compared to the length of the swimbladder or fish body (Foote and Francis, 2002). The technique should thus give acceptable accuracy for all the fish bodies studied here, but will be near the limits of its applicability for the small swimbladder of the myctophid. To calculate the backscatter from a tilted ellipsoid, containing a nonconfocal and ellipsoid core, is not a simple problem, and we have modeled the swimbladder and fish body separately following Furusawa (1988).

B. Fluid sphere models

Because many of the targets in Fig. 1 are likely to be small (groups **c** and **d**) and therefore at the limit of the applicability of the Kirchhoff method, an alternative modeling approach, appropriate to small scatterers, was also implemented. We assume that these small targets are much smaller than the wavelength of the incident acoustic radiation (4 cm for our *in situ* data). This means that the overall shape of the target is not as important as the volume in determining its scattering properties (Strasberg, 1953; Love, 1978). The

TABLE I. Physical parameters of the spheroidal fish models and sea water.

Parameter	Units	Oreo	Oreo	Roughy	Roughy	Myctophid	Myctophid	Water
Organ	...	bladder	body	bladder	body	bladder	body	sea
Length	cm	5.2	22.5	7.5	30.5	1.5	6.8	...
Height	cm	2.3	13.5	1.85	16.5	0.5	1.1	...
Width	cm	2.3	5.0	1.85	7.1	0.5	0.65	...
Speed of sound	m/s	335	1548	1525	1548	335	1548	1484
Density	Kg/m^3	126	1060	903	1050	126	1060	1032
Orientation	deg	22	...	12.0	...	12	...	...
Total length	cm	...	30.0	...	...	...	7.5	
Standard length	cm	...	...	...	35.0	...	...	...

small prolate spheroid swimbladder of our myctophid (with an aspect ratio of 3:1) can be expected to have resonance frequency and peak target strength responses (in broadside orientation) to within 6% of those of a sphere of the same volume (Feuillade and Werby, 1994). We therefore take all our plankton model targets to be spherical and composed of fish flesh or gas-filled. As we have only acoustic information about the real targets, a spherical model seems a reasonable first approximation.

We used the approach of Anderson (1950) to determine the complex target strength of the spherical targets of fish flesh and Love (1978, 1993) for the spherical air bladders. Love (1978) derived an expression for the complex target strength of a gas bladder encased within a viscous, heat-conducting shell of fish flesh. However, in a later paper (Love, 1993) he showed that, for blue whiting, the outside diameter of the shell had little effect on the target strength and that thermal and surface tension effects were negligible. He went on to produce a simplified formula that is effectively that of Anderson (1950) with the addition of viscous damping in the flesh surrounding the gas bladder (Medwin and Clay, 1998). Although this model has the disadvantage that it only applies to the gas bladder oscillating in its fundamental mode, for the small targets considered here the sound radiation in higher-order modes is several orders of magnitude below that of the fundamental (Strasberg, 1953). Our spherical gas bladders (effectively bubbles) are therefore modeled using Love's (1993) simplified theory. This approach takes account of resonance in the gas bladder, while the Kirchhoff does not.

C. Generation of plots in complex target space

As noted above, the data in Fig. 1 were generated with a 38 kHz transmission of 12 cycles duration; this short pulse length being used to maximize resolution of single targets. By computing the complex spectrum, $P(f)$, of the pulse and assuming the transducer response, $T(f)$, to be the same as that given by MacLennan (1981), we can derive the complex response, $R(f)$, of a fish from Eq. (4) below:

$$R(f) = T(f)P(f)G(f), \qquad (4)$$

where $G(f)$ is an estimate of the complex target strength of the model fish, $\mathcal{L}_{bs}(f)$, of Eq. (1), evaluated using the Kirchhoff method. By taking the inverse Fourier transform of $R(f)$, and complex demodulating, we obtain the

pulse response of the model fish to insonification by the ping.

Figure 3 shows the pulse response of a model black oreo of 30 cm total length, when insonified in the dorsal aspect at zero tilt. The solid trace shows the amplitude response and the dashed trace the phase response. However, whereas the numerical results presented in Fig. 3 are continuous, the experimental pulse responses of Barr et al. (2000) were sampled at 10 kHz. As a result, measurements of the ping response, as plotted in Fig. 3, were only recorded every 7.5 cm (100 μs). Typical sample points at this resolution are shown by the open circles on the continuous amplitude and phase curves.

To simulate the in situ data in Fig. 1, we used Monte Carlo simulations assuming fish lengths and tilt angles in the fish population were normally distributed with mean length, L, and standard deviation, σ_L, and mean tilt, χ, with standard deviation, σ_χ. Fish were selected at random from this distribution, their ping response computed at a resolution of 7.5 cm and A_{max} and ϕ'_{max} estimated for each fish to give a single point. Many tens of thousands of points are needed to produce displays comparable to Fig. 1 and computation speed proved a limitation. To speed up the simulation, the algorithm used by Barr et al. (2000) was modified and the number of points used to fit the phase response reduced from five to three. From Fig. 3 it can be seen that reducing the number of points with the simulated responses will make little difference to the estimates of ϕ'_{max}. The simulation also allowed
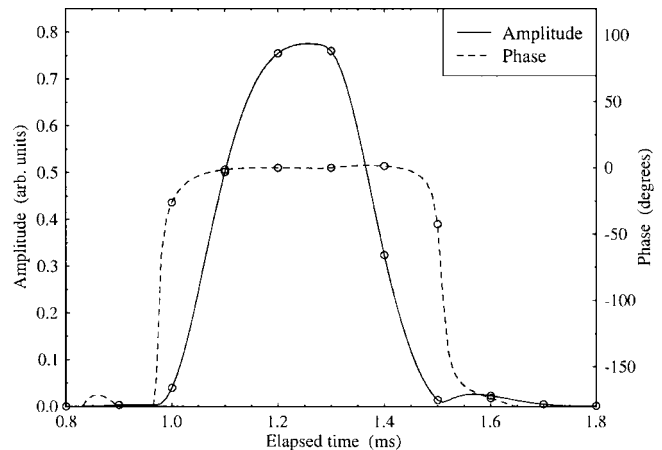
FIG. 3. Amplitude and phase response of a model black oreo of 30 cm total length, insonified in the dorsal aspect at zero tilt angle.
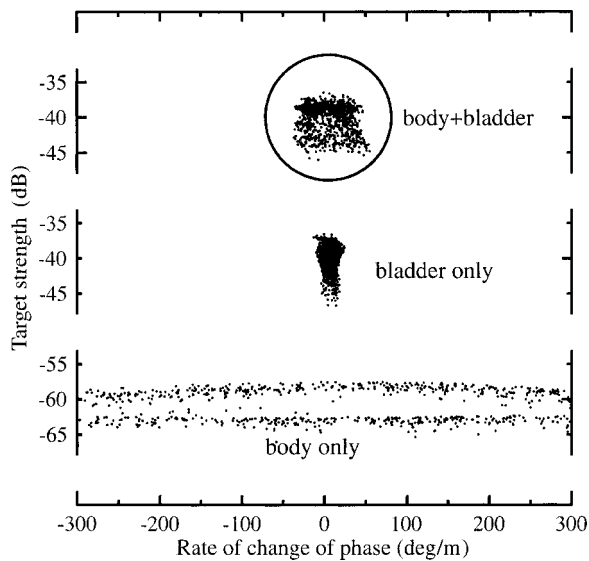
FIG. 4. Model black oreo targets plotted in complex target space. The circle has the same amplitude and phase as the circle around group **a** in the upper panel of Fig. 1.

FIG. 5. Complex target space plots of four black oreo distributions with different length and tilt standard deviations (σ_L and σ_χ, respectively).

for a uniform random movement of fish targets within the 7.5 cm resolution range.

IV. MODELING RESULTS

A. Black oreo

Figure 4 shows the complex target response for three black oreo simulations, where the mean length was taken to be 33.7 ± 1.8 cm and tilt angle $17 \pm 8°$ (Coombs and Barr, 2004), with the other parameters from Table I. The upper group of results, centered at ≈ -40 dB target strength and 0 deg/m rate of change of phase, is the result of modeling the combined effects of the swimbladder and the fish body. The shape of the group is very similar to, but somewhat less dispersed than, group **a** in the upper panel of Fig. 1, which was thought by Barr *et al.* (2000) to represent black oreo. The circles drawn around the groups in Figs. 1 and 4 have the same target strength and rate of change of the phase range to assist comparisons. It is possible that some of the extra spread in the *in situ* results is due to signal contamination by small targets at a level below that required to trigger rejection by the single target detection algorithms used.

The center group in Fig. 4 represents the response of the black oreo swimbladder alone. Comparing this with the combined response of the swimbladder and fish body above it, we can see that the main contribution of the fish body is to increase the spread in the rate of change of phase. The lowest trace in Fig. 4 is the response of the fish body alone and is typically ≈ 20 dB below that of the swimbladder. This ratio is much greater than that reported by Foote (1980) for cod and mackerel and is probably due to the difference in shape between these species. Oreos are much deeper (dorsal to ventral) in wavelength terms than cod, resulting in less coherence in the returned echo. The response of the fish body alone shows a very large spread on the phase axis for a very small change in target strength, producing an arc-shaped feature similar to curve **a** in the lower panel of Fig. 1, which
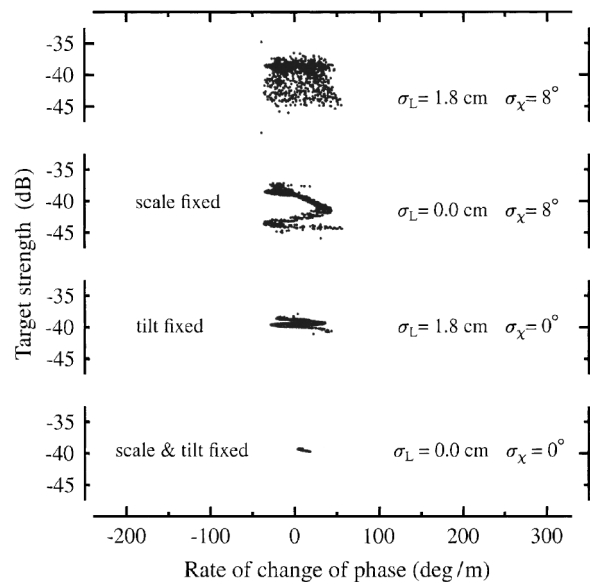
Barr *et al.* (2000) attributed to orange roughy. Based on the groupings in Figs. 1 and 4, we conclude that it is necessary to include the fish body when modeling the black oreo complex target strength.

Figure 5 shows the effects of different features of the black oreo model. The top group is a copy of the upper plot in Fig. 4 for a complete model of the fish body plus swimbladder. The standard deviations of the fish length and tilt angle distributions, from Coombs and Barr (2004), are shown beside each group. The second group in Fig. 5 uses the same complete fish model but the fish length has been fixed at 30 cm and only the tilt angle varied. It can be seen that the tilt alone is sufficient to cause the same spread in target strength and rate of change of phase as when both tilt and length vary, but the group shows a clearly defined structure. The third group down shows the effect of keeping the fish tilt angle constant at 0° and letting the fish length vary. This group has about the same phase spread as the upper two but the amplitude spread is reduced by a factor of 4. The bottom group in Fig. 5, for a fish of constant total length 30 cm and of constant tilt angle 0°, shows the target spread produced only by the 7.5 cm uncertainty in the start of digitization. The spread from this source is negligible in comparison with the variability in fish tilt and length. From Figs. 1 and 5 we conclude that black oreo tilt is of much more significance in determining target strength than fish length.

B. Orange roughy

The overall form of the target strength-length response of the spheroidal orange roughy body produced by the Kirchhoff method is almost identical to that generated by the slab-cylinder method of Clay and Horne (1994) [see Fig. 6(c) of Barr, 2001]. However, the magnitude is around 4 dB less for the Kirchhoff method. On the other hand, the target strength values for the swimbladder, with its assumed circular cross section, are within 1 dB when computed by the two different techniques. We believe the values given by the Kirchhoff

R. Barr and R. F. Coombs: Fish identification using target phase

FIG. 7. Orange roughy targets plotted in complex target space. Results are computed separately for the swimbladder and fish body. Note the increased *x*-axis spread compared to Fig. 6.
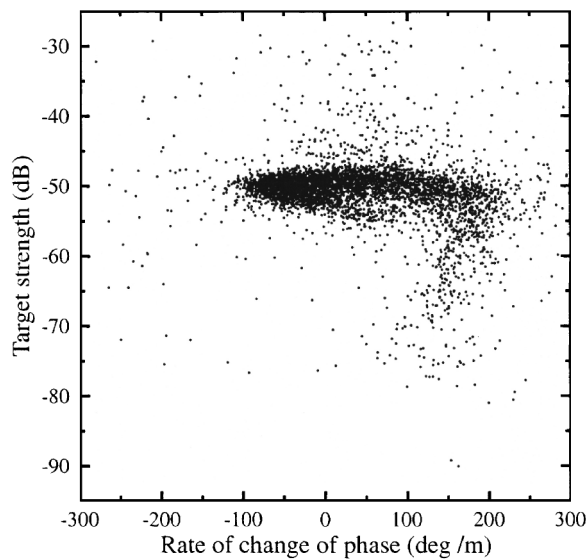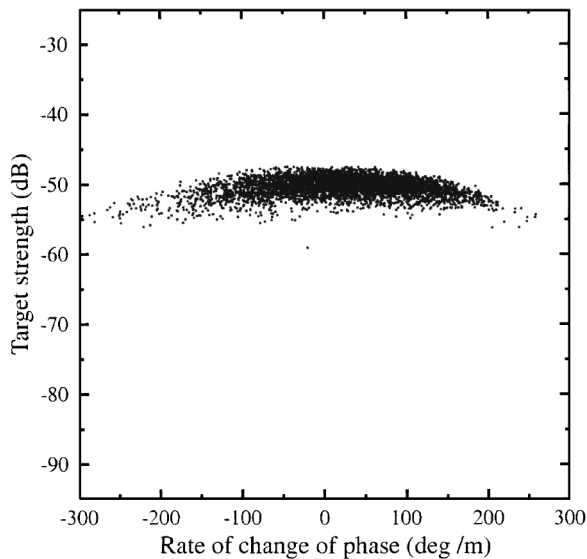
FIG. 6. Acoustic targets plotted in complex target space. Upper panel: Data computed for spheroidal model roughy. Lower panel: *in situ* data file d74 recorded 19 June 1999 during the orange roughy survey.

technique are to be preferred because the spheroidal model approximates the shape of an orange roughy better than the slab-cylinder model.

In order to compensate for the reduced contribution of the fish body to the total roughy target strength we have increased the contribution from the swimbladder to produce a target strength of similar magnitude to those measured experimentally (McClatchie *et al.*, 1999; Barr *et al.*, 2000). This has been accomplished by reducing the swimbladder tilt angle to 12° compared to the 18° assumed in Barr (2001). A close inspection of the frozen saggital section of the male orange roughy in Phleger and Grigor (1990) shows that while the rear half to two-thirds of the swimbladder is tilted ≈18° above the horizontal, the front half to a third is almost horizontal. A compromise tilt of 12° for the whole bladder therefore seems reasonable. With this tilt our simple model produces target strength values within 0.5 dB of the most recent *in situ* measurements (Doonan *et al.*, 2001), for roughy of ≈34.5 cm standard length.

The lower panel in Fig. 6 shows more than 4700 *in situ* targets plotted in complex target space recorded on the Chatham Rise in June 1999 at depths between 680 and 820 m. The upper panel shows 5000 targets produced using our spheroidal orange roughy model with parameters as given in Table I except for the swimbladder tilt angle, which was reduced to 12°, as described above. The only free parameters in the model were the model fish length and tilt distribution. To produce the results plotted in the upper panel of Fig. 6 required, a 0° mean fish tilt angle with a standard deviation of 2° and a fish length standard deviation of 0.5 cm. The tilt angle standard deviation appears unrealistically small at 2°, being only half that found to be necessary to model smooth oreo target strength distributions (Coombs and Barr, 2004) and a quarter that required for black oreos. The 0.5 cm fish length standard deviation required to fit the *in situ* data is also well below the 2.8 cm typical of roughy caught by trawling on acoustically surveyed marks by Barr *et al.* (2000). However, it must be remembered that our fish models are perfect spheroids. It is possible that by incorporating some irregularities on the surface of the fish model, as used, for example, by Clay (1991), larger standard deviations will be required to model the *in situ* data.

Figure 7 shows the individual contributions of the fish body and the swimbladder to the overall complex target strength presented in the upper panel of Fig. 6. Note that the range of the *x* axis in Fig. 7 is twice that in Fig. 6. It is clear from Fig. 7 that the dominant contributor to the arc-shaped phase response in Fig. 6 is the fish body, but that the very rapid phase variation for the fish body alone is moderated by the presence of the swimbladder. The dimensions of the swimbladder are much smaller in terms of acoustic wavelength than those of the fish body. The comet-like tail on the swimbladder response results from the variation in swimbladder response with tilt angle [see Fig. 6(a) of Barr 2001].

While the "dot plots" in Figs. 1, 6, and 7 give a rough idea of target distributions in complex target space, it is difficult to get any idea of relative target concentration. This
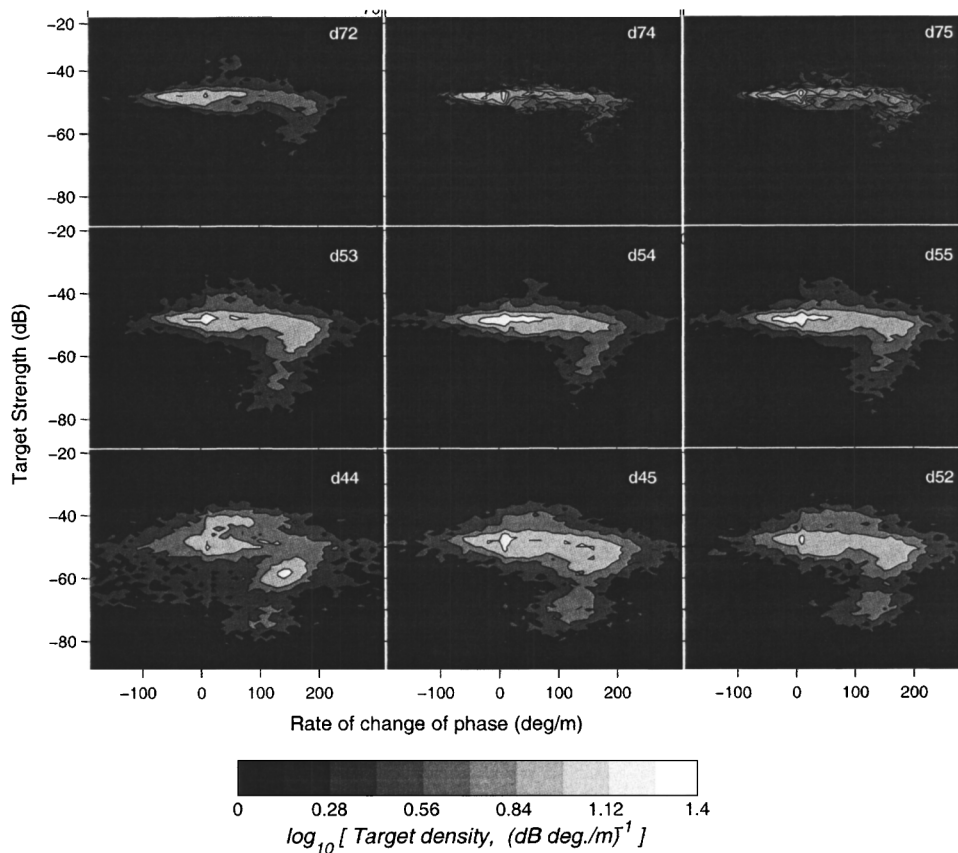
FIG. 8. *In situ* data recorded during the 1999 orange roughy survey. The targets are plotted in complex target space with the gray scale indicating the target density.

problem has been overcome in Fig. 8, where the *in situ* data for nine further target strength runs are plotted as gray shaded contours, equally spaced in the logarithm of target density. The highest target concentrations are white and the lowest black. All these data were recorded during the 1999 orange roughy survey, over marks that were later found by trawling to contain orange roughy. It was initially thought that the arc-like bands around −50 dB were necessarily the roughy but this feature, as described later, can also be produced by a different target type. The data in panels d44, d52, and d53 also show a clear group near −40 dB, probably produced by oreos. The weaker groupings (−60 to −70 dB) with large positive rates of phase change (100 to 150 deg/m) will be discussed later.

Figure 9 shows contours of target density in complex target space computed solely from our spheroidal orange roughy models. The results in each panel are computed for 10,000 roughy, described largely by the data in Table I, but with fish length (L) and fish tilt angles (χ) normally distributed with the parameters given in the respective panel. The arc-shaped group in the central panel (cl3) has a shape and size very similar to the arc-shaped plot in the center panel (d54) in Fig. 8, which appears to confirm the arc-shaped signature as orange roughy. However, there is still some doubt about the small spread in fish lengths (34.3±0.5 cm) needed to model the *in situ* data. The mean fish length determined from trawling, over the same path as the acoustic transects, was 34.2±2.7 cm, a spread some five times greater than in the model.

The models in Fig. 9 have been arranged to show the effects, in complex target space, of fish length, fish length

standard deviation, tilt angle, and tilt angle standard deviation. The three center panels, from left to right, show the effect of a 1.4 cm increase in mean fish length: as fish length increases, the peak in target density moves from positive to negative values of the rate of the change of phase. The three central panels running from bottom to top of Fig. 9 show the effect of increasing the standard deviation of the fish length. As might be expected from the effect of changing fish length, this increases the length of the arc.

Moving diagonally in Fig. 9 from top left to bottom right shows the effect of changing the mean fish tilt angle. This appears to have little effect on the arclength, but a significant effect on the spread of target strength with positive tilt angles, giving a larger spread. Moving diagonally in Fig. 9 from bottom left to top right shows the effect of increasing the fish tilt angle standard deviation, which in a similar fashion as previous, has only a slight effect on the arclength but produces an increased spread in target strength.

In summary, our simple orange roughy model is able to give a plausible explanation of the arc-shaped target distributions, as shown in the lower panel of Fig. 1, but seems to require fish distributions with unrealistically small standard deviations of fish length and tilt angle.

C. Myctophid

Figure 2 shows the shape of a typical myctophid in comparison with an orange roughy and a black oreo. Myctophids are much longer in comparison to their body depth than either oreos or roughy but the main difference is their size (see Table I); myctophids at around 7.5 cm are significantly
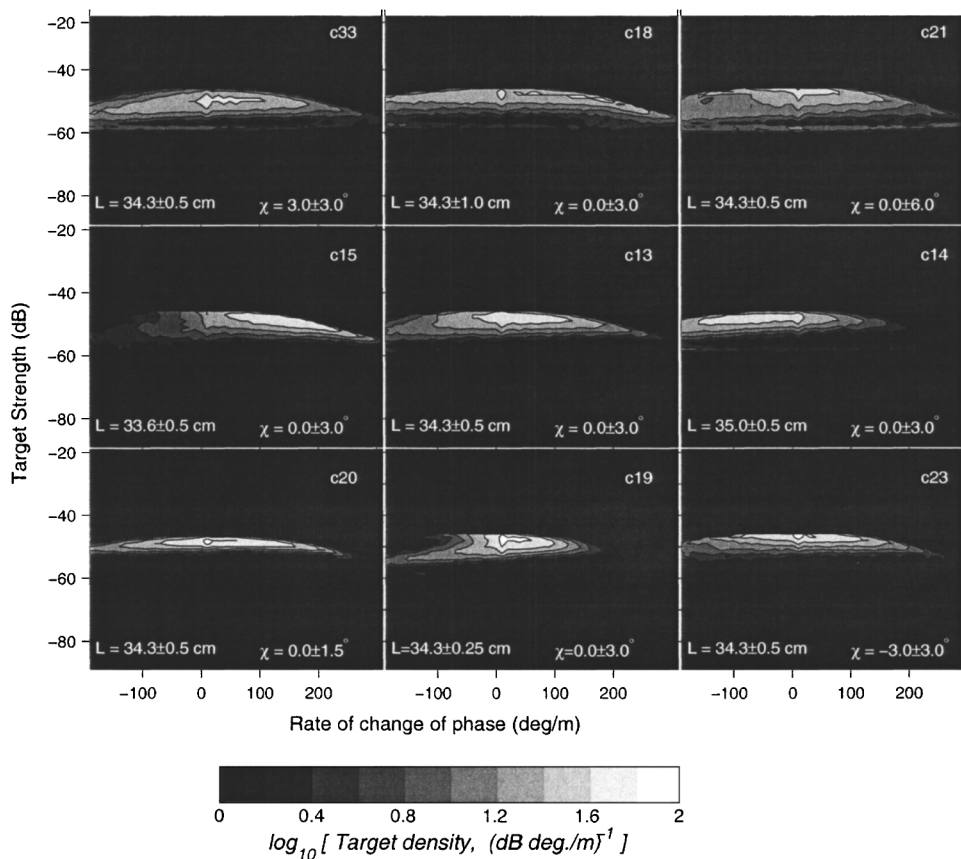
FIG. 9. Spheroidal model orange roughy targets plotted in complex target space with the gray scale indicating the target density.

smaller than the other two. "Myctophid" is used somewhat loosely here and is intended to include all small fishes with gas bladders. The aim of the myctophid model is to identify the target groups **b** in Fig. 1.

The myctophid target strength-length relationship from modeling is shown in Fig. 10. The computations assume all fish dimensions scale linearly but the target strength values, especially those of the swimbladder, are less reliable as the dimensions of the fish decrease and become comparable to
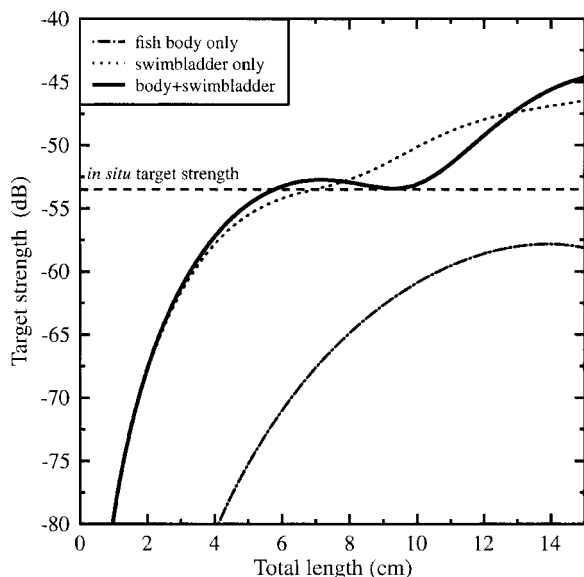
FIG. 10. Target strength of the model myctophid compared with the *in situ* target strength of groups **b** in Fig. 1.

the 38 KHZ wavelength ($\approx$4 cm). For fish lengths between 6 and 10 cm, the total fish target strength is almost constant at $\approx$$-53$ dB, independent of fish length. This constancy in target strength results from interference between sound reflected from the fish body and from the swimbladder and is within 0.5 dB of the value of -53.5 dB (the dashed horizontal line in Fig. 10) found by Barr *et al.* (2000) to be representative of more than 30 000 targets in groups such as **b** in Fig. 1, strong evidence that these are myctophid-like fishes. To further test this assignment, Monte Carlo analyses were made of the complex target strength of 5000 myctophid-like targets selected from fish distributions with different length and tilt angle distributions. Figure 11 shows the results for four different fish distributions. The upper group (inside the ellipse) was produced by myctophids with lengths uniformly distributed between 5.6 and 10.5 cm, swimming horizontally with zero tilt angle. The ellipse around the distribution has the same dimensions as the ellipse **b** in the lower panel of Fig. 1 and is located at the same position on the target strength scale. However, on the rate of change of phase scale (*x* axis) it is in a position intermediate between those of the ellipses in the upper and lower panels of Fig. 1.

Although this is a remarkably good fit, it is not clear how realistic the assumptions of uniform length distribution (5.6–l0.5 cm) and zero tilt angle are. While myctophid-like fishes have been caught in this length range in shallower waters on the Chatham Rise (McClatchie, 2002), the distributions were more Gaussian than uniform with a spread from 3-10 cm. The zero fish tilt angle required to fit the *in situ* data is unrealistic (Bone, 1973) and probably results from inad-
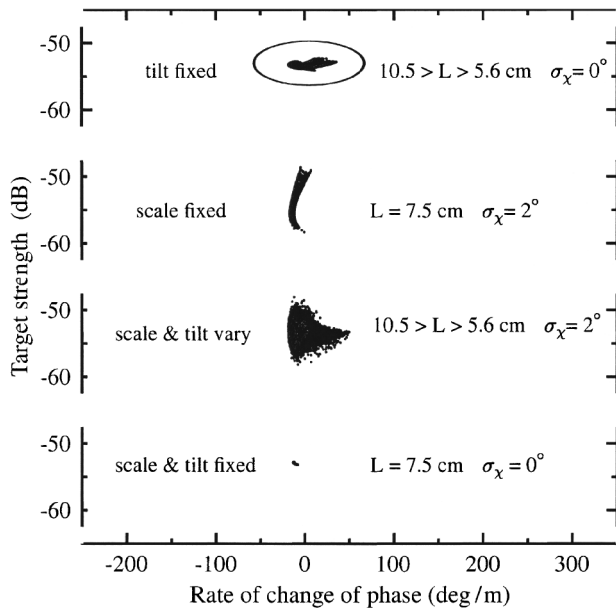
FIG. 11. Complex target space plots of four myctophid distributions with different length and tilt distributions.

FIG. 12. Complex target space plots of two simulated small target distributions. These targets have lengths uniformly distributed from 1% to 100% of the standard myctophid target (7.5 cm total length). The black dots apply to a gas bladder alone and the gray dots to a fish body alone. Targets are assumed to have zero mean tilt with a standard deviation of 5°.

equacies of the Kirchhoff method for targets small in comparison to the wavelength.

The complex target strengths of three further myctophid distributions are presented in Fig. 11. Comparing these with the uppermost group (uniformly varying length and fixed tilt), it is apparent that none accurately models the behavior of the observed fish distribution (groups **b** in Fig. 1). The second group from the top (fixed scale length of 7.5 cm and zero tilt angle with 2° standard deviation) clearly has too much spread in the target strength and insufficient spread in the rate of change of phase. Third from the top (both scale and tilt varying) produces a realistic spread in the rate of change of phase, but has too wide a spread in amplitude, especially for low values of the rate of change of phase. The lowest group (fixed fish tilt and scale length) clearly has insufficient spread in either amplitude or phase.

D. Small targets, Kirchhoff

To try and determine the practical limits of our myctophid model, we attempted to use it to reproduce the low target strength groups, **c** and **d**, in Fig. 1. We assumed that these targets were either gas bladders or fish flesh and that they were uniformly distributed within the length range from 1%–100% of our standard myctophid target (Table I) since this had a target strength at the upper limit of those in groups **c** and **d**. The black and gray patches in Fig. 12 show the complex target response of gas- and fish flesh- based "plankton," respectively. Although these cover the same target strength range as groups **c** and **d** in Fig. 1, they differ significantly with respect to phase. The *in situ* data show a clear progression of the increasing rate of change of phase with increasing target strength that is absent in the modeled results. The latter show the rate of change of phase to be almost constant and independent of target strength. As the majority of planktonic targets must be either flesh or gas
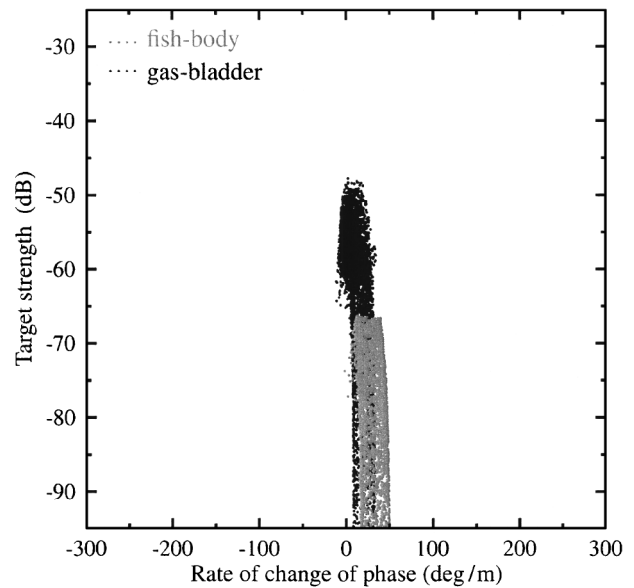
dominated, this is further evidence of the weakness of the Kirchhoff model for smaller targets.

E. Small targets, fluid sphere

In this section we present the results of the alternative fluid sphere model for small targets. The gray trace in the center and toward the bottom of Fig. 13 shows the response of 20 000 spherical fish flesh targets with radii uniformly distributed in the range 0.001–10.0 mm. The fish flesh and sea water parameters are those for the oreo body in Table I. The response is similar to that for the myctophid fish body in Fig. 12, with very little change in the rate of change of phase
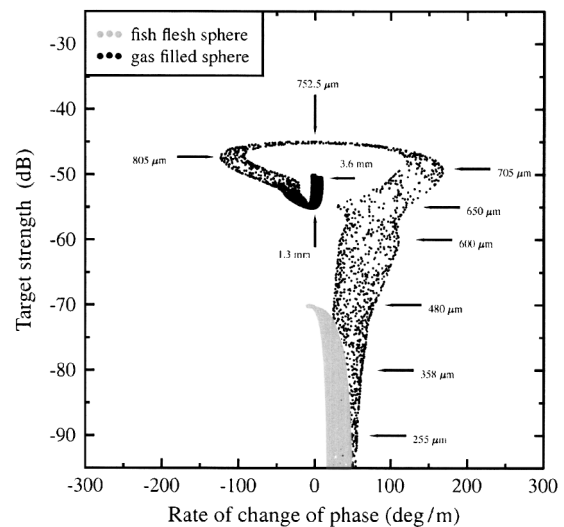
FIG. 13. Complex target space plots of two theoretical small spherical target distributions. The gray dots apply to a uniform distribution of fish flesh targets of radii in the range 0.001 to 10.0 mm. The black dots apply to a uniform distribution of gas filled targets with radii in the range 0.001 to 3.6 mm.

for a large change in target strength. Neither set of fish-flesh target responses exhibit the same increase in rate of change of phase with increasing target strength shown by groups **c** and **d** in Fig. 1 and, if anything, they show the reverse trend. It is unlikely therefore, that any but the smallest targets in Fig. 1 could be of a similar density to fish flesh.

What then produces the weak targets, as in group **c**, in the upper panel of Fig. 1? The answer is in the black target response in Fig. 13. This shows the complex target strength of a uniform distribution of l0 000 spherical gas bubbles of radius 0.001 to 3.6 mm, located at a depth of 800 m. The bubbles are assumed to be surrounded by a medium of zero viscosity and so either Anderson's (1950) or Love's (1993) formulations can be used to compute the response and this comparison was made as a means of checking the validity of our calculations. For the first time we can now see a gradual increase in the target rate of change of phase with increasing target strength, suggesting that the small targets in Fig. 1 are produced by small gas inclusions in either plankton or fishes. The large and rapid change in phase about the maximum target strength in Fig. 13 is the result of resonance of the gas spheres near 38 kHz. The largest-amplitude response occurs near the zero rate of change of phase, typical of a resonant system.

Printed around the black response curve in Fig. 13 are markers giving the radius of the air bladder at that region of the plot. It can be seen that to model groups **c** and **d** in Fig. 1 (between −85 and −50 dB) requires gas bladders between 0.7 and 1.4 mm diameter.

It is possible that the arc-shaped curve **a**, in the lower panel of Fig. 1, thought to be orange roughy by Barr *et al.* (2000), might, in fact, be produced by the resonant swim-bladders of myctophids. As noted earlier, myctophids have a complicated swimbladder morphology, but all of the juveniles had thin-walled gas-filled swimbladders. From Butler and Pearcy (1972), a 23 mm specimen had a volume equivalent to that of a sphere of radius ≈0.5 mm, whereas from Marshall (1960) a 75 mm myctophid had a radius of ≈3.6 mm. Thus, most of the question-mark-shaped curve in Fig. 13 could also be produced by a group of myctophids of varying species and size and consequently so could groups **c** and **d** in the Fig. 1 Chatham Rise data.

The rate of change of phase for spheres of radius greater than ≈1.3 mm radius is almost zero and remains so up to a radius of 3.6 mm, the limit of the calculations. At the same time the target strength increases from around −55 to −50 dB. The bubble model thus produces target strengths with similar features to the myctophid model using the Kirchhoff method (Figs. 10 and 11), and it is reassuring that both techniques produce similar results. The target strength values at the upper limit of the bubble model's validity are in reasonable agreement with those of the myctophid model at the lower limit of its validity.

Barr *et al.* (2000) warned against attributing rapid changes in phase purely to targets without gas-filled swim-bladders like orange roughy because similar signatures were found at much shallower depths where there were no orange roughy. The center panel in Fig. 14 shows one such distribution in complex target space recorded over a thick scattering layer on the Chatham Rise at depths between 400 and 500 m. These results were attributed to very dense targets overwhelming the single target algorithm, especially when individual targets were spaced at multiples of the sounding wavelength, in a similar manner to that suggested by Soule *et al.* (1996). Effectively, two small targets such as myctophids, separated in the vertical direction, were thought to be mimicking the reflections from the upper and lower surface of a fish without a gas-filled swimbladder. However, we now think it more likely that these arc-shaped targets were produced by resonant air bladders in either myctophids or siphonophores. The upper and lower panels in Fig. 14 show the response in a complex target space of 10 000 gas-filled spheres, with radii uniformly distributed between 0.35 and 0.65 mm, immersed in fish flesh. The lower panels are for fish flesh with zero viscosity and the upper panel a viscosity of 2 kg m^{-1} s^{-1} (20 Poise).

It is clear from Fig. 14 that resonant gas-filled spheres have signatures very similar to those of the *in situ* acoustic targets. The model results in the upper panel of Fig. 14 match the target strength of the *in situ* data very well but slightly underestimate the total phase spread in the arc-shaped region. On the other hand, the arc-shaped region in the lower panel, which was computed ignoring the viscosity of the fish flesh, has target strengths that are ≈4 dB too large and overestimates the spread in the rate of change of phase. A fish flesh viscosity of 1.0 kg m^{-1} s^{-1} (10 Poise) fits the spread in the rate of change of phase of the *in situ* data almost exactly, but slightly overestimates the target strength. The slightly high values of the simulated target strength could result from the assumption that the gas bladders are spheres when they are actually prolate spheroids. The latter, with aspect ratios of 3:1 and 10:1 and the same volumes as the spheres, would have target strengths reduced by ≈0.5 and 1.5 dB, respectively (Feuillade and Werby, 1994; Ye, 1997).

Finally, Fig. 15 shows the effect of the depth of immersion in the ocean on the response of gas-filled spheres enclosed in fish flesh for depths of 100, 330, and 1000 m. The viscosity of the fish flesh has been taken as 1.0 kg m^{-1} s^{-1} which was the value found to best fit the *in situ* data (Fig. 14). It can be seen that at all depths the shape in complex phase space is roughly that of a question mark and that the target strength at resonance increases monotonically with depth. This is just a feature of the larger spheres needed to support resonance at greater depths. However, the rate of change of phase, and hence the Q of the resonance, maximizes at ≈330 m. The depth at which this occurs depends on the relative contributions to damping caused by reradiation (scattering) and viscous effects (see Medwin and Clay, 1998). The resonant radius, r_{res}, the sphere radius giving maximum target strength, and the zero rate of change of phase, is shown for each depth in Fig. 15.

V. DISCUSSION

Our aim in this paper was to reproduce similar target responses to those in Fig. 1 using simple fish models, and to
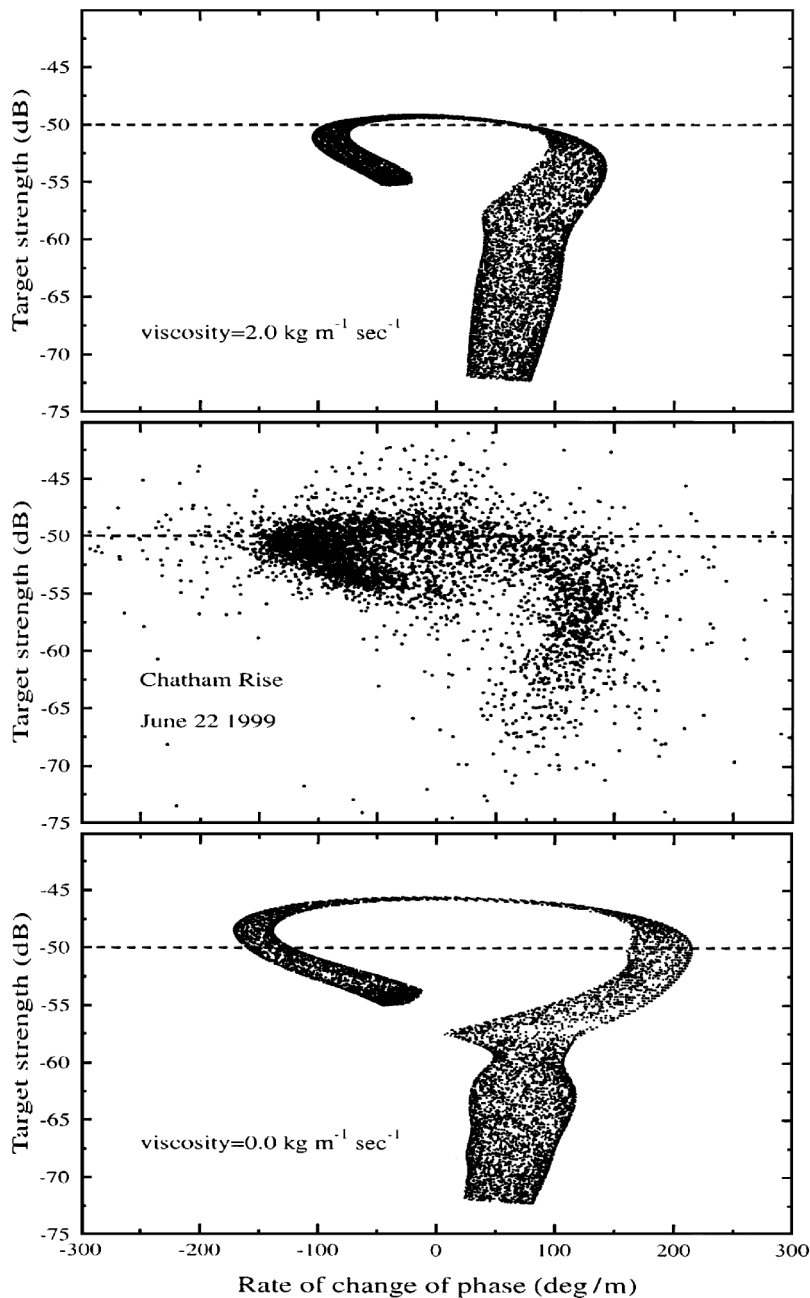
FIG. 14. Center panel: complex target space plot of *in situ* data recorded in a midwater scattering layer at a depth of ≈400 m during the orange roughy survey. Upper and lower panels: complex target space plots of l0 000 small gas-filled spherical targets with radii uniformly distributed between 0.35 and 0.65 mm. The results in the lower panel are for gas spheres encased in fish flesh of zero viscosity. Those in the upper panel are for gas spheres encased in flesh of viscosity $2.0 \ \text{kg m}^{-1} \text{s}^{-1}$.

determine how the parameters of these models affect the target response—how successful were we and what practical insights have we gained?

The black oreo model, a typical large deepwater fish with a gas-filled swimbladder, provides a good fit to the *in situ* data. In this case, much of the greater part of the scattering is from the swimbladder and this is naturally shaped like a prolate spheroid and is large relative to the wavelength, so the result is not unexpected. Figure 5 shows that the tilt angle has a bigger effect on the target strength than the fish length, so to improve estimates of oreo target strength for biomass estimation, more information on fish attitudes in the wild is needed. The large influence of tilt probably results from the large offset angle of the swimbladder in the body (Coombs and Barr 2004) and its large size relative to the wavelength. The fish length was found to have

a greater effect on the phase response rather than the amplitude.

For orange roughy the results were less clear cut. The arc-shaped response results primarily from interference between the signals scattered from the upper and lower surfaces of the fish body, but is too great without the moderating effect of scattering from the swimbladder. With this fish model we have been able to realistically match the pattern of the *in situ* data with respect to both amplitude and phase. However, this was only possible by using a fish length distribution that was much smaller than that of the fish caught in trawl samples and it was also necessary to adjust the angle of the swimbladder in a somewhat arbitrary fashion. Orange roughy are much more complex scatterers than oreos. The contrast in scattering between the body and swimbladder (Fig. 7) is not large, the swimbladder is only approximately
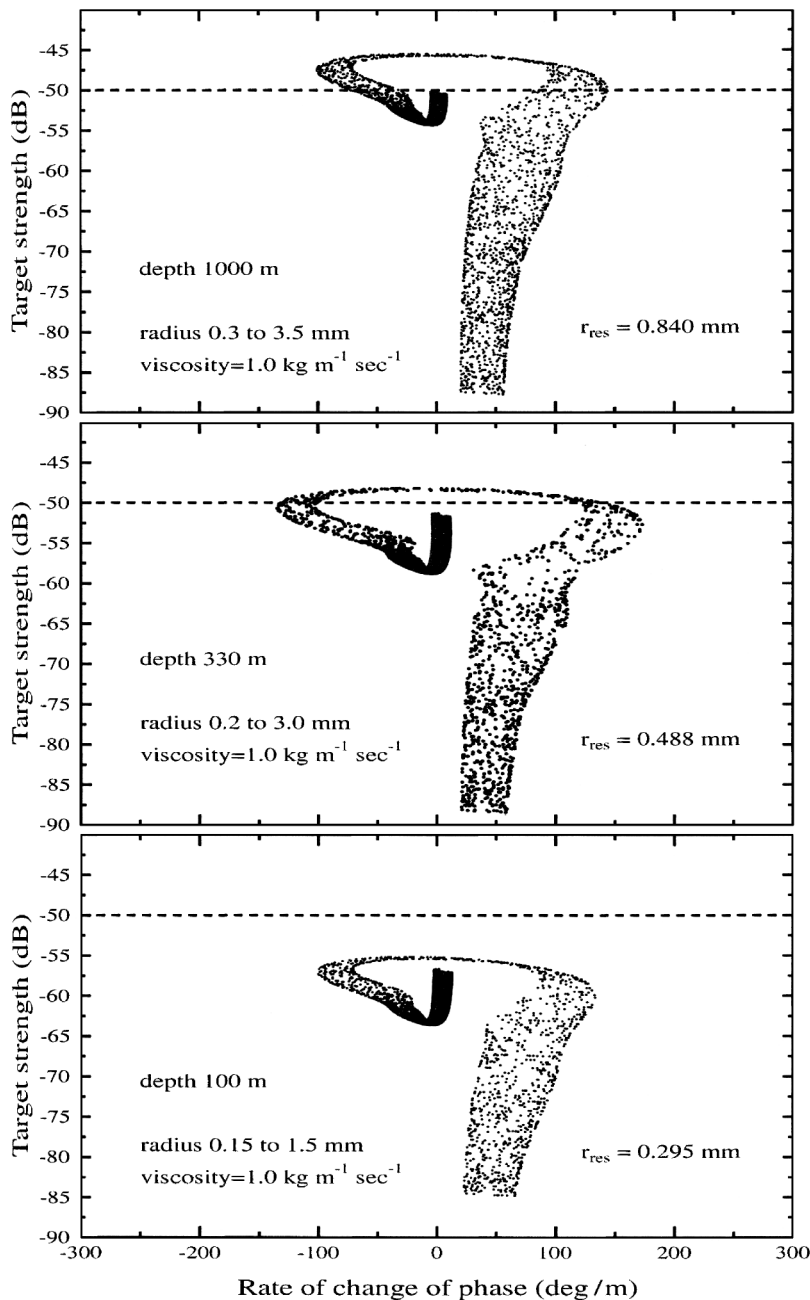
FIG. 15. Complex target space plots of 10 000 small gas-filled spherical targets with radii uniformly distributed between the limits indicated showing how resonance changes with depth. r_{res} is the radius of the gas sphere that is resonant at the given depth. The gas spheres are encased in flesh of viscosity 1.0 kg m^{-1} s^{-1}.

spheroidal (Phleger and Grigor 1990) and there are significant deposits of wax esters in the head as well as the swimbladder. Thus, to better model orange roughy scattering it will be necessary to adopt an approach such as that of Macaulay (2002), who used finite difference modeling together with density information from CAT scans to model roughy. However, while this technique will remove the computational uncertainties, it will still have practical modeling problems associated with the need to recover specimens from great depth. The uncertainties in the orange roughy modeling perhaps creates some doubt as to whether the targets exhibiting rapid rates of phase change are in fact produced by orange roughy, and the small target results showed that a similar pattern could be produced by resonating gas bubbles.

The results for myctophids were also less clear cut than for oreos. We were able to reproduce target groups **b** in Fig.

1 effectively but only by assuming a zero tilt angle showing no variation. While the former is not unreasonable, the latter is unrealistic (Bone, 1973). This is almost certainly because these targets are small relative to the wavelength and the Kirchhoff method is at the limits of its applicability. While it is frustrating not to be able to accurately model complex target strength data with a realistic fish distribution model, the lack of agreement shows the value of the extra constraints placed on the fish models by having to simultaneously fit both amplitude and phase.

Our Kirchhoff modeling of small targets (between −50 and −80 dB in Fig. 1) duplicated the range of target strengths but not the trend in phase of increasing rate of change of phase with increasing amplitude evident in Fig. 1, again demonstrating the value of the extra dimension of phase to the analysis of target strength. Modeling small fish-flesh targets as fluid spheres gave similar results to Kirchhoff

and fitted only the weakest targets (≈ -80 dB in Fig. 1). Modeling small targets as spherical gas bubbles encased in viscous fish flesh was much more successful and was able to duplicate the trend of increasing rate of change of phase with increasing target strength. This model suggests that the gas spheres responsible for groups **c** and **d** in Fig. 1 have a maximum radius of ≈ 0.6 mm and gas bubbles of this size were found in siphonophores by Barham (1963), who suggested that they were probably responsible for a deep acoustic backscattering layer observed in the San Diego trough at the time the samples were collected. The bubble sizes were measured in the laboratory three weeks after the siphonophores were removed from deep water, and it is uncertain what the size of the bubbles was at great depth in the ocean, but it was probably smaller. Even so, our results suggest that group **d** in Fig. 1 could be siphonophores and these have been caught in sediment traps at depths of 300 and 1000 m on the Chatham Rise (Nodder, 2002), where the *in situ* data were collected. The resonant gas targets in Fig. 13 could also have been produced by small myctophids with spheroidal swimbladders of equivalent size. The diameter of these was 1.5 mm, which is a reasonable swimbladder dimension for some families of small myctophids, especially those invested with lipids (Butler and Pearcy, 1972; Neighbors and Nafpaktitis, 1982).

The gas bubble model also showed a region of very rapid phase change with a comparatively small-amplitude change, resulting from the acoustic resonance of the gas bubbles (Fig. 13). This signature is very similar to the arc-shaped pattern of the rate of change of phase that we have ascribed to orange roughy above (group **a** in the lower panel of Fig. 1). Of course, these two mechanisms are not mutually exclusive and both may be present in our *in situ* data. To further investigate this question, a more sophisticated approach will be needed, such as simultaneous measurements of complex target strength at multiple frequencies [similar to those made by Kloser *et al.* (2002) using amplitude only], or a sounding system using frequency modulated (chirp) pulsing techniques (Barr, 2001). New modeling techniques such as that of Macaulay (2002) should also help. However, regardless of the mechanism involved, these signatures with rapid phase variation again clearly show the value of the extra dimension provided by the phase component of the complex target strength.

To model resonant targets at shallower depths on the Chatham rise, it was found necessary to use a value of 1.0 kg m^{-1} s^{-1} (10 Poise) for the viscosity of the medium surrounding the gas bladder. This is significantly lower than the value of 500 Poise used by Love (1993) in modeling backscatter from blue whiting and is more similar to a thick oil rather than fish flesh, possibly suggesting that the swimbladder is surrounded by lipids of this viscosity rather than fish flesh as such.

A measurement of the complex target response is useful when estimating fish biomass acoustically. Large gas-filled swimbladders and small resonant bubbles can both exhibit similar amplitude responses leading to an overestimate of the abundance of the large fish when in the presence of smaller resonant targets. However, the clear difference between the phase responses of the two types of target show that it is possible to separate them using complex target plots. It may be that biomass estimates for some species will have to be made at multiple frequencies to avoid this resonant target problem.

This work also implies that fisheries acoustic systems used in this way must be calibrated with respect to phase as well as amplitude, using, for example, the response relative to that of a tungsten carbide calibration sphere. The phase response of all system components must also be considered, for example, both the amplitude and phase response of transducers are now known to change, sometimes significantly, with the depth of immersion in the ocean (Coombs *et al.*, 2003; Kloser, 1996).

All of this work was carried out at 38 kHz and the specific results apply only at that frequency. However, the rate of change of phase is applicable generally to all fish and plankton target strength work at all frequencies.

VI. CONCLUSIONS

The target rate of change of phase adds a valuable new dimension to studies of fish and plankton target strength and to acoustic biomass estimation of fish generally. Our results show that the patterns seen in the target strength-rate of change of phase plots of *in situ* data can be explained in terms of relatively simple models of fish of different sizes and with and without gas-filled swimbladders.

Anderson, V. C. (**1950**). "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431.

Barham, E. G. (**1963**). "Siphonophores and the deep scattering layer," Science **140**, 826–828.

Barr, R. (**2001**). "A design study of an acoustic system suitable for differentiating between orange roughy and other New Zealand deep-water species," J. Acoust. Soc. Am. **109**, 164–178.

Barr, R., Coombs, R. F., and Macaulay, G. J. (**2000**). "Can we discriminate between different deepwater fishes using a standard acoustics target strength ping?," in *Proceedings of the 10th International Symposium on Acoustic Remote Sensing of the Atmosphere and Oceans and Associated Techniques (ISARS)*, Auckland, New Zealand, November **2000**, ISBN 0-473-07313-7, edited by Stuart Bradley, pp. 77–80.

Bone, Q. (**1973**). "A note on the buoyancy of some lantern-fishes (myctophoidei)," J. Mar. Biol. Assoc. U.K. **53**, 619–633.

Butler, J. L., and Pearcy, W. G. (**1972**). "Swimbladder morphology and specific gravity of myctophids off Oregon," J. Fish. Res. Board Can. **29**, 2168–2179.

Clay, C. S. (**1991**). "Low-resolution acoustic scattering models: Fluid-filled cylinders and fish with swimbladders," J. Acoust. Soc. Am. **89**, 164–178.

Clay, S. C., Horne, J. K. (**1994**). "Acoustic models of fish: The Atlantic Cod (Gadus morhua)," J. Acoust. Soc. Am. **96**, 1661–1668.

Coombs, R. F., and Barr, R. (**2004**). "Acoustic remote sensing of swimbladder orientation and species mix in the oreo population on the Chatham Rise," J. Acoust. Soc. Am. **115**, 1516–1524.

Coombs, R. F., Macaulay, G. J., Knol, W. and Porritt, G. (**2003**). "Configurations and calibrations of 38 kHz fishery acoustic survey systems, 1991–2000," N. Z. Fish. Assess. Rep. **2003/49**, p. 24.

Do, M. A. and Coombs, R. F. (**1989**). "Acoustic measurements of the population of orange roughy (*Hoplostethus atlanticus*) on the north Chatham Rise, New Zealand," N.Z.J. Mar. Freshwater Res. **23**, 225–237.

Doonan, I., Bull, B., Dunford, A., Coombs, R., Tracey, D., and Hart, A. (**2001**). "Acoustic estimates of the biomass of aggregations of orange roughy in the spawning box and on the northeastern and easter hills, Chatham Rise, July 2000," N. Z. Fish. Assess. Rep. **2001/70**, p. 24.

Ehrenberg, J. E. (**1979**). "A comparative analysis of *in situ* methods for directly measuring the acoustic target strength of individual fish," IEEE J. Ocean. Eng. **OE-4**, 141–152.

Feuillade, C., and Werby, M. F. (**1994**). "Resonances of deformed gas bubbles in liquids," J. Acoust. Soc. Am. **96**, 3684–3692.

Foote, K. G. (**1980**). "Importance of the swimbladder in acoustic scattering by fish: A comparison of gadoid and mackerel target strengths," J. Acoust. Soc. Am. **67**, 2084–2089.

Foote, K. G., and Francis, D. T. I. (**2002**). "Comparing Kirchhoff-approximation and boundary-element models for computing gadoid target strength," J. Acoust. Soc. Am. **111**, 1644–1654.

Furusawa, M. (**1988**). "Prolate spheroidal models for predicting general trends of fish target strength," J. Acoust. Soc. Jpn. (E) **9**, 13–23.

Kloser, R. J. (**1996**). "Improved precision of acoustic surveys of benthopelagic fish by means of a deep-towed transducer," ICES J. Mar. Sci. **53**, 407–413.

Kloser, R. J., Ryan, T., Sakov, P., Williams, A., and Koslow, J. A. (**2002**). "Species identification in deep water using multiple acoustic frequencies," Can. J. Fish. Aquat. Sci. **59**, 1065–1077.

LeFeuvre, P., Rose, G. A., Gosine, R., Hale, R., Pearson, W., and Kahn, R. (**2000**). "Acoustic species identification in the Northwest Atlantic using digital image processing," Fisheries Res. **47**, 137–147.

Love, R. H. (**1978**). "Resonant acoustic scattering by swimbladder-bearing fish," J. Acoust. Soc. Am. **64**, 571–580.

Love, R. H. (**1993**). "A comparison of volume scattering strength data with model calculations based on quasisynoptically collected fishery data," J. Acoust. Soc. Am. **94**, 2255–2268.

Macaulay, G. J. (**2002**). "Anatomically detailed acoustic scattering models of fish," Bioacoustics **13**, 275–277.

MacLennan, D. N. (**1981**). "The theory of solid spheres as sonar calibration targets," Scott. Fish. Res. Rep. **22**, 17.

Marshall, N. B. (**1960**). "Swimbladder structure of deep-sea fishes in relation to their systematics and biology," *Discovery Reports Vol. XXXI* (Cambridge University Press, Cambridge), pp. 1–22.

McClatchie, S., Macaulay, G., Coombs, R. F., Grimes, P., and Hart, A. (**1999**). "Target strength of an oily deep-water fish, orange roughy (*Hoplostethus atlanticus*). I. Experiments," J. Acoust. Soc. Am. **106**, 1–12.

McClatchie, S. (**2002**). (private communication).

Medwin, H., and Clay, S. C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, London).

Neighbors, M. A. and Nafpaktitis, B. G. (**1982**). "Lipid compositions, water contents, swimbladder morphologies and buoyancies of nineteen species of midwater fishes (18 myctophids and 1 neoscopelid)," Mar. Biol. (Berlin) **66**, 207–215.

Nodder, S. (**2002**). (private communication).

Paul, L. J., (**1986**). *New Zealand Fishes: An Identification Guide* (Reed Methuen, Auckland, New Zealand).

Phleger, C. F., and Grigor, M. R. (**1990**). "Role of wax esters in determining buoyancy in *Hoplostethus atlanticus* (Beryciformes: Trachichthyidae)," Mar. Biol. (Berlin) **105**, 229–233.

Rose, G. A., and Leggett, W. C. (**1988**). "Hydroacoustic signal classification of fish schools by species," Can. J. Fish. Aquat. Sci. **45**, 597–604.

Scalabrin, C., Diner, N., Weill, A., Hillion, A., and Mouchot, M. (**1996**). "Narrowband acoustic identification of monospecific fish shoals," ICES J. Mar. Sci. **53**, 181–188.

Soule, M. Barange, M., and Hampton, I. (**1995**). "Evidence of bias in estimates of target strength obtained with a split-beam echosounder," ICES J. Mar. Sci. **52**, 139–144.

Soule, M. Barange, M., Solli, H., and Hampton, I. (**1997**). "Performance of a new phase algorithm for discriminating single and overlapping echoes in a split-beam echosounder," ICES J. Mar. Sci. **54**, 934–938.

Soule, M., Hampton, I., and Barange, M. (**1996**). "Potential improvements to current methods of recognizing single targets with a split-beam echosounder," ICES J. Mar. Sci. **53**, 237–243.

Strasberg, M. (**1953**). "The pulsation frequency of nonspherical gas bubbles in liquids," J. Acoust. Soc. Am. **25**, 536–537.

Ye, Z. (**1997**). "Low-frequency acoustic scattering by gas-filled prolate spheroids in liquids," J. Acoust. Soc. Am. **101**, 1945–1952.

Statistical description of matched field processing ambiguity surfaces

Brian H. Tracey
MIT Lincoln Laboratory, 244 Wood Street, Lexington, Massachusetts 02420

The relationship between array design and ambiguity surface characteristics is not straightforward for matched field processing (MFP). Because MFP makes use of multipath propagation, ambiguities are a function of the environment as well as the array. This paper develops a statistical approach that seeks to provide an analytic link between array design and metrics describing the MFP output. Approximate expressions are derived for the probability distribution of power output across conventional MFP ambiguity surfaces. The validity of the expressions is examined through numerical simulation. This approach can be used as a design tool for comparing the expected performance of different array geometries. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2000750]

I. INTRODUCTION

For plane-wave and range-focused beamforming, the relationship between array geometry and the array response is well understood.[1] Quantities such as array resolution and sidelobe levels can be calculated in a straightforward manner. A set of similar guidelines linking the array geometry to array response does not currently exist for matched field processing (MFP). MFP uses full-field propagation models to account for multipath when constructing steering vectors for beamforming.[2] A major difficulty in deriving guidelines is that the array response depends on the environment for MFP. Thus array design for MFP is often studied through numerical simulations. While these studies can be of significant value, analytic tools for array design are also desirable.

Since MFP seeks to exploit multipath, an ideal receive array would provide orthogonal sampling of the propagating normal modes. For a vertical line array (VLA), this means that the array should span as much of the water column as possible, with depth sampling sufficient for sampling the highest trapped mode.[3,4] Horizontal line arrays (HLAs) rely on mode orthogonality in the range dimension to separate the propagating modes, leading to requirements on array length.[5,6] The mode correlation matrix (MCM), which describes cross-talk between modes, provides a useful way to characterize an array's ability to resolve multipath. HLA and VLA geometries can be directly compared using metrics that describe the extent to which the MCM is diagonally dominant.[7,8]

MFP ambiguity surfaces are characterized by a complicated set of local maxima and minima that shift with source location and environmental parameters. As a result, they have often been described by statistical measures such as peak-to-average sidelobe ratio and deflection index.[9] The approach taken in several previous studies of MFP array design[7,8] is to simulate MFP output for different arrays, calculate statistics describing the ambiguity surfaces, and correlate these statistical measures with metrics describing the diagonality of the mode correlation matrix. In contrast, the work below seeks to develop an *analytic* link between mode resolution and a statistical description of the MFP ambiguity surface. The analytic approach can help give additional insight into the array design problem. The results below show how ambiguity surface statistics depend on the mode shapes and attenuations across the ranges and depth searched, in addition to the mode correlation matrix.

The first two sections below show the MFP ambiguity structure for several example arrays and present an expression relating the conventional MFP (CMFP) output to the mode correlation matrix. The following sections derive the predicted sidelobe distribution for narrow-band and broadband CMFP. Monte Carlo simulations are then used to explore the accuracy of the predictions as a function of array geometry. Criteria are developed to identify when good quantitative agreement can be expected. Examples below also suggest that the analytic approach can provide better qualitative ranking of array performance than commonly used metrics such as MCM diagonality.

II. DEPENDENCE ON ARRAY GEOMETRY

As motivation, the effects of array geometry on MFP are examined for a representative shallow water environment. Simulations were generated based on environmental data gathered for the Santa Barbara channel experiment (SBCX).[10] This environment consists of a 209-m-deep water column with a downward refracting sound speed profile, as shown in Fig. 1. The seabed in the SBCX site was described as a series of sediment layers, with properties given in Table I. The KRAKEN normal mode model[11] was used for the simulations, with the environment assumed to be range independent.

Simulations were carried out for four array geometries. Properties of the arrays are listed in Table II. The first array geometry was a vertical line array (VLA) spanning 80% of the water column. This array span is similar to those actually used during the SBCX data collection. A second VLA was modeled that spanned only 40% of the water column. Two
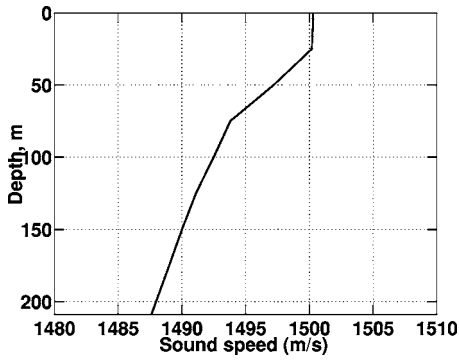
FIG. 1. Sound speed profile used for simulations, taken from SBCX measurements.

TABLE II. Parameters for arrays used in simulations. For vertical arrays, "depth" refers to the bottommost hydrophone in the array.

Type	No. of sensors	Spacing (m)	Length (m)	Depth (m)
80% VLA	50	3.4	166.6	208
40% VLA	50	1.7	83.3	208
1 km HLA	201	5	500	100
500 m HLA	101	5	1000	100

horizontal line arrays (HLAs), of 1 km and 500 m length respectively, were modeled. Both HLAs were oriented at endfire to the source, as multipath resolution and MFP performance are best for HLAs at this orientation. Narrow-band simulations were done for a frequency of 120 Hz. This frequency is in the range of projected tones used in SBCX and other MFP experiments.

Conventional, or Bartlett, MFP surfaces were generated to demonstrate the localization possible with each array. No environmental mismatch was introduced. The signal was normalized so that the peak of the ambiguity surface equaled one. The conventional MFP output power is given by

$$b(\Theta) = \left| \frac{\mathbf{p}_{\text{look}}^H(\Theta)}{|\mathbf{p}_{\text{look}}(\Theta)|} \cdot \frac{\mathbf{x}(\Theta_S)}{|\mathbf{x}(\Theta_S)|} \right|^2, \quad (1)$$

where $\Theta = (r, z, \theta)$ is the MFP steering direction in three dimensions, Θ_S is the source location, $\mathbf{x}$ is the data, and $\mathbf{p}_{\text{look}}$ is the predicted pressure field on the array. The vector $\mathbf{p}_{\text{look}}$ and its normalization are kept separate, rather than being combined into a replica vector $\mathbf{v}$, for use in calculations below. In the absence of mismatch, the data vector $\mathbf{x}$ is a scaled version of the replica vector at the source location.

Simulated CMFP surfaces are shown in Fig. 2. The simulated noise source was located at 70-m depth and at 10-km range from the center of the receiver array. The 80% spanning VLA provides a good localization of the source. Significant ambiguities exist for the other arrays.

While the two HLAs clearly have high sidelobes, two points are worth noting. First, the horizontal aperture provides azimuthal discrimination that is not available with the VLAs. If only a single source is present along each bearing, it may be possible to track the MFP output over time to

TABLE I. Parameters for geoacoustic model of SBCX test site: z=depth from surface; c_c=compressional sound speed; ρ=density; α_c =compressional wave attenuation.

z (m)	c_c (m/s)	ρ (g/cm^3)	α_c (dB/λ)
209	1607	1.95	0.37
309	1702	1.95	0.37
309	1862	1.98	0.035
609	2374	1.98	0.035
609	2374	2.03	0.04

localize the source even though an instantaneous MFP output is highly ambiguous. Adaptive beamforming and broadband processing can also help to reduce ambiguities.

The HLA simulations also show that sidelobe levels are lower near the surface than at depth. This results from a combination of two effects. First, the HLAs can better resolve the high-angle modes associated with surface sources, as these arrive at conical angles closer to broadside. Second, the downward-refracting sound speed profile traps the lowest-order modes at depth.

III. MODE CORRELATION MATRIX AND MFP OUTPUT

In this section the mode correlation matrix is defined and its link to the output of a Bartlett processor is shown. The acoustic source is assumed to be in the far field of an array of L sensors. The range r_i between each sensor and the source is given by

$$r_i = r_s + \Delta_i, \quad (2)$$

where r_s is the range from the source to the center of the array and Δ_i is the range increment to each sensor. For a perfectly straight VLA, $\Delta_i = 0$ for all sensors. For a uniform straight HLA, $\Delta_i(\theta) = (x_i - x_{\text{ref}})\cos \theta$, where x is distance along the array and θ is bearing to the source. Using a normal mode description of the sound field, the pressure received on sensor i is given by

$$p(r_i, z_i, \theta) = \frac{i\sqrt{2\pi} e^{-i\pi/4}}{\sqrt{r_s}} \sum_n \frac{\psi_n(z_s)\psi_n(z_i)}{\sqrt{k_n}} e^{ik_n[r_s + \Delta_i(\theta)]}, \quad (3)$$

where ψ_n is the mode shape for mode n, k_n is the horizontal wave number for the mode, and z_s and z_i are the source depth and ith receiver depth, respectively. For matched field processing, replica vectors describing the response of the array to every possible source location are generated. Assuming that no mismatch exists, the replicas may be generated by using Eq. (3), evaluated for all look directions (r, z, θ).

For the far-field case, complex modal correlation matrices can be defined that include the range differences between elements, but not the absolute range:

$$e_{nm} = \sum_{i=1}^{L} \psi_n(z_i)\psi_m(z_i) e^{i(k_n\Delta_i(\theta_s) - k_m^*\Delta_i(\theta))}. \quad (4)$$

Since the modes are orthogonal, the mode correlation matrix will be diagonal for a fully spanning VLA.

This definition can be used to rewrite the CMFP output power given by Eq. (1). First, Eq. (3) is used to express the data vector $\mathbf{x}$, neglecting mismatch [mismatch can be in-

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Brian H. Tracey: Statistics of matched field ambiguity surfaces   1373
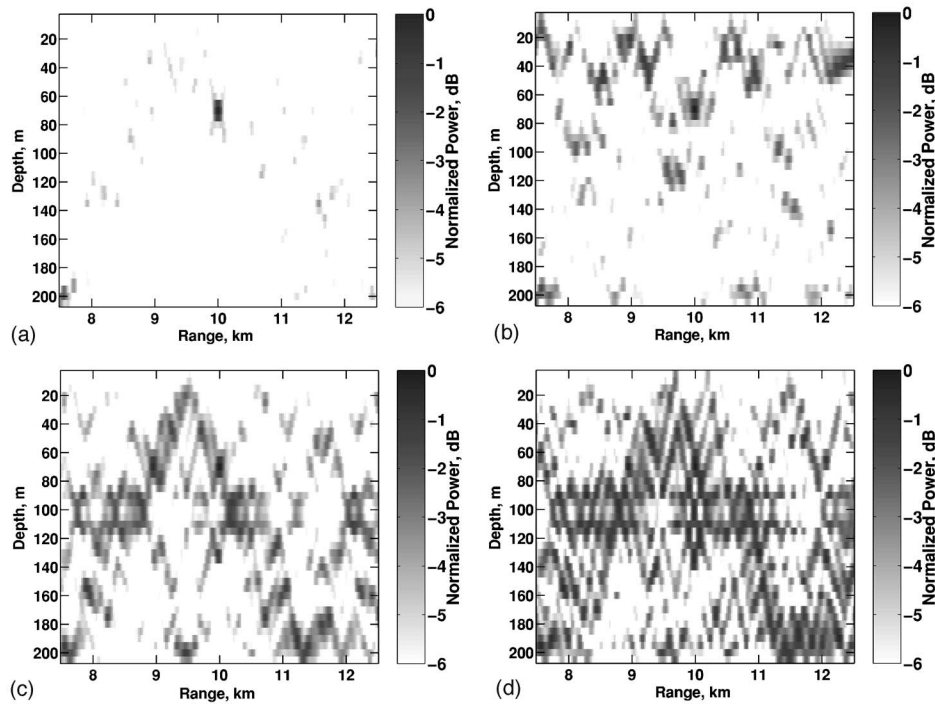
FIG. 2. Simulated CMFP ambiguity surfaces in the SBCX environment for the arrays considered. A 120-Hz source is located at 10-km range and 70-m depth, oriented at endfire to the horizontal arrays. Although HLAs have much higher sidelobes, they also have an ability to separate sources based on bearing that VLAs do not.

cluded by perturbing the modal horizontal wavenumbers in Eq. (3)]. All terms involving the array geometry are combined using the mode correlation in Eq. (4). The remaining terms may be grouped together, giving the output as

$$b(\Theta) = \left| \sum_{n,m} Y_{nm} e_{nm} \right|^2, \quad (5)$$

where Y_{nm} contains all information about the source location and look direction:

$$Y_{nm} = \frac{2\pi}{\sqrt{k_n k_m}} \frac{\psi_n(z_s)\psi_m(z)e^{i(k_n r_s - k_m^* r)}}{|\mathbf{p}_{\text{look}}(\Theta)||\mathbf{x}(\Theta_S)|}. \quad (6)$$

The conjugates on mode eigenvalues ensure that attenuation is handled correctly.

Before proceeding, it is helpful to study the mode correlation matrix in more detail. Figure 3 shows mode correlation matrices for the arrays defined in Sec. II. The plots are normalized so diagonal elements are equal to one. Modes 15 and higher have high loss and do not contribute to the acoustic field at long range. As expected, the 80% spanning VLA provides excellent resolution of the modal arrivals, while increased cross-talk is seen for the 40% VLA and the two HLAs.

A basic difference in the mode correlation patterns for vertical and horizontal arrays is seen in Fig. 3. For VLAs, the cross-talk between modes remains roughly similar as mode number increases, while for HLAs cross-talk decreases with mode order. The improved resolution of HLAs for higher modes results because higher modes impinge on the array at steeper vertical angles. These modes therefore arrive at coni-

cal angles that are closer to broadside, where the array resolution is better, and can be more easily separated.

While Fig. 3 shows results for endfire, Eq. (4) applies for any bearing angle. As the look direction bearing and source bearing diverge, the values of all components in the mode correlation matrix will drop. This reflects the filtering of source energy introduced by the conventional beampattern of the mode filter. Thus Eq. (4) and the results in the following sections can be used to determine the output power statistics as a function of the azimuthal difference from the source.

IV. NARROW-BAND CMFP STATISTICS

Equation (5) above expressed CMFP output power in terms of cross-modal correlations. That relationship is used here to derive approximate analytical results for the probability distribution of matched-field sidelobes. A statistical approach is based on the following assumptions.

(i)   Statistics for the MFP surface are assembled by randomly sampling different MFP cells or "look directions." Thus the look direction can be treated as a random variable.

(ii)   The true source location is unknown, so the source depth and range are assumed to be random variables.

The look direction depths are assumed to be distributed with uniform probability over depths $[D_{l1}, D_{l2}]$, while the source depths are distributed over $[D_{s1}, D_{s2}]$. Both source ranges and look direction ranges are assumed to be uniformly distributed over the full MFP search space, from R to $R+\Delta R$.
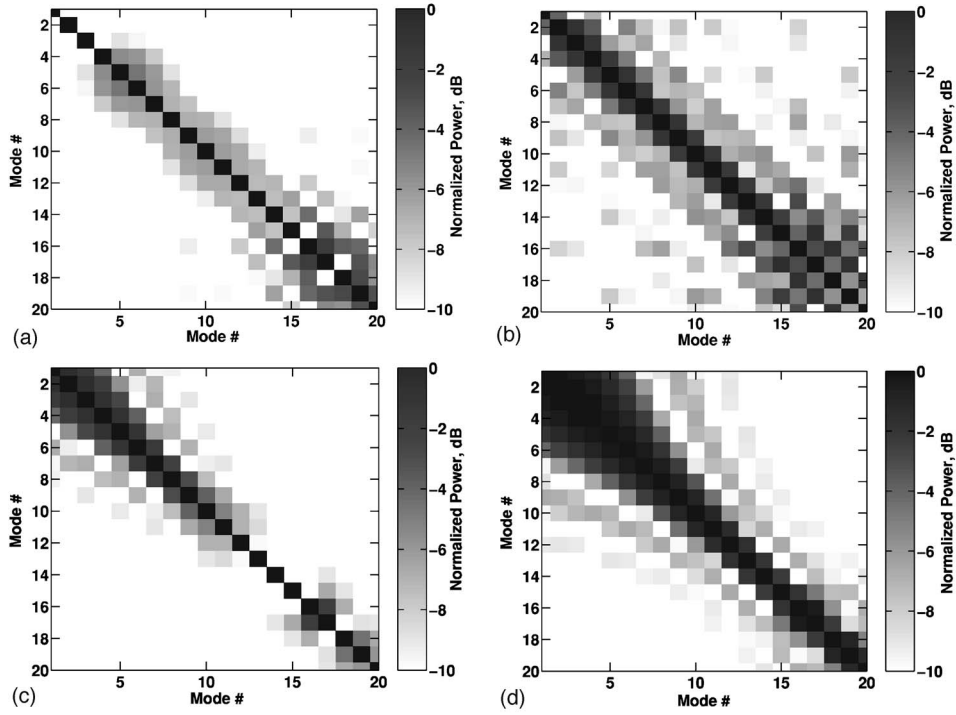
FIG. 3. Far-field mode correlation matrices for the arrays considered. Note the difference in structure between horizontal and vertical arrays.

Given the assumptions above, the terms Y_{nm} from Eq. (5) are zero-mean complex random variables (note that the mode correlation terms e_{nm}, which depend on array geometry, are considered deterministic). A further assumption is made that the random terms Y_{nm} are Gaussian distributed. Because many such terms are summed, inaccuracies in this assumption are reduced by the central limit theorem. The Y_{nm} are zero-mean, so only their variance needs to be calculated. The variance of the real part of Y_{nm} is found as (see the Appendix)

$$\sigma_{nm}^2 = \frac{2\pi^2}{k_n k_m} \int \frac{\psi_m^2(z)\psi_n^2(z_s)}{|\mathbf{p}_{\text{look}}(\Theta)|^2 |\mathbf{x}(\Theta_S)|^2} \times e^{-2\alpha_n r_s} e^{-2\alpha_m r} p_{\mathbf{r}}(\mathbf{r}) \, d\mathbf{r}, \tag{7}$$

where $\mathbf{r}$ represents the independent random variables r_s, z_s, r, and z. A further assumption is required to represent the normalizing terms $|\mathbf{p}_{\text{look}}|^2$ and $|\mathbf{x}|^2$. For an array that resolves the modes well, off-diagonals will not contribute strongly to the intensity, and the norm-squared can be approximated by

$$|\mathbf{p}(\Theta)|^2 \approx \sum_{l=1}^{N_p} \frac{2\pi}{k_l} \psi_l^2(z) e_{ll} e^{-2\alpha_l r_{\text{mid}}}, \tag{8}$$

where N_p is the number of trapped (or proper) modes and r_{mid} is the mid-range of the MFP search region whose statistics are being calculated. This approximate normalization is discussed in more detail in the Appendix.

Numerical results below indicate that the assumption of a Gaussian distribution is reasonable. The approximation made in Eq. (8) appears to be a cause of inaccuracy for

arrays with limited vertical aperture. Unfortunately, a more accurate (but analytically tractable) approximation has not been identified.

The Appendix shows that the assumptions above result in the following expression:

$$\sigma_{mn}^2 \approx \frac{4\pi^2}{2k_n k_m} Z_m(D_{l1}, D_{l2}) Z_n(D_{s1}, D_{s2}) \times R_m(R, \Delta R) R_n(R, \Delta R), \tag{9}$$

where Z_m and R_m account for depth and range integrations. The range integral R_m accounts for mode attenuation over the range of the ambiguity surface:

$$R_m(R, \Delta R) = \frac{e^{-2\alpha_m R}}{2\alpha_m \Delta R} [e^{-2\alpha_m \Delta R} - 1], \tag{10}$$

while the depth integrals are given by

$$Z_m(D_1, D_2) = \int_{D_1}^{D_2} \frac{\psi_m^2(z) \, dz}{\Delta D \sum_{l=1}^{N_p} (2\pi/k_l) \psi_l^2(z) e_{ll} e^{-2\alpha_l r_{\text{mid}}}}, \tag{11}$$

where D_1 and D_2 are the lower and upper depths of integration, and $\Delta D = D_2 - D_1$. Separate integrals over depth are required for the source and replicas, as the region of possible source depths can be different than the MFP search region.

Equation (9) gives the variance for the real part of a single Y_{nm} term from Eq. (5). These terms are multiplied by e_{nm} and summed to give another complex zero-mean Gaussian random variable, denoted Y. The variances of the real and imaginary parts of Y are equal and are given by

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Brian H. Tracey: Statistics of matched field ambiguity surfaces   1375

$$\sigma_Y^2 = \sum_{n,m} \sigma_{nm}^2 |e_{nm}|^2. \tag{12}$$

Finally, the absolute value-squared is taken as shown in Eq. (5), yielding an exponential probability distribution:

$$P_{NB}(b) = \frac{1}{\mu} e^{-b/\mu}, \tag{13}$$

where

$$\mu = 2 \sum_{n,m} \sigma_{nm}^2 |e_{nm}|^2. \tag{14}$$

The interpretation of this result is that the ambiguity surface distribution is controlled by the amount of mode cross-talk, as measured by e_{nm}, weighted by a factor σ_{mn}^2 that captures the mode excitation and attenuation.

Equation (13) is an approximate result and can give non-physical behavior (i.e., nonzero probability of $b > 1$). A metric that describes its range of validity is therefore required. One useful approach is to require that the probability of non-physical behavior is below a threshold; for example $\Pr(b > 1) < 0.01$. This metric is demonstrated in the numerical results below.

Several results can be found directly from Eq. (13). The percentage of the ambiguity surface that has a value higher than some threshold γ is given by

$$\Pr(b > \gamma) = e^{-\gamma/\mu}. \tag{15}$$

For example, the fraction of the surface within 3 dB of the peak is found by setting $\gamma = 0.5$. Finally, the expected value of the exponential distribution is μ. This means that the average peak-to-sidelobe ratio, which has been used as a statistic for describing MFP ambiguity surfaces, is given by μ.

V. STATISTICS FOR INCOHERENT BROADBAND CMFP

When the source being localized has broadband energy, frequency averaging may be used to reduce ambiguities in the output. Frequency averaging is beneficial because the location of MFP ambiguities shifts over frequency while the true source location remains fixed. Thus high ambiguities at one frequency are averaged together with lower output at other frequencies. Linear averaging has been widely applied to MFP:[9,12]

$$I_{BB} = \frac{1}{M} \sum_{i=1}^{M} b(f_i). \tag{16}$$

Other options include averaging cross-frequency components[13] or averaging matched field outputs in decibels rather than in linear power.[14] Only linear averaging is analyzed here.

The results in Sec. IV can be used to find the PDF for broadband CMFP. In doing so, the assumption is made that the frequencies being averaged are spaced widely enough that their sidelobe locations can be treated as uncorrelated. Conditions for this to be true are discussed below. If more closely spaced frequencies are added, little benefit in terms

of sidelobe reduction will result. The PDF for the sum of M uncorrelated ambiguity surfaces is found using moment generating functions:

$$P(b) = \mathcal{L}^{-1} \left(\prod_{n=1}^{M} \mathcal{L}(P_{NB}(b|f_n)) \right), \tag{17}$$

where $\mathcal{L}$ denotes the Laplace transform. Using the transformed narrow-band distribution $\mathcal{L}(e^{-b/\mu}/\mu) = (1/\mu)[1/(s + 1/\mu)]$ along with the definition of the inverse transform, the PDF for the averaged surface is found as

$$P_{BB}(b) = \frac{M}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \left(\prod_{n=1}^{M} \frac{1}{\mu_n} \right) \left(\prod_{n=1}^{M} \frac{1}{s + 1/\mu_n} \right) e^{sbM} \, ds, \tag{18}$$

where the factors of M in the leading fraction and the exponent account for the number of frequencies being averaged. The narrow-band terms μ_n will vary with frequency and will be unique. The integral above can then be solved as a contour integration with each frequency giving rise to a simple pole in the complex plane. Summing the residues of these poles gives the final result:

$$P_{BB}(b) = M \left(\prod_{n=1}^{M} \frac{1}{\mu_n} \right) \sum_{n=1}^{M} \left(\prod_{m=1, m\neq n}^{M} \frac{\mu_m \mu_n}{\mu_n - \mu_m} \right) e^{-bM/\mu_m}. \tag{19}$$

Although the narrow-band coefficients μ_n are unique, they often vary slowly with frequency. This can give rise to numerical instabilities. If the μ_n are approximated as constant, the broadband distribution is a central χ^2 distribution with $2M$ degrees of freedom:

$$P_{BB}(b) = \frac{M}{\sigma_Y^{2M} 2^M \Gamma(M)} (bM)^{M-1} e^{-bM/2\sigma_Y^2}, \tag{20}$$

or, in terms of μ,

$$P_{BB}(b) = \frac{M}{\mu^M \Gamma(M)} (bM)^{M-1} e^{-bM/\mu}. \tag{21}$$

The higher-order χ^2 distributions are characterized by a reduced probability of either very large or very small values. Frequency averaging fills in the troughs of the sidelobes in addition to averaging down the peaks. The *average* peak to sidelobe ratio for this form of incoherent averaging is μ, the same as for narrow-band processing, but the shape of the distribution is different.

The derivation above assumes that the sidelobe locations are uncorrelated across frequency. Waveguide invariant arguments can be used to estimate the frequency spacing required for this to be true.[12,15] The migration of CMFP sidelobes with frequency is described by

$$\frac{\delta\omega}{\delta r} = \beta \frac{\omega}{r - R}, \tag{22}$$

where R is the true range to the source and β is the waveguide invariant. β is roughly 1 for many shallow water scenarios. For the CMFP sidelobes to shift by one CMFP range main lobe L_r, a frequency shift of

Brian H. Tracey: Statistics of matched field ambiguity surfaces
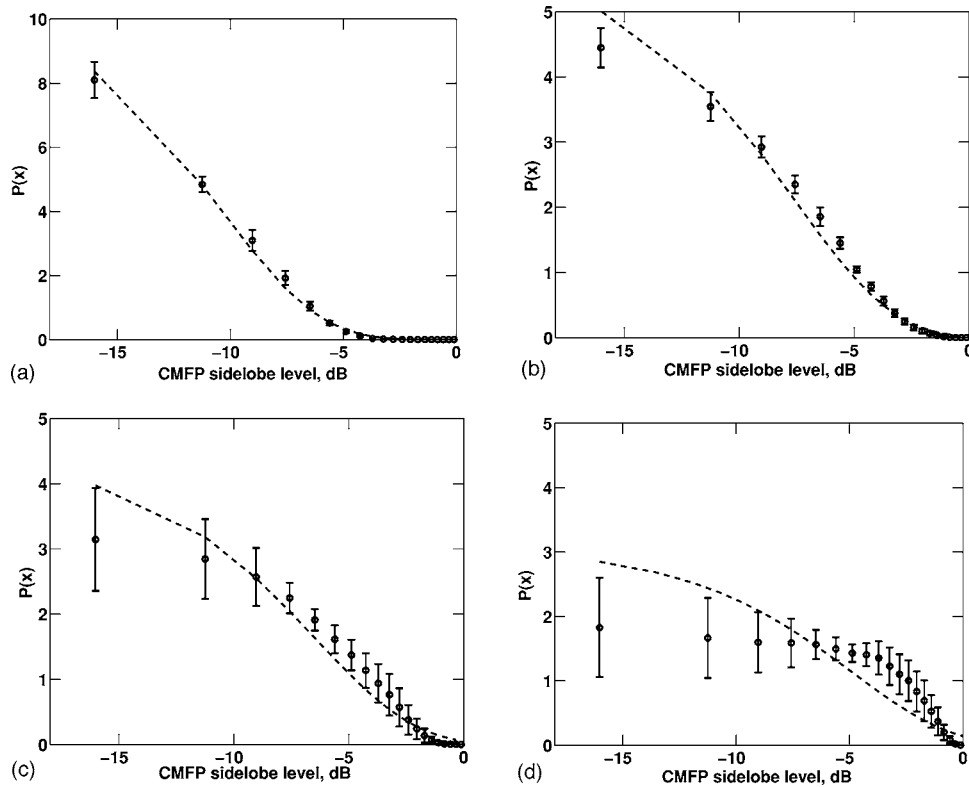
FIG. 4. Comparison of predicted PDFs (dashed lines) with distributions estimated from Monte Carlo runs. Histograms were generated for individual Monte Carlo runs and used to produce the averaged distribution ± one standard deviation. Agreement is close for the 80% spanning VLA and degrades as effective vertical aperture is reduced.

$$\Delta f \geqslant \frac{fL_r}{r - R} \qquad (23)$$

is required. The main lobe size in range is given by $L_r = 2\pi/(k_1 - k_N)$.[10] Since sidelobes shift faster with increasing distance from the source, frequency averaging reduces MFP levels away from the source more rapidly. Setting $r - R = 1$ km in Eq. (23) suggests, for the scenarios considered below, a frequency spacing of roughly 10 Hz. Adding frequencies at a finer spacing is not expected to help reduce ambiguities in the MFP output.

VI. NUMERICAL RESULTS

Simulations were carried out to test the results above. CMFP power was simulated for an ensemble of random source ranges and depths. The output power from these surfaces was used to generate histograms which were normalized to estimate PDFs. A mean distribution and standard deviations around it were calculated and are compared to the predicted distribution.

Figure 4 shows a comparison of simulated and predicted distributions for narrow-band CMFP. For these cases, the source and look direction depths and ranges were allowed to vary uniformly over the MFP search grid shown in Fig. 2. For the 80% spanning VLA the agreement to the overall shape is quite good. Of the arrays considered, this one best satisfies the assumption used to derive the approximate normalization in Eq. (8).

The match between predictions and simulation degrades for the less capable arrays. As discussed above, one possible

metric for the accuracy of Eq. (13) is the probability of non-physical outputs $[\Pr(b > 1)]$. This probability is calculated to be 1.1e-5 for the 80% spanning VLA, 0.003 for the 40% spanning VLAs, 0.0094 for the 1 km HLA, and 0.046 for the 500 m HLA. Reasonable matches to the analytic model are seen for cases where $\Pr(b > 1) < 0.01$. It is worth noting that the relevant parameter for mode cross-talk is the array length in wavelengths. Thus the agreement for the 500-m array at 240 Hz should be comparable to the agreement seen for the 1-km-long array at 120 Hz.

The simulations in Fig. 2 showed that the HLA apertures could provide some discrimination between surface and submerged depth zones. The ability of the analytic results to predict this is tested by restricting the look directions D_{l1} and D_{l2} in the calculation. A comparison between calculated and simulated results is shown in Fig. 5 for the 1-km-long HLA example. The "surface" depth zone is defined as 0–20 m, with the "submerged" zone defined as the rest of the water column. The analytic prediction is successful in predicting the surface/submerged trends in output power distribution.

Simulations are next used to examine the output distribution for incoherent broadband MFP. The analysis above showed that a simplification is possible if the narrow-band variances are fairly flat with frequency. These variances were examined and found to vary by less than 10% over the frequency range considered. cases considered. Thus Eq. (21), which assumes constant variance with frequency, was used for the calculation.

Figure 6 shows a comparison of the predicted and simulated distributions for incoherent broadband MFP. The calcu-
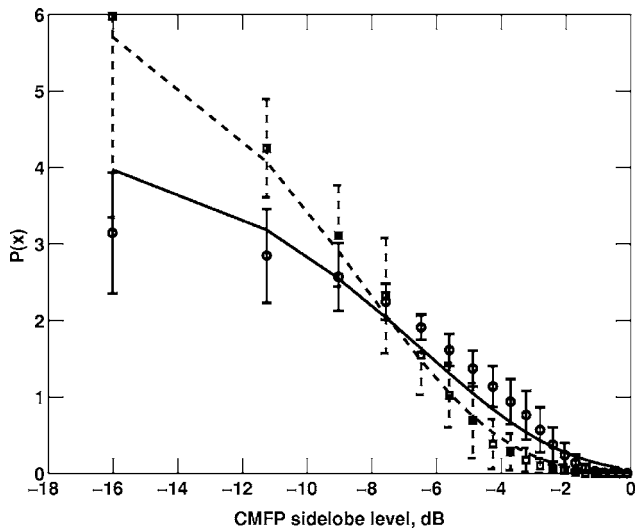
FIG. 5. Predicted PDF for depth zones [1–20 m] and [20–209 m] compared to Monte Carlo results, 1 km HLA. Predictions are shown with lines, while Monte Carlo results are shown with error bars. Near-surface zone results are shown as dashed lines, while submerged zone results are solid. Note improved agreement between the model and Monte Carlo for the near-surface zone due to improved sampling of the high-order modes.
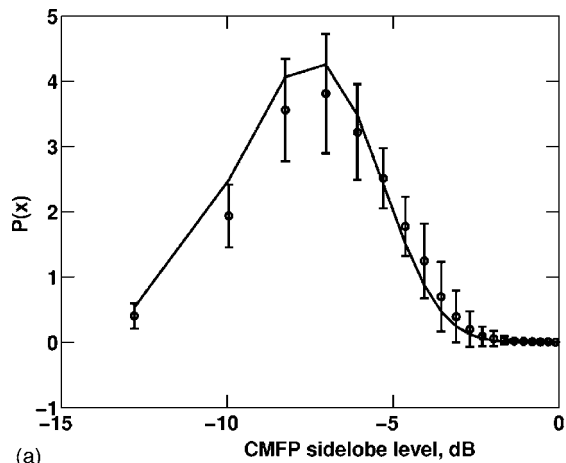
(a)
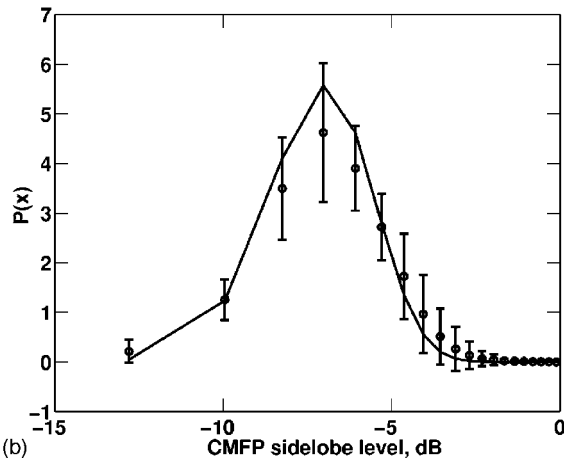
(b)

FIG. 6. Comparison of predicted PDFs for incoherent broadband averaging with distribution estimated from Monte Carlo runs. Predictions are shown with dashed lines, while Monte Carlo results are shown with error bars.
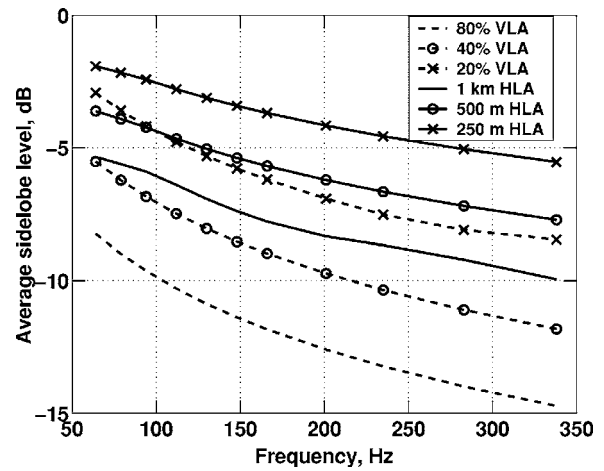
FIG. 7. Average sidelobe level (μ parameter) for narrow-band CMFP versus frequency and array geometry. The improved resolution and increased number of modes at high frequencies act to reduce average sidelobe levels. For HLAs, the sidelobe levels calculated are for the source bearing (endfire) only; lower sidelobes are present on other bearings.

lations were done for the 1-km-long HLA for bandwidths of 20 and 40 Hz. As the bandwidth increases, the distribution becomes more peaked; both high sidelobes and deep nulls in the frequency-averaged output become less likely. It is interesting to note that the agreement between the predicted and simulated distributions is closer than it was in the narrow-band case. This is presumably because variations in the steering vector normalization average out over frequency, making the approximation in Eq. (8) more accurate.

VII. SIDELOBE LEVELS VERSUS APERTURE AND FREQUENCY

The numerical results above have helped to validate the analytic results derived. In this section the analytic results are used to compare performance of several arrays versus frequency. The derived mean sidelobe level μ is also compared to a metric measuring MCM diagonality for use as a predictor of array performance.

Figure 7 shows the average sidelobe ratio calculated for single-frequency CMFP. This ratio is given by the parameter μ from Eq. (13), assuming that source power is normalized to one. The ratio is calculated for a set of tones in the range 64–338 Hz that correspond to projected tones used in SBCX. For HLAs, the results shown are for the endfire beam only. Results are shown for the four arrays discussed above and for two shorter arrays: a 250-m-long HLA (101 phones at 2.5-m spacing) and a 20% spanning bottom-mounted VLA. A trend is seen toward lower sidelobe levels at higher frequencies, where more modes can be coherently summed to reduce ambiguities.

As discussed above, a useful criteria for quantitative accuracy of the predictions is that the probability of nonphysical results be less than some threshold. Discussion following Fig. 4 suggested a criteria of $\Pr(b > 1) < 0.01$, which corresponds to an average sidelobe level of −6.6 dB or less. A less restrictive criteria of $\Pr(b > 1) < 0.05$ would suggest reasonable agreement for arrays with average sidelobe level of −4.76 dB or lower.
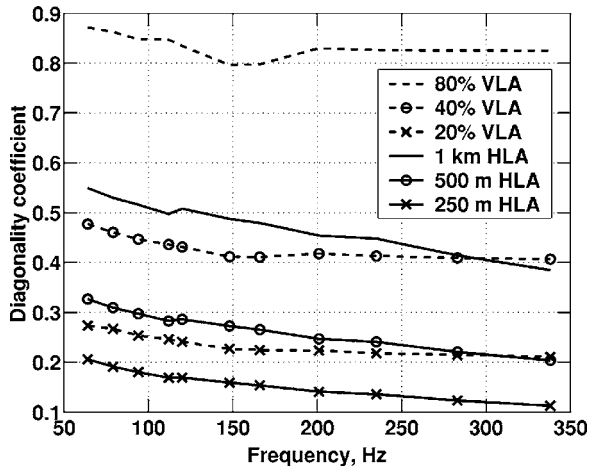
FIG. 8. Diagonality coefficient calculated for mode correlation matrices (MCMs); more diagonal is better. This metric roughly tracks the relative performance of the different arrays, but does not account for the different structure of VLA and HLA MCMs.

Previous studies have used metrics describing the diagonality of the mode correlation matrix (MCM) to compare arrays. The importance of the mode correlation matrix was discussed above. One problem with using a diagonality metric is that the structure of the MCM is not captured.

The MCM can be described using the following metric, which compares the power in the matrix diagonal to the power in all matrix elements:

$$d = \frac{|\mathrm{diag}(\mathbf{E})|^2}{|\mathbf{E}|_{\mathrm{frob}}^2}, \tag{24}$$

where $\mathbf{E}$ is the MCM and "frob" indicates the Frobenius norm.

Figure 8 shows this metric for the arrays and frequency ranges shown above. This metric gives a ranking which is similar to that seen in Fig. 7, but some noticeable differences exist. The diagonality coefficient shows the 1-km HLA and 40% VLA as being extremely similar, with the 1-km HLA having a slightly more diagonal MCM. Figures 7 and 2 both show the 1-km HLA as having slightly *worse* performance, in spite of its higher diagonality. The explanation is that the MCM metric does not capture the MCM structure for HLAs, which typically have worse resolution for the low-order modes that can dominate submerged source response. In contrast, Eq. (13) weights the MCM values by factors that account for mode excitation and attenuation, leading to improved qualitative agreement with direct numerical ranking of array performance.

VIII. SUMMARY

This article has addressed the prediction of sidelobe levels for conventional matched field processing. An approach was developed for analytically predicting the distribution of output power in conventional MFP ambiguity surfaces as a function of array aperture and processing bandwidth. These investigations are intended to provide tools for MFP array design that complement the use of numerical simulations and array design trade studies.

APPENDIX: VARIANCE CALCULATION

In this appendix the variance of the Y_{mn} terms shown in Eq. (8) is calculated. This variance is used to calculate the parameter μ describing the exponential distribution for narrow-band CMFP.

To review, the quantities in the CMFP power output which are randomly varying under the assumptions listed in Sec. IV are collected into the term

$$Y_{nm} = \frac{2\pi}{\sqrt{k_n k_m}} \frac{\psi_n(z_s)\psi_m(z)e^{i(k_n r_s - k_m^* r)}}{|\mathbf{p}_{\mathrm{look}}(\Theta)||\mathbf{x}(\Theta_S)|}. \tag{A1}$$

The exponential terms in the random quantity Y_{nm} can be rewritten as

$$e^{i(k_n r_s - k_m^* r)} = e^{i\phi}e^{-\alpha_m r}e^{-\alpha_n r_s}. \tag{A2}$$

The phase associated with the random source and look direction ranges has been collected into a phase term $e^{i\phi}$. Although this phase actually depends on r and r_s, treating it separately separates out the very rapid fluctuation in phase from the much slower change in attenuation. Y_{nm} is a complex zero-mean random variable. Since the argument to the exponential can range over $0-2\pi$, the real and imaginary parts have equal variance, given by

$$\sigma_{nm}^2 = \frac{4\pi^2}{k_n k_m} \int \frac{\psi_m^2(z)\psi_n^2(z_s)}{|\mathbf{p}_{\mathrm{look}}(\Theta)|^2 |\mathbf{x}(\Theta_S)|^2}$$
$$\times \cos^2[\phi(r)]e^{-2\alpha_n r_s}e^{-2\alpha_m r}p_{\mathbf{r}}(\mathbf{r})\,d\mathbf{r} \tag{A3}$$

with $\mathbf{r}$ representing the independent random variables r_s, z_s, r, z, and ϕ, and $p_{\mathbf{r}}(\mathbf{r})$ representing their probability density function. The $\cos^2[\phi(r)]$ term comes from the real part of the complex exponential. A similar equation containing a $\sin^2[\phi(r)]$ term is obtained for the variance of the imaginary part. These terms both average to $\frac{1}{2}$, giving the result quoted in Sec. IV.

The normalization of the pressure fields is required next. For an array that resolves the modes well, off-diagonals will not contribute strongly to the intensity, and the norm-squared for the look direction vector is

$$|\mathbf{p}(\Theta)|^2 \approx \sum_l \frac{2\pi}{k_l}\psi_l^2(z)e_{ll}e^{-2\alpha_l r}. \tag{A4}$$

Attenuation acts to eliminate contributions from high angle modes, so almost all the contribution at longer ranges is from the trapped or "proper" modes. This allows a further simplification:

$$|\mathbf{p}(r,z)|^2 \approx \sum_{l=1}^{N_p} \frac{2\pi}{k_l} \psi_l^2(z) e_{ll} e^{-2\alpha_l r_{\text{mid}}}, \qquad (A5)$$

where r_{mid} is the mid-range of the MFP search region whose statistics are being studied. Numerical tests indicated that this further simplification did not introduce errors for the cases studied.

Next, the approximation in Eq. (A5) is substituted into Eq. (A4). Because the source location and look directions are independent random variables, integration over these variables can be done separately:

$$\sigma_{nm}^2 = \frac{4\pi^2}{k_n k_m} \left[\frac{1}{\Delta_{Dl}} \int_{D_{l1}}^{D_{l2}} dz \frac{\psi_m^2(z)}{\sum_{l=1}^{N_p} (a/k_l) \psi_l^2(z) e_{ll} e^{-2\alpha_l r_{\text{mid}}}} \right]$$

$$\times \left[\frac{1}{\Delta_{Ds}} \int_{D_{s1}}^{D_{s2}} dz_s \frac{\psi_n^2(z_s)}{\sum_{p=1}^{N_p} (a/k_p) \psi_p^2(z_s) e_{pp} e^{-2\alpha_p r_{\text{mid}}}} \right]$$

$$\times \left[\frac{1}{\Delta R} \int_{R}^{R+\Delta R} dr\, e^{-2\alpha_m r} \right]$$

$$\times \left[\frac{1}{\Delta R} \int_{R}^{R+\Delta R} dr_s\, e^{-2\alpha_n r_s} \right]. \qquad (A6)$$

The first two terms in brackets are the Z_m terms from Sec. IV, while the two range integrals can be evaluated to give the R_m terms. The equation above then gives the result quoted in Eq. (10).

[1]W. Burdic, *Underwater Acoustic Systems Analysis*, 2nd ed., (Prentice-Hall, Englewood Cliffs, NJ, 1991).

[2]A. Baggeroer, W. Kuperman, and P. Mikhalevsky, "An overview of matched field methods of ocean acoustics," IEEE J. Ocean. Eng. **18**(4), 401–424 (1993).

[3]H. C. Song, "Performance bounds on the passive localization of a moving source for ocean acoustics," Ph.D. thesis, Massachusetts Institute of Technology, September 1990.

[4]J. Buck, J. Preisig, and K. Wage, "A unified framework for mode filtering and the maximum a posteriori mode filter," J. Acoust. Soc. Am. **103**, 1813–1824 (1998).

[5]V. Premus, "Mode scintillation index: a physics-based statistic for acoustic source depth discrimination," J. Acoust. Soc. Am. **105**, 2170–2180 (1999).

[6]A. Baggeroer and H. Cox, "Passive sonar limits upon nulling multiple moving ships with large aperture arrays," 33rd Asilomar Conference on Signals, Systems and Computers, pp. 103–108 (1999).

[7]C. 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).

[8]S. Tantum and L. Nolte, "On array design for matched-field processing," J. Acoust. Soc. Am. **107**, 2101–2111 (2000).

[9]G. Smith, C. Feuillade., and D. DelBalzo, "Matched-field processing enhancement in a shallow-water environment by incoherent broadband averaging," J. Acoust. Soc. Am. **91**, 1447–1454 (1992).

[10]L. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," J. Acoust. Soc. Am. **113**, 2719–2731 (2003).

[11]M. Porter, "The KRAKEN normal mode program," SACLANT Undersea Research Centre Technical Report SM–245, 1991.

[12]R. Brienzo and W. Hodgkiss, "Broadband matched-field processing," J. Acoust. Soc. Am. **94**, 2821–2831 (1993).

[13]E. Westwood, "Broadband matched-field source localization," J. Acoust. Soc. Am. **91**, 2777–2789 (1992).

[14]A. Baggeroer, W. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**, 571–587 (1988).

[15]A. Thode, W. Kuperman, G. D'Spain, and W. Hodgkiss, "Localization using Bartlett matched-field processor sidelobes," J. Acoust. Soc. Am. **107**, 278–286 (1990).

Experimental demonstration of adaptive reverberation nulling using time reversal

H. C. Song, W. S. Hodgkiss, W. A. Kuperman, P. Roux, and T. Akal[a]
Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0238

M. Stevenson
NATO Undersea Research Centre, 19138 La Spezia, Italy

The concept of environmentally adaptive reverberation nulling using a time reversal mirror (TRM) recently has been described [Song *et al.*, J. Acoust. Soc. Am. **116**, 762–768 (2004)]. In this paper, monostatic reverberation nulling is demonstrated experimentally at 850 and 3500 Hz using data from a shallow water experiment conducted off the west coast of Italy in April 2003. The active transmission of a seafloor spatial null from a vertical source array is shown to result in the attenuation by 3–5 dB of prominent reverberation features with their levels being reduced to that of the more diffuse reverberation background. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1984990]

I. INTRODUCTION

Bottom reverberation severely limits the detection of low-speed targets by active sonar systems in a shallow water environment.[1] As a consequence of waveguide propagation, some regions of the seafloor are strongly ensonifed, resulting in high levels of backscatter observed at specific times by the receiver even if the seafloor is relatively flat over the region of interest. It would be desirable to avoid ensonifying the seafloor while at the same time maintaining coverage of the water column for target detection purposes.

Recently, a series of ocean acoustic experiments has been carried out confirming the robustness and potential utility of time reversal mirrors (TRM) in underwater acoustics.[2–4] The focusing capability of a TRM without *a priori* knowledge of the environment suggests two different approaches for active target detection in reverberation-limited environments. First, a TRM can focus acoustic energy on a target while shadowing the boundaries below and above the focus in a waveguide, thereby reducing reverberation. The resulting echo-to-reverberation enhancement has been demonstrated experimentally in shallow water using a TRM in the 3–4-kHz band.[5]

The other approach to enhance active target detection is reverberation nulling recognizing that focusing and nulling are complementary. Lingevitch *et al.*[6] recently have shown in simulation that backscattering from the rough water-bottom interface can serve as a surrogate probe source in time reversal. The concept is based on the fact that a time-gated portion of reverberation provides an estimate of the transfer function vector between a TRM array and the range cell along the bottom. The transfer function vector then could be used to refocus energy to the corresponding bottom interface. On the other hand, the idea behind reverberation nulling is to mini-

mize the acoustic energy incident on the boundary in a specific range cell while more energy is projected into the water column possibly illuminating targets. To achieve this, a weight vector which is in a complementary subspace orthogonal to the focusing vector is applied on the TRM prior to retransmission. A theoretical study of reverberation nulling is described in Ref. 7 along with numerical simulations.

As a companion to Ref. 7, this paper presents results of reverberation nulling from an April 2003 TRM experiment conducted in shallow water off the west coast of Italy. We are not aware of other mid-frequency, adaptive reverberation nulling approaches and related experimental results published in the literature. Section II briefly reviews the concept of reverberation nulling. Section III presents experimental results at 850 and 3500 Hz. Conclusions then follow in Sec. IV.

II. REVERBERATION NULLING: DESCRIPTION

The theory of reverberation nulling is described in detail in Ref. 7. In this section, we briefly overview the procedure for reverberation nulling. It is assumed that roughness at the water/bottom interface is the dominant backscattering mechanism and we confine our interest to narrow-band signals. It also is assumed that both the environment and bottom roughness are axisymmetric.

The fundamental idea behind reverberation focusing/nulling is based on the fact that a time-windowed segment of reverberation provides an estimate of the transfer function vector $\mathbf{h}(f)$ between a TRM array and the corresponding range cell along the bottom.[6] The reverberation then can be treated as a single (extended) source problem in array processing. This suggests that a single snapshot would permit obtaining an estimate of the transfer function vector when

---

[a] T. Akal was with NATO Undersea Research Centre, 19138 La Spezia, Italy.
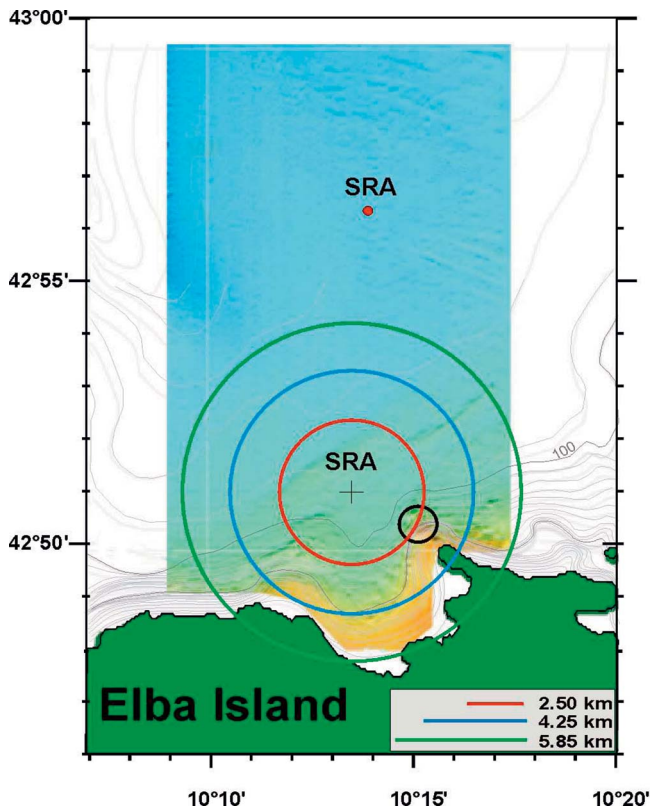
FIG. 1. The FAF-03 experiment took place north of Elba Island, Italy. The monostatic reverberation portion of the experiment involving the SRA only was conducted in two different areas at two different frequencies: 850 and 3500 Hz. For 850 Hz, the SRA was moored about 6 nautical miles north of Elba in a relatively flat area of water depth 117 m (upper part). For 3500 Hz, the SRA was positioned closer to Elba (about 2 miles north) in water depth of 105 m having a shoaling bathymetry towards Elba (lower part with circles).

there is a strong reverberation return, i.e., the high reverberation-to-noise and/or reverberation-to-echo case. In fact, we will use a single snapshot to demonstrate reverberation nulling in Sec. III.
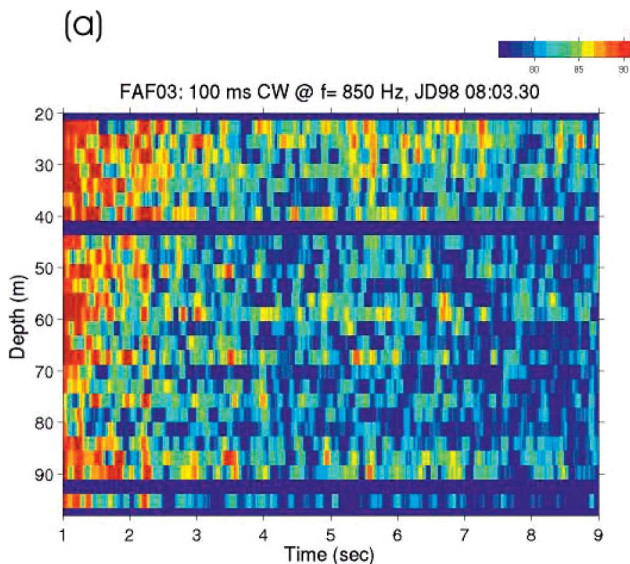
Typically, a single transducer does not have sufficient source level to induce significant backscattering so that simultaneous excitation (i.e., a beam) with maximum power is desirable. For instance, a broadside transmission refers to an excitation with equal amplitudes across the array element. Although one can use arbitrary beams, in practice it will be difficult to excite high-angle beams efficiently (which corresponds to the higher order modes) since this requires both a dense sampling and long aperture of the TRM array. Usually multiple snapshots $\mathbf{h}_i$ are collected from either different beams or repeated use of the same beam in order to construct a data covariance matrix $\mathbf{R} = \Sigma \mathbf{h}_i \mathbf{h}_i^{\dagger}$ where $\dagger$ denotes a conjugate transpose (Hermitian) operation.

Since focusing and nulling are complementary problems, the success of nulling depends on obtaining a transfer (focusing) vector. This can be done by selecting a time window as described in Ref. 7 which suggests a time window less than twice the signal pulse length. The time-gated reverberation then can reduce to a single source problem such that the effective rank of $\mathbf{R}$ is equal to 1. The complex conjugate of the first eigenvector spanning the signal subspace is used as a new excitation weight vector on the TRM for retransmission to focus the energy onto the range cell, i.e., $\mathbf{w}_{focus} = \mathbf{v}_1^{*}.$[7]

Once a focusing vector is obtained, nulling can be achieved by choosing an arbitrary vector in the $(N-1)$-dimensional noise subspace orthogonal to the one-dimensional focusing vector where N is the number of elements of the TRM array. The natural question is then how to select a nulling weight vector optimally in the sense that maximum energy is projected into the water column. Unfortunately, we cannot further optimize the nulling vector $\mathbf{w}_{null}$ since no other information is available from the reverberation return except the transfer function between the array and the corresponding range cell.
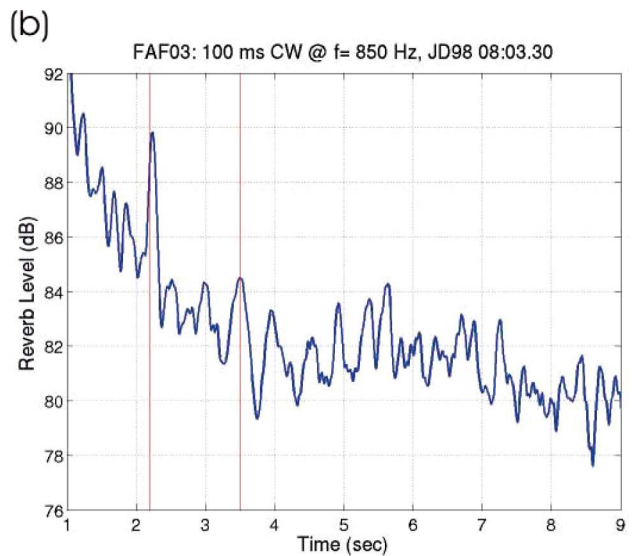
FIG. 2. (a) Monostatic reverberation return from an 850-Hz, 100-ms broadside transmission from the SRA. Four transducers (channels 1, 3, 21, and 29 numbered from the bottom) were not operating. (b) Reverberation time series obtained by incoherent averaging the active 25 channels. The reverberation is strong initially and decreases in about 7 s to the ambient noise level. Note the two strong reverberation returns at 2.2 s and 3.5 s.
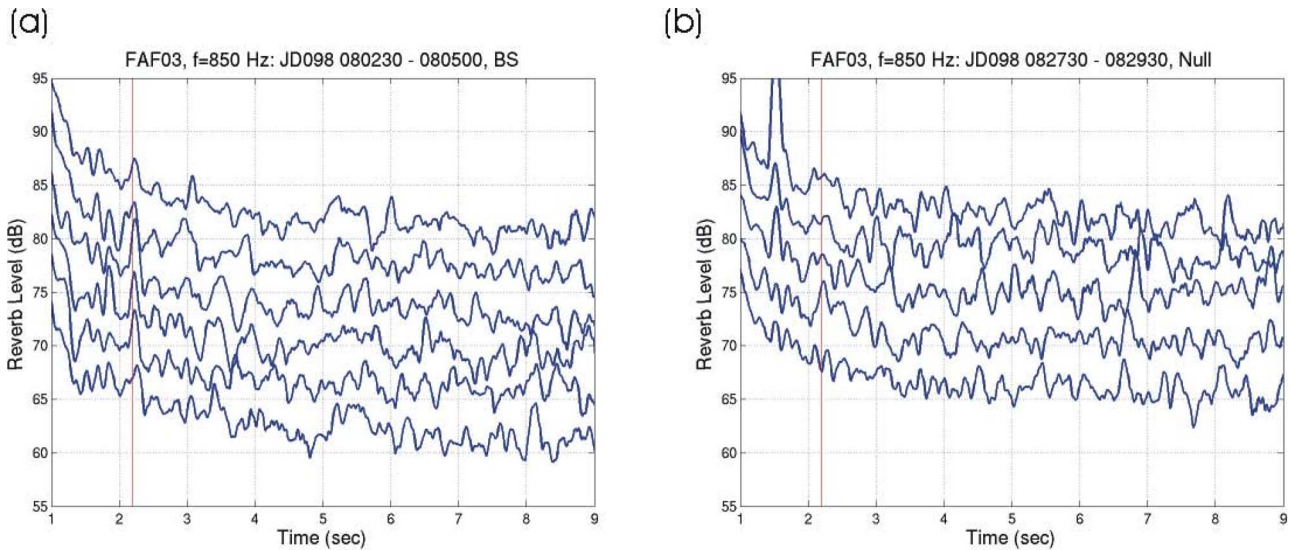
FIG. 3. Ping-to-ping variability of reverberation returns from an 850-Hz, 100-ms transmission over several consecutive pings spaced every 30 s. For convenience, adjacent pings are displaced relative to one another by 4 dB and the upper curve represents the absolute level. (a) Broadside transmission and (b) reverberation nulling.

III. REVERBERATION NULLING: EXPERIMENT

A. Reverberation experiment

A time-reversal experiment was carried out in April 2003, north of Elba Island off the west coast of Italy (Fig. 1). The monostatic reverberation portion of the experiment involving the source-receiver array (SRA) was conducted at two different sites using two different frequencies, 850 and 3500 Hz, respectively. Figure 1 (upper part) shows the SRA located about 6 nautical miles north of Elba in a relatively flat area of water depth 117 m for the 850-Hz transmissions. For the 3500-Hz transmissions, the SRA was positioned closer to Elba (about 2 miles north) in water depth of 105 m having a shoaling bathymetry towards Elba as shown in Fig. 1 (lower part with circles). Both 850 and 3500 Hz SRAs had 29 transducers spanning a 78-m aperture with 2.786-m inter-element spacing. The SRAs covered the water column from
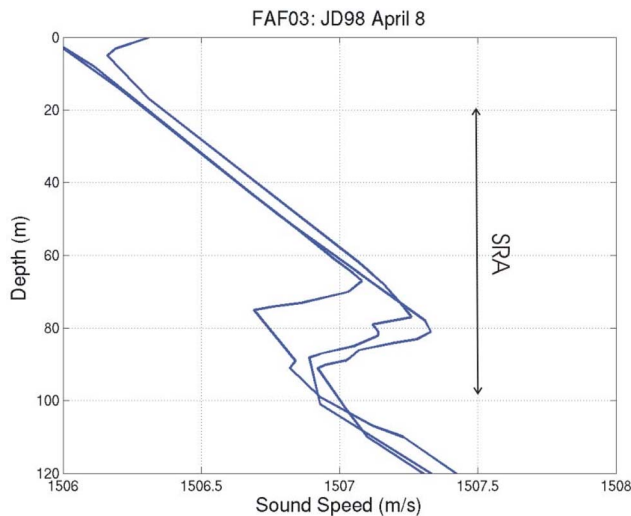
FIG. 4. Measurements of the sound speed profile were obtained from CTD casts taken by the *R/V Alliance* on JD98 (8 April) during the 850-Hz reverberation experiment. The SRA covered the water column from 20 to 98 m.

20 to 98 m. The vector time series transmitted by the SRA is normalized such that the maximum value across all elements and time is equal to the maximum source level of 182 and 178 dB *re* 1 μPa at 850 and 3500 Hz, respectively. A detailed description of the TRM hardware and 3500-Hz transducers can be found in Ref. 8. The SRAs were moored for stable operations on 8 April (JD98) for 850 Hz and 13 April (JD103) for 3500 Hz.

B. Reverberation nulling

The nulling experiments were conducted as follows.

(1) Transmit a 100-ms CW broadside transmission (BT) from the SRA at either 850 or 3500 Hz and record the returning reverberation time series.
(2) Select a time window of 100 ms corresponding to the range (or time) of interest and apply a Fourier transform at the center frequency of either 850 or 3500 Hz to obtain an estimate of the array transfer function vector $\mathbf{h}$ from the seafloor to the SRA elements. Thus, $\mathbf{h}^*$ is the focusing vector where * denotes a conjugate operation.
(3) A nulling weight vector $\mathbf{w}_{null}$ is obtained by projecting the broadside beam $\mathbf{e}$ onto the noise subspace such that

$$\mathbf{w}_{null} = [\mathbf{I} - \mathbf{h}^*\mathbf{h}^T/|\mathbf{h}|^2]\mathbf{e}, \qquad (1)$$

where $\mathbf{I}$ is an N-dimensional identity matrix, $\mathbf{e}$ is an N-dimensional column vector with entires equal to 1, and T denotes a transpose operation.
(4) Generate a vector time series from the excitation weight vector $\mathbf{w}_{null}$ (i.e., 100-ms CW with different amplitudes and phases across the array) and broadcast the resulting time series.

Note that steps 1 and 2 can be repeated by exciting the SRA with different beams or the same broadside beam at multiple times to construct a sample covariance matrix $\hat{\mathbf{R}} = (1/K)\sum \mathbf{h}_i\mathbf{h}_i^\dagger$ where K is the number of snapshots involved.
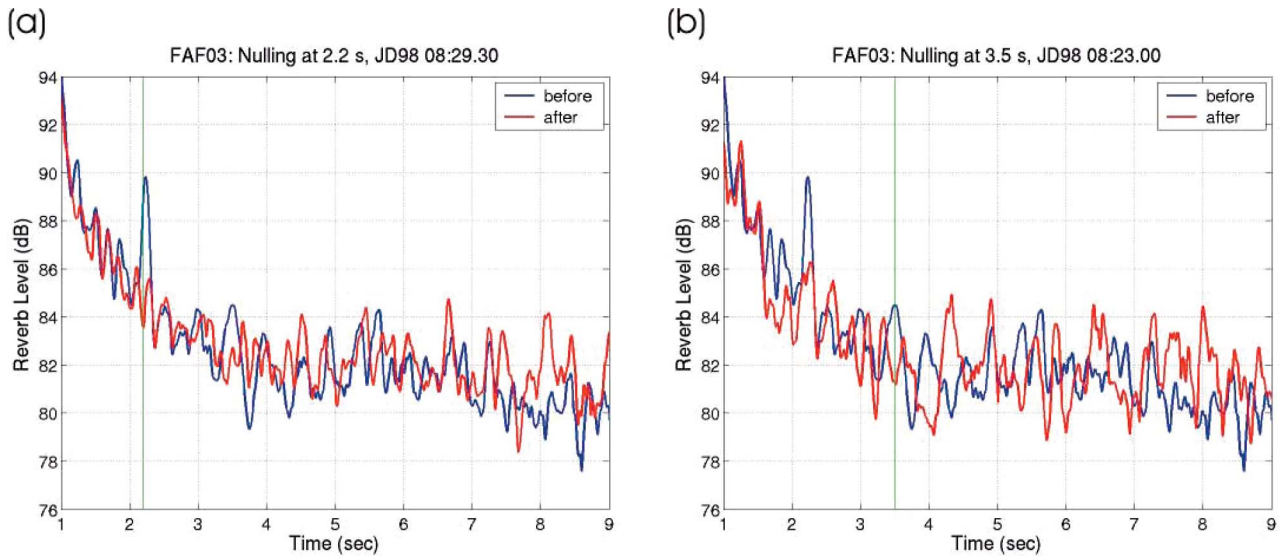
FIG. 5. Reverberation return resulting from a new excitation weight vector for nulling when applied to range windows at (a) 2.5 s and (b) 3.5 s, respectively. The original reverberation return from a broadside transmission in Fig. 2(b) is superimposed for comparison purposes. While (a) indicates as much as 5-dB reduction in the reverberation level compared to the background level at 2.2 s, (b) shows about 2-dB decrease. Note that suppression at a specific time increases the level at other times.

In our nulling experiment, however, we collected just a few reverberation returns from broadside transmissions and selected a single reverberation return which apparently exhibited a strong level at the range (or time) of interest for processing. In other words, a single snapshot was used to extract an estimate of the transfer function vector from the seafloor when there is a strong reverberation return, i.e., the high reverberation-to-noise case.

Note that there is freedom to choose any of the eigenvectors in the noise subspace or an arbitrary combination of them as described in Sec. II. In our nulling experiment, an additional constraint in step 3 has been imposed such that the nulling weight vector is the projection of the broadside beam onto the subspace spanning the noise eigenvectors. This was done for two reasons. First, one can compare the reverbera-

tion nulling resulting from the new weight vector with the original reverberation return from a broadside transmission. Second, a broadside beam will excite lower order modes efficiently given the aperture of the array and element spacing.

1. f = 850 Hz

Figure 2(a) shows a reverberation return from a 100-ms broadside transmission from the SRA which is selected out of six consecutive pings over the 3-min period in Fig. 3(a). For convenience, adjacent curves are displaced relative to one another by 4 dB and the upper curve represents the absolute level. Note the ping-to-ping variability illustrating the repeatability over the time period. During this measurement, four of the SRA transducers (channels 1, 3, 21, and 29 numbered from the bottom) were not operating properly and were disabled (thick horizontal blue lines) in Fig. 2(a). The reverberation time series displayed in Fig. 2(b) was obtained by incoherent averaging the active 25 channels. The reverberation is strong initially and decreases in about 7 s to the ambient noise level. Figure 4 shows the sound speed profiles collected during the reverberation experiment on 8 April.

For reverberation nulling, we concentrated on the two strong peaks around 2.2 and 3.5 s in Fig. 2(b). Resulting reverberation returns (after) due to a new excitation weight for nulling are shown in Figs. 5(a) and 5(b) when applied to 2.2 and 3.5 s, respectively. The original broadside transmission (before) is superimposed for comparison purposes. While Fig. 5(a) shows as much as 5-dB reduction in the reverberation level compared to the background at 2.2 s, Fig. 5(b) indicates the level at 3.5 s is reduced by about 3 dB. Note that suppression of reverberation at a specific time, however, increases the reverberation level at other times. Figure 3(b) indicates that reverberation nulling is maintained over four out of five pings in the 3-min period.
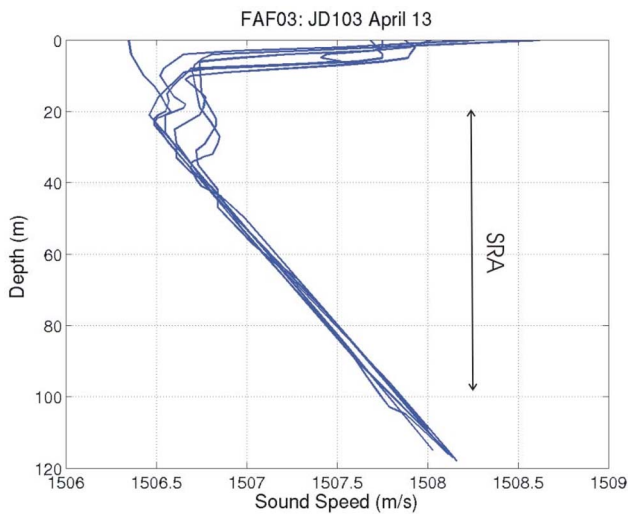
FIG. 6. Sound speed measurements on 13 April during the 3500-Hz reverberation experiment. The SRA covered the water column from 20 to 98 m. The profile is similar to a deep sound channel with the channel axis around 20 m.

1384   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Song *et al.*: Experimental demonstration of reverberation nulling
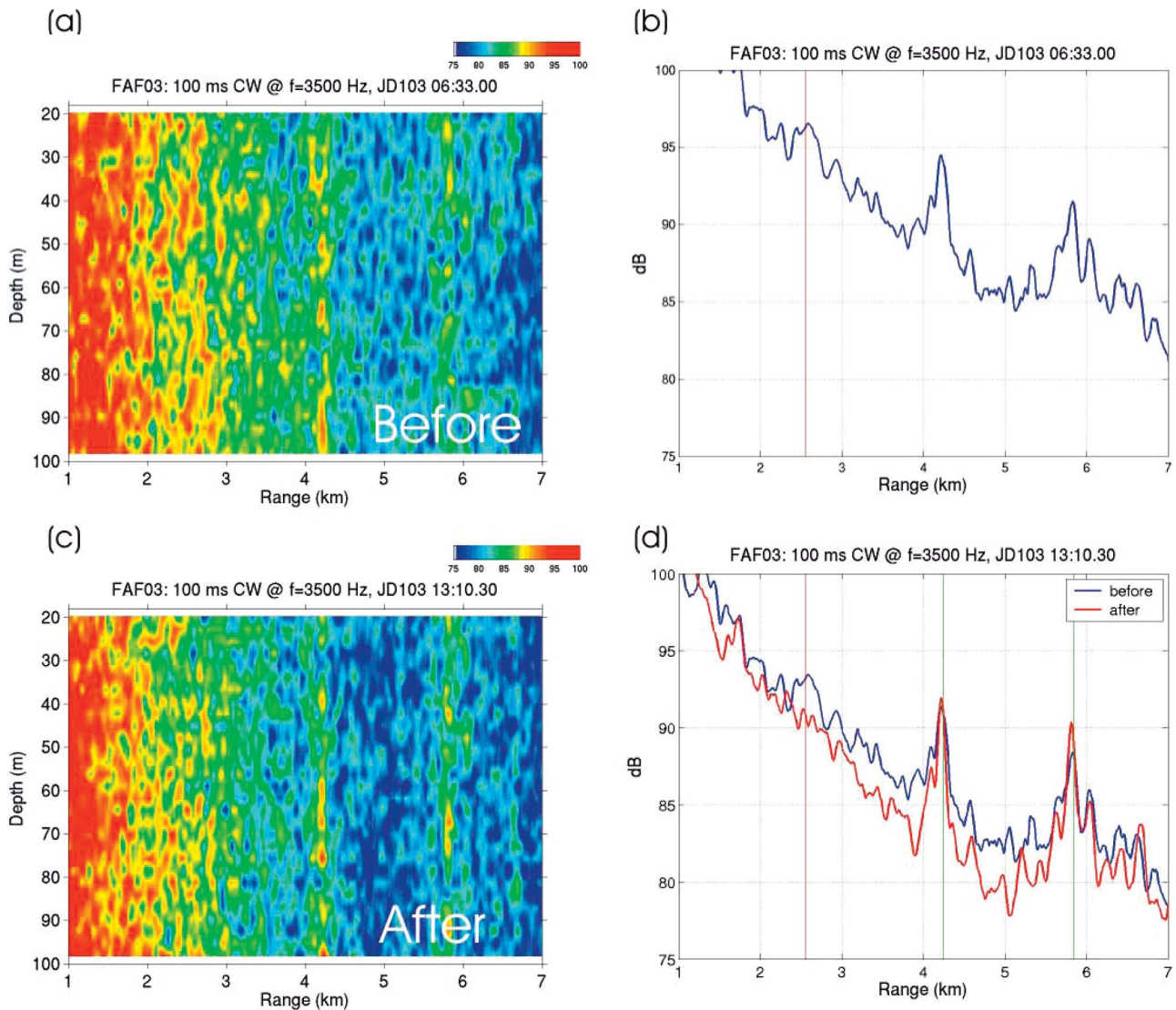
FIG. 7. (a) Monostatic reverberation return (06:33:00) from a 3500-Hz, 100-ms broadside SRA transmission as a function of range (converted from time). (b) Reverberation level obtained by incoherent averaging across the channels. Due to the proximity of the SRA to Elba, the two prominent peaks around 4.25 and 5.85 km result from the interaction with Elba corresponding to the outer concentric circles denoted in the Fig. 1(b). The peak at 2.5 km, however, is due to a seamount at the corresponding range which is visible in the bottom topography in Fig. 1(b) denoted by a black circle. (c) and (d) Reverberation return after nulling (13:10:30). The reverberation level at 2.5-km range is reduced by about 3 dB to the background level. Note that the reverberation level at 5.8-km range has increased.

It is interesting to observe from Fig. 5 that the reverberation nulling intended for either 2.2 or 3.5 s leads to nulling at both times. This implies that the two peaks are related through the propagation characteristics of the waveguide.

2. f=3500 Hz

At 3500 Hz, the SRA was deployed in 105-m water depth close to Elba Island as shown in Fig. 1 (lower part). Figure 6 shows sound speed profiles collected during this portion of the experiment on 13 April which look similar to a deep sound channel with a channel axis around 20 m. A typical example of the reverberation return from a broadside transmission is shown in Fig. 7 while Fig. 8(a) shows ping-to-ping variability over the 3-min period as in Fig. 3. Figure 7(b) displays the incoherent average of the reverberation level across the SRA. Note that the time axis is replaced by range r using the relationship of $r=2t/c$ with $c=1500$ m/s.

As expected, the reverberation level at 3500 Hz is higher than the level at 850 Hz by about 10 dB at 3 s. Considering the transmitted source level at 3500 Hz which is 4 dB less than that at 850 Hz, the overall reverberation level is approximately 18 dB higher.

Due to the proximity of the SRA to Elba Island, the two prominent peaks around 4.25 and 5.85 km result from the interaction with Elba Island corresponding to the outer concentric circles denoted in Fig. 1. The broad peak at 2.5 km, however, is due to a seamount at the corresponding range which is visible in the bottom topography (black circle) shown in Fig. 1. Thus the peak at 2.5-km range provides a good candidate for reverberation nulling.

The results after implementing reverberation nulling are displayed in Figs. 7(c) and 7(d), indicating reduction of the reverberation level by about 3 dB to the background level at 2.5 km. On the other hand, the reverberation return from the

Song *et al.*: Experimental demonstration of reverberation nulling
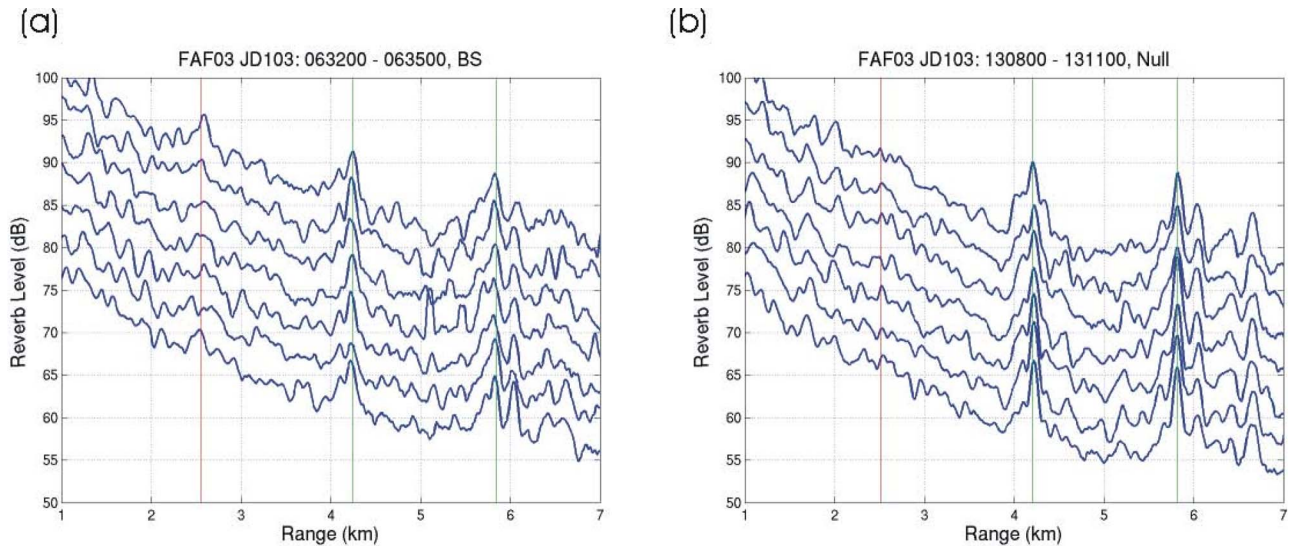
FIG. 8. Ping-to-ping variability of reverberation returns from a 3500-Hz, 100-ms transmission over several consecutive pings spaced every 30 s. For convenience, adjacent pings are displaced relative to one another by 4 dB and the upper curve represents the absolute level. (a) Broadside transmission and (b) reverberation nulling.

interaction with Elba has increased at 5.8 km range. Figure 8(b) indicates that we can achieve reverberation nulling most of the time over the 3-min period. It should be mentioned that the reverberation level after nulling has been increased by 2.4 dB to facilitate comparison with the original broadside reverberation return. This has been done to accommodate the normalization described in Sec. III A where the normalized amplitudes for all of the elements are equal to 1 for a broadside transmission whereas only a single element has amplitude equal to 1 for reverberation nulling. Since the nulling experiment shown in Fig. 7(c) and 7(d) was conducted almost 7 h after the capture of the reference broadside reverberation return (06:33:00) in Fig. 7(a), it is interesting to observe a reverberation return from a broadside transmission right after nulling (13:10:30). Figure 9 suggests that the

reverberation structure at 13:28:00 remains unchanged with a consistent sharp peak at 2.5-km range.

IV. CONCLUSIONS

In this paper, the feasibility of reverberation nulling using a time reversal mirror without knowledge of the environment was demonstrated experimentally in shallow waters at two different frequencies. At 850 Hz, the reverberation level was reduced up to 5 dB. At 3500 Hz, the prominent reverberation return from a nearby seamount was reduced approximately 3 dB.
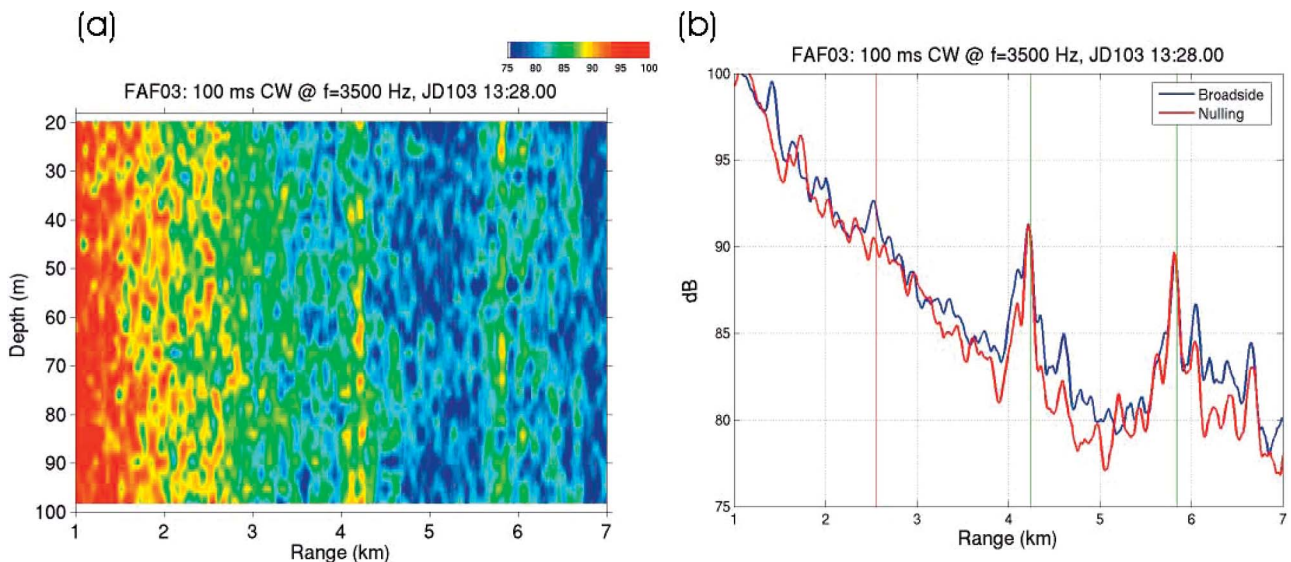
FIG. 9. (a) Monostatic reverberation return from a 3500-Hz, 100-ms broadside SRA transmission (13:28:00) as a function of range which was observed after the nulling experiment (13:10:30) shown in Fig. 7(c). (b) Reverberation level obtained by incoherent averaging across the channels.

[1] H. Cox, "Space-time processing for suppression of bottom reverberation," in *Proceedings of 29th Asilomar Conference*, 1995, pp. 1296–1299.

[2] W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **102**, 25–40 (1998).

[3] W. S. Hodgkiss, H. C. Song, W. A. Kuperman, T. Akal, C. Ferla, and D. R. Jackson, "A long range and variable focus phase conjugation experiment in shallow water," J. Acoust. Soc. Am. **105**, 1597–1604 (1999).

[4] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**, 820–829 (2001).

[5] S. Kim, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, G. F. Edelmann, and T. Akal, "Echo-to-reverberation enhancement using a time reversal mirror," J. Acoust. Soc. Am. **115**, 1525–1531 (2004).

[6] J. F. Lingevitch, H. C. Song, and W. A. Kuperman, "Time reversed reverberation focusing in a waveguide," J. Acoust. Soc. Am. **111**, 2609–2614 (2002).

[7] H. C. Song, S. Kim, W. S. Hodgkiss, and W. A. Kuperman, "Environmentally adaptive reverberation nulling using a time reversal mirror," J. Acoust. Soc. Am. **116**, 762–768 (2004).

[8] W. S. Hodgkiss, J. D. Skinner, G. E. Edmonds, R. A. Harriss, and D. E. Ensberg, "A high frequency phase conjugation array," in *Proceedings of Oceans 2001*, 2001, pp. 1581–1585.

A continuous wave method for ultrasonic characterization of liquid materials

Yves Deblock,[a] Pierre Campistron, and Bertrand Nongaillard

IEMN-DOAE, UMR CNRS 8520, BP 69, Cite Scientifique, 59652, Villeneuve d'Ascq, France

This paper describes a method for measuring the velocity and attenuation of high-frequency sound waves traveling through liquid materials. This method is based on the measurement of the forward scattering parameter (S_{21}) of the acoustic system composed of two transducers and the liquid between them. The acoustic parameters are obtained from the analysis of S_{21} in the time domain. Various experiments using distilled water as the liquid sample have been carried out. They show the influence on the measurement sensitivity of receiver detection bandwidth, liquid thickness, and difference of temperature between the room and the test cell. The method was used for the measurement of the attenuation and velocity of mixtures of alcohol and water. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000771]

I. INTRODUCTION

Ultrasonic liquid characterization generally means sensitive measurement of the acoustic properties, i.e., velocity and attenuation coefficient, as a function of frequency. These properties are often related to viscoelastic mechanisms, which involve relaxation phenomena. Broadband frequency studies are of great use in understanding these phenomena.

Usually, one can distinguish two common ways for characterizing a liquid through purely ultrasonic measurements. The major difference lies in the excitation type of the ultrasonic devices: pulsed or continuous wave signals.

The pulse method, first described by Pellam *et al.*,[1] is often used in the biomedical field. Short bursts of ultrasonic waves travel out into the liquid from one transducer to another. The length of the pulses is sufficiently short to prevent overlapping of the different reflected waves in the system. The amplitude of the received pulse as a function of the distance between the transducers gives the attenuation coefficient of sound in the liquid. The pulse transit time as a function of the distance gives the sound velocity.

Interferometric methods that use continuous wave signals derive from optics and were introduced in the acoustic field by Pierce.[2] Standing waves are set up in the liquid between the transducer and a reflector (or another transducer). The detected signal of the transducer presents a succession of flat maxima and sharp minima as the distance between the transducer and the reflector is increased. The distance between successive minima or maxima is half the wavelength in the liquid, therefore the velocity can be determined. The attenuation is obtained from the decrease in the detected signal with the increase of the distance between the transducer and the reflector.

Many different techniques using the basic methods have been developed (see, for example, the comprehensive review of Breazeale *et al.*[3]). More recently a very powerful pulse

method has been developed by Kushibiki *et al.*:[4] attenuation and velocity of sound waves in liquids are obtained from very precise and well-established experimental procedures in the very- and ultra-high frequency ranges.

A method is proposed in this paper for the measurement of attenuation and velocity using continuous wave (CW) signals. This method uses a similar acoustic system as in Kushibiki's method but the velocity and the attenuation of sound waves are obtained from the measurement of the scattering parameter S_{21} of the acoustic system, using a network analyzer. The acoustic system is composed of the liquid sample inserted between the faces of two rods, having transducers on their opposite face. The advantage of this present method over other published methods lies in the ability to use a narrow bandwidth detection of S_{21} that improves the signal-to-noise ratio and makes the measurement of velocity and attenuation more sensitive.

II. MEASUREMENT PRINCIPLE

Figure 1 shows the acoustic system used to determine the velocity and the attenuation of sound waves in liquids. A transducer unit is composed of a rod having a transducer bonded to one end. It is shown in this section how the velocity and the attenuation are obtained from the measurement of the transfer function of the acoustic system (i.e., the scattering parameter S_{21}) using a vector network analyzer.

Briefly, a vector network analyzer consists of a sweeping signal source, a test set to separate forward and reverse test signals, and a multi-channel high-sensitivity receiver. The S-parameters of a network are obtained by comparing these test signals with a reference signal (amplitude ratio and phase difference). The reference signal is obtained by splitting the output signal.

When one transducer is driven by a continuous wave signal, standing waves are set up in the rods and the liquid. Assuming a one-dimensional model for the propagation of acoustic waves in the transducers, the rods, and the liquid, S_{21} can be written as

---

[a]Electronic mail: yves.deblock@iemn.univ-lille1.fr
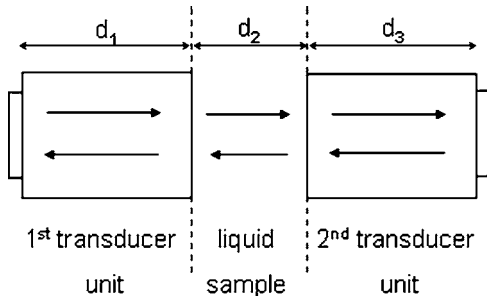
FIG. 1. The schematic representation of the acoustic system for velocity and absorption coefficient measurements of liquid samples.

$$S_{21}(f) = \sum_{n,p,q} K_{npq} \exp[-\alpha(nd_1 + qd_3) - \alpha_l pd_2]$$

$$\times \exp(-j2\pi f t_{npq}) \tag{1}$$

with

$$t_{npq} = \frac{nd_1 + qd_3}{v} + \frac{pd_2}{v_l}. \tag{2}$$

The subscripts n, p and q are odd integers, corresponding to the number of trips the wave travels in the first rod, the liquid, and the second rod, respectively. v and v_l represent the acoustic velocity in the rod and the liquid, respectively. α and α_l represent the attenuation of waves in the rod and in the liquid, respectively. d_1 and d_3 are the rod lengths of the first and second transducer units. d_2 is the gap distance between the transducer units, that is, the liquid thickness. K_{npq} are coefficients that include electromechanical conversion effects of both transducers and the reflection and transmission coefficients of acoustic waves at the interfaces.

Over a narrow frequency range the coefficients K_{npq} and the absorption coefficients α and α_l can be assumed independent of the frequency compared with the complex exponentials $\exp(-j2\pi f t_{npq})$ of Eq. (1). Consequently the discrete inverse Fourier transform applied to $S_{21}(f)$ gives a series of discrete impulses convolved with the inverse Fourier transform of the rectangle window of measurement and $S_{21}(t)$ can be written as

$$S_{21}(t) = A_{npq} \sum_{n,p,q} \frac{\sin \pi(t - t_{npq})\Delta f}{\pi(t - t_{npq})\Delta f} \exp[j2\pi(t - t_{npq})f_{\text{cent}}] \tag{3}$$

with

$$A_{npq} = \Delta f K_{npq} \exp[-\alpha(nd_1 + qd_3) - \alpha_l pd_2]. \tag{4}$$

Δf is the frequency window over which the inverse Fourier transform is performed and f_{cent} is its center frequency.

$S_{21}(t)$ is a superposition of impulses having the shape of the $(\sin x)/x$ function, of amplitude A_{npq} and located at times t_{npq}. When the impulses are well separated from each other, the phase of $S_{21}(t)$ linearly increases versus time with a slope given by $2\pi f_{\text{cent}}$. Separation of the first impulse is achieved by selecting a frequency range that satisfies
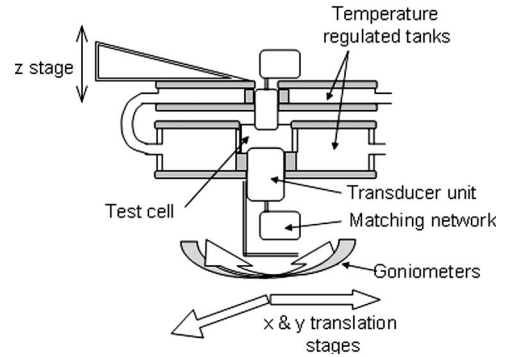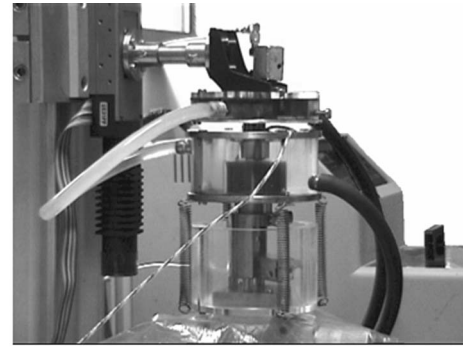
FIG. 2. The experimental arrangement for velocity and attenuation coefficient measurements of liquid samples.

$$\Delta f > \max\left(\frac{v}{2d_1}, \frac{v_l}{2d_2}, \frac{v}{2d_3}\right). \tag{5}$$

On the other hand, Δf must be narrow enough to assume that the K_{npq} are independent of frequency. Δf has been estimated through experiments as

$$\Delta f < \frac{1}{10}W, \tag{6}$$

where W is the transducer bandwidth.

Finally the attenuation and the velocity of sound waves are obtained by measuring the variations of A_{111} and t_{111} of the first impulse of $S_{21}(t)$ as the liquid thickness is changed by Δd_2:

$$\alpha_l = \frac{\Delta A_{111\ \text{dB}}}{\Delta d_2}, \tag{7}$$

with $\Delta A_{111\ \text{dB}} = 20 \log(\Delta A_{111})$ and $\Delta A_{111} = A_{111}(d_2 + \Delta d_2)/A_{111}(d_2)$, and

$$v_l = \frac{\Delta d_2}{\Delta t_{111}} \tag{8}$$

with $\Delta t_{111} = t_{111}(d_2 + \Delta d_2) - t_{111}(d_2)$.

III. MEASUREMENT SYSTEM AND METHOD

Figure 2 shows a schematic drawing and a photograph of the main part of the measurement system. Both ultrasonic units are inserted in the inner cylinder of the cylindrical shaped tanks. The tanks are connected together and temperature regulated by water circulation. A cryostat holds the temperature of water constant in the tanks to within 0.1 °C. As it is shown in Fig. 2, a cavity is formed by the inner cylinder of
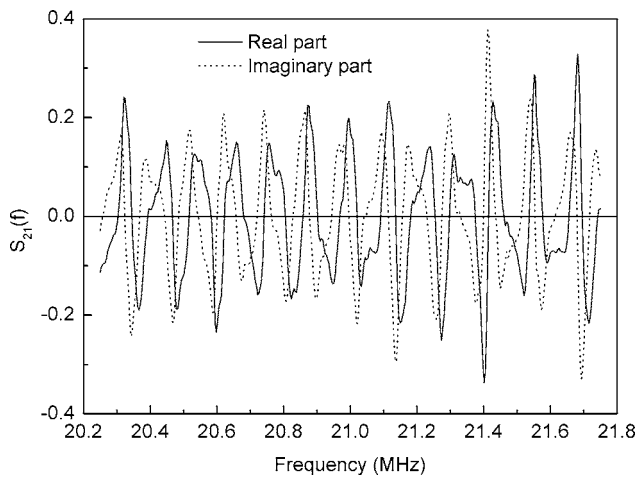
FIG. 3. An example of the scattering parameter S_{21} of the acoustic system as a function of frequency.

FIG. 4. The amplitude of the discrete inverse Fourier tranform of the scattering parameter shown in Fig. 3.

the lower tank and by the ultrasonic unit. This cavity constitutes the test cell for the liquid under investigation. The upper tank and its ultrasonic unit are fixed on a motorized z-axis stage of resolution 1 μm, allowing the distance between the transducer units to be set. The lower tank and its ultrasonic unit are fixed on a complex manual positioning stage including two goniometers and two x- and y-axis stages, which allows the parallelism and alignment between the ultrasonic units to be set.

An ultrasonic unit consists of a piezoelectric LiN_bO_3 plate bonded to one end of a SiO_2 rod. The plate thickness corresponds to one half wavelength of the longitudinal wave traveling in the plate at the resonant frequency. The plate diameter and the rod length are related in such a way that the limit between the Fresnel and Fraunhofer zones of the acoustic field in the rod corresponds to the rod length, and so the diffraction effects are minimized in the rod.

Four sets of ultrasonic units have been designed, capable of measurements from 20 to 100 MHz. Electrical matching of all transducers was done using two-reactance matching networks providing better electromechanical conversions. The complete description of the matching procedure has been given elsewhere.[5]

A network/spectrum analyzer (HP 4195A) was used, in the network mode, to measure the scattering parameter S_{21}. The frequency was swept over a span of Δf and 401 samples were measured and sent to a computer for processing. The inverse Fourier transform was calculated with a sampling resolution of 1 ns.

The mechanical positioning system plays a very important role in the measurement accuracy. First of all, alignment and parallelism are roughly adjusted by eye as the transducers are brought closer but without contact. Afterwards, fine adjustment is performed by maximizing the amplitude of the detected signal, using a radio frequency burst as the driving signal of the emitter transducer.

Figure 3 shows typical records of the components of S_{21}. In this experiment, two transducers of 21-MHz center frequency were used. The transducers were bonded to rods having lengths of 10 and 20 mm. The distance between the transducer units was arbitrarily set to 5.6 mm. Measurements
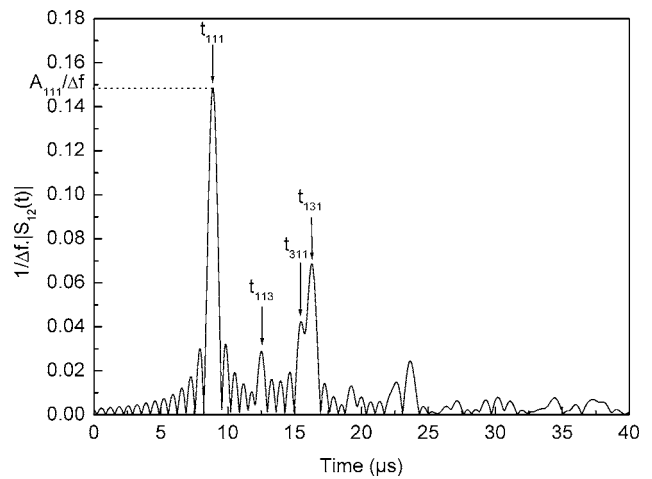
were done over a frequency range of 1.5 MHz, centered around 21 MHz. Figure 4 shows $|S_{21}(t)|$, the amplitude of the inverse Fourier transform of $S_{21}(f)$. Impulses come from multiple reflections of the acoustic waves in the two rods and in the water. They are identified by the times t_{npq} defined by Eq. (2). When the impulses are well separated from each other Eq. (3) predicts that the phase of $S_{21}(t)$ linearly increases versus time with a slope given by $2\pi f_{cent}$, and shifts correspondingly with the successive impulses. This must be checked by subtracting the phase $\phi(t) = 2\pi f_{cent}t$ from the phase of $S_{21}(t)$ and studying whether the resulting phase difference is constant. Figure 5 plots this difference around the first impulse. As expected the phase is roughly constant over the main lobe with shifts of π in the side lobes.

The attenuation and the velocity of sound waves in liquids are obtained by measuring the variations of A_{111} and t_{111} as the gap distance is changed by Δd_2, according to Eqs. (7) and (8). Finding the time corresponding to the absolute maxima of $|S_{21}(t)|$ is imprecise, thus t_{111} is redefined as the nearest time in the neighborhood of the maximum of $|S_{21}(t)|$ for which the phase goes through zero.
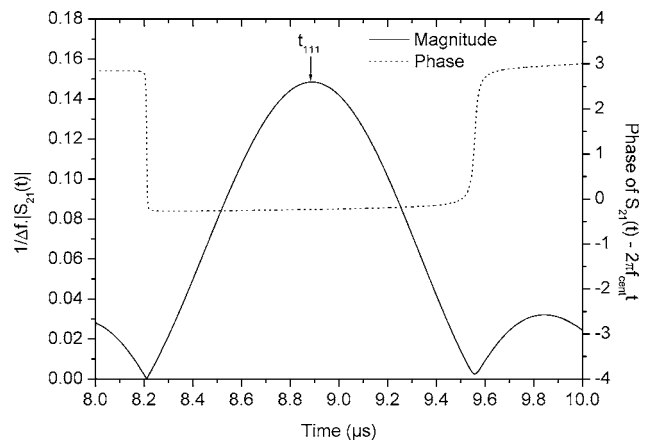
FIG. 5. The amplitude and subtracted phase of the discrete inverse Fourier transform of S_{21} observed on a small time range around the first transmitted impulse.

Deblock *et al.*: Method of ultrasonic characterization of liquids

IV. EXPERIMENTS AND DISCUSSIONS

The purpose of this section is to show the reliability of the measurement system. Acoustic properties of distilled water are well known over several experimental conditions and thus it was used as the test liquid.[6,7] Data reported here were not subjected to any mathematical process such as averaging. The various experiments which were performed show the important characteristics of the measurement method. The influence on the measurement accuracy of the bandwidth of the intermediate frequency filter of the network analyzer and the thickness of the liquid sample were studied. Efficiency of the temperature control system was tested.

As an example, the method was applied to the measurement of the acoustic properties of ethyl alcohol and water mixtures.

A. Influence of the bandwidth of the intermediate frequency filter

It is well known that reducing the bandwidth of a receiving system improves the signal-to-noise ratio and makes measurements more sensitive. In the current technique, the bandwidth narrowing effect is illustrated through the measurements of the variations of A_{111} and t_{111} as a function of the water thickness, for different bandwidths (RBW) of the intermediate frequency filter of the network analyzer. Four selectable bandwidths were chosen, ranging from 0.3 to 300 kHz, and experiments were performed at 21 MHz (the central frequency of a set of transducer units) and at a temperature of 18.6 °C. Results obtained using the 300- and 0.3-kHz bandwidths are plotted in Figs. 6(a) and 6(b) and Figs. 7(a) and 7(b), respectively, with the corresponding linear regression fits. A comparison of Fig. 6(a) with Fig. 7(a) clearly shows that narrowing the bandwidth causes a reduction in noise. The standard deviation of the fit decreases from 0.02 to 0.002 dB. However, the standard deviation of the fit is constant of the order of 3 ns irrespective of the selected bandwidth [cf. Figs. 6(b) and 7(b)].

The slopes of the linear regression fits give the absorption and the velocity of sound waves in water. Data of absorption and velocity and the parameters describing the linear regression, for various bandwidths, are summarized in Table I.

The measured absorption and velocity respectively decreases and increases as the detector bandwidth decreases. However, no clear explanation could be found to account for these evolutions. However, it is clear that the measured absorption and velocity using the narrowest detector bandwidth comes very close to the accepted literature values of absorption and velocity at 18.6 °C and 21 MHz, which are 0.1027 dB/mm and 1478 m/s, respectively. The small deviation from the accepted values is due to three factors: the misalignment and nonparallelism between the transducers and the accuracy of the z-stage mechanical displacement.

Great care must be taken in positioning the transducers. Imperfect alignment and parallelism produce different linear fits of the plots $(\Delta A_{111}, \Delta d_2)$ and $(\Delta t_{111}, \Delta d_2)$, providing different values of attenuation and velocity to be determined.
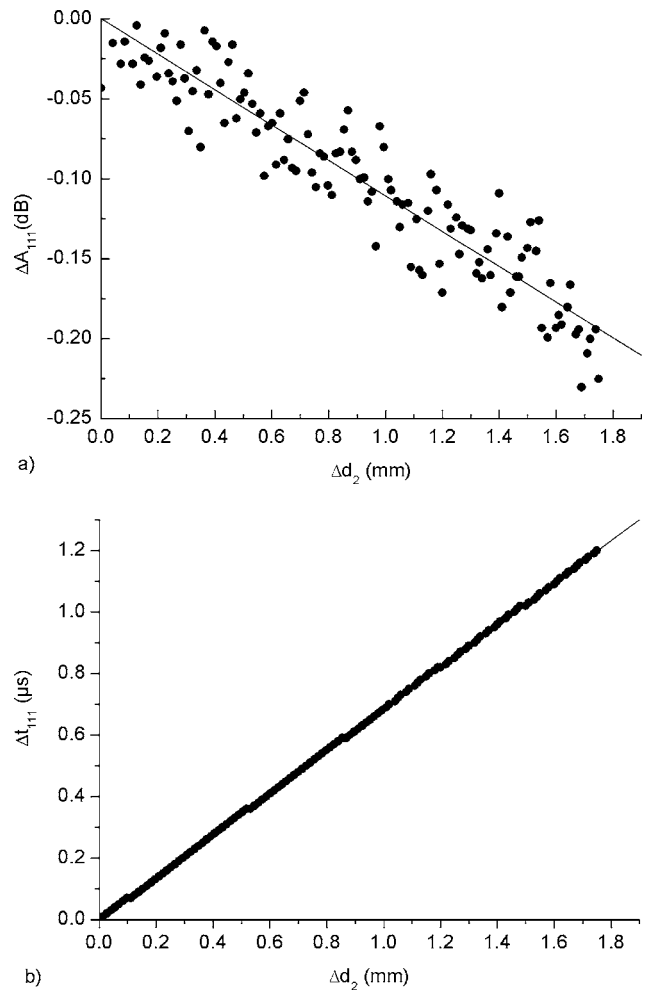
FIG. 6. Variations as a function of the liquid thickness Δd_2 (a) of the amplitude A_{111} of the first impulse and (b) of its transit time. The bandwidth (RBW) of the intermediate frequency filter and the temperature are set to 300 kHz and 18.6 °C, respectively.

B. Influence of the variation of thickness

The thickness variation of the liquid sample is of considerable importance for the accuracy of the measurement method. The larger the thickness, the better the accuracy. This is true as long as the signal-to-noise ratio of the measured parameter S_{21} is good. The attenuation and the velocity of sound waves were measured at various frequencies as a function of liquid thickness and compared with referenced data from Pinkerton[6] and Kroebel,[7] respectively. Relative errors on attenuation and velocity calculated from referenced and measured data are reported in Figs. 8(a) and 8(b), respectively. These measurements were performed using four pairs of transducer units and the 300-Hz bandwidth filter. Errors on both velocity and absorption decrease as the thickness variation increases up to 35 wavelengths. Above this thickness limit the errors become roughly constant to within 0.08% and 2.5%, respectively. Under this limit the error on attenuation decreases versus frequency since the attenuation, and ΔA_{111}, increase versus frequency. These errors are mainly attributed to the displacement accuracy of the z-axis stage.
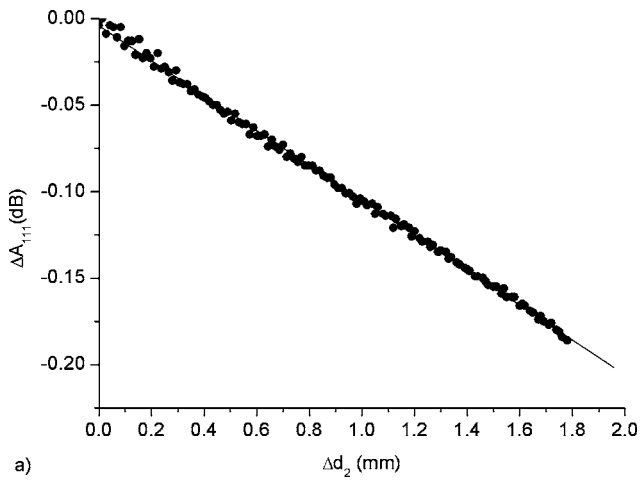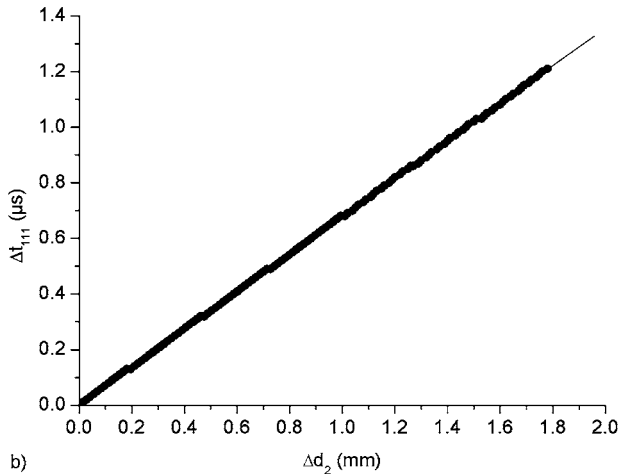
FIG. 7. Variations as a function of the liquid thickness Δd_2 (a) of the amplitude A_{111} of the first transmitted impulse and (b) of its transit time. The bandwidth (RBW) of the intermediate frequency filter and the temperature are set to 0.3 kHz and 18.6 °C, respectively.

FIG. 8. Relative errors compared with Pinkerton data of absorption (a) and Kroebel data of velocity (b) as a function of liquid thickness normalized to the wavelength. The temperature is set arbitrarily to 23 °C.

C. Influence of the temperature control system

The cavity filled by the liquid sample is open to air in a way that is dependent on the transducer spacing. Thus it is difficult to keep it at a constant temperature, particularly when the temperature is several degrees different from the room temperature, and when the distance between the transducers units is changed. The velocity and the attenuation of sound waves were measured at the arbitrarily frequency of 58 MHz while the temperature of the liquid was changed from 5 °C to 50 °C. The variation of the thickness of the

TABLE I. Absorption and velocity of sound wave in water at $T=18.6$ °C as a function of the bandwidth (RBW) of the intermediate frequency of the filter of the network analyzer.

RBW (kHz)	Absorption (dB/mm)	SD[a] (dB)	R[b]	Velocity. (m/s)	SD (ns)	R
300	0.1107	0.0204	0.9411	1460		0.9996
30	0.1021	0.0068	0.9917	1474	3	0.9995
3	0.1074	0.0030	0.9986	1478		1
0.3	0.1009	0.0020	0.9991	1477		1

[a]Standard deviation of the fit.
[b]Correlation coefficient.

liquid sample was set to 10 and 100 wavelengths, with a room temperature of 22 °C. Errors on both attenuation and velocity have been plotted as a function of the temperature difference between the room and the cavity in Figs. 9(a) and 9(b). Over a temperature range of about 30° around the room temperature, the control system is able to keep errors on attenuation and velocity lower than 5% and 0.4%, respectively.

D. Ultrasonic absorption and velocity in mixtures of ethyl alcohol and water

Storey[8] made measurements of absorption on mixtures of ethyl alcohol and water using a pulse mode method. For comparison, the same measurements were made using the current method and in addition velocity was also investigated.

The absorption coefficient and velocity were measured for mixtures with compositions ranging from pure water to 94% alcohol. For comparison with the data of Storey the quantity α/f^2 is plotted as a function of the composition of the liquid in Fig. 10(a). Velocity is plotted in Fig. 10(b).
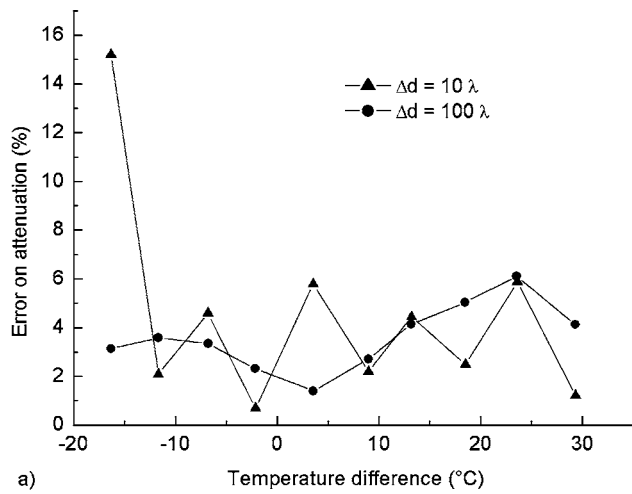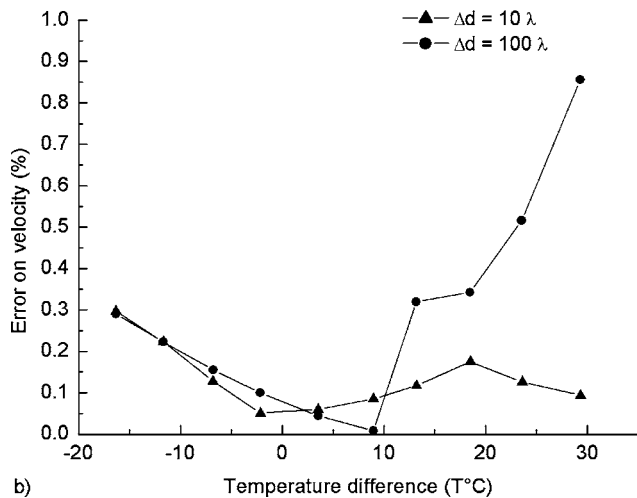
FIG. 9. Relative errors compared with Pinkerton data of absorption (a) and Kroebel data of velocity (b) as a function of the temperature difference between the liquid sample and the room temperature. The frequency is set arbitrarily to 58 MHz.

As it was found by Storey, the quantity α/f^2 reaches a maximum at a composition between 20% and 30% alcohol. For pure water (0%) and pure alcohol (100%) α/f^2 is independent of frequency. At and around the maximum α/f^2 depends on the frequency.

The additional velocity measurements show that the velocity exhibits a maximum at a different composition between 15% and 20% alcohol. Whatever the composition, the velocity may be considered as frequency independent and only dependent on temperature.

V. CONCLUSION

A method for measuring the velocity and attenuation of sound waves in liquids has been developed. The specificity of this method consists in driving the emitter transducer in a continuous wave mode and using a detector of narrow bandwidth. This enhances the signal-to-noise ratio and makes measurements very sensitive. Various experiments using distilled water have been performed indicating the efficiency of this method. However, the resulting data does not constitute
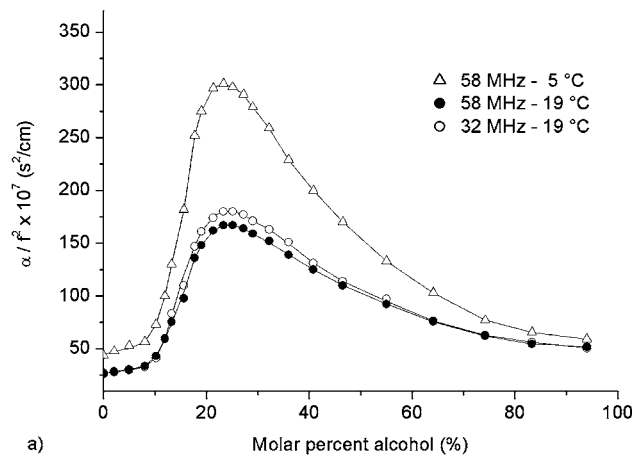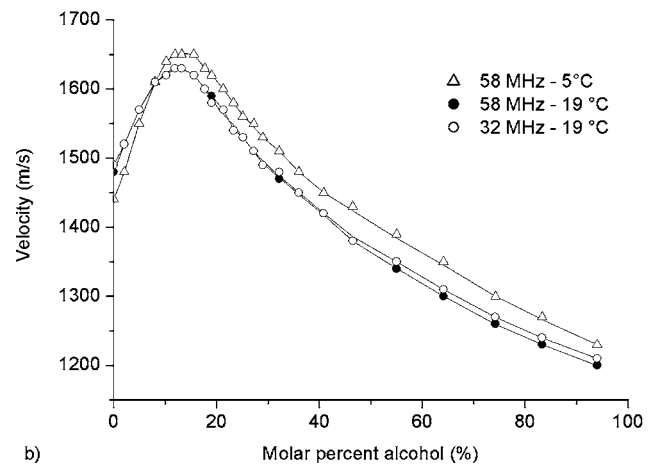
FIG. 10. Variations as a function of the composition of alcohol and water mixtures, of the absorption (a) and the velocity (b).

intrinsic characteristics of the measurement method capability since they are liquid dependent. Finally some materials were investigated for which absorption was already measured by another method and in addition velocity was measured.

[1] J. R. Pellam and J. K. J. Galt, "Ultrasonic propagation in liquids: I. Application of pulse technique to velocity and absorption measurements at 15 megacycles," J. Chem. Phys. **14**, 608 (1946).
[2] G. W. Pierce, "Piezoelectric crystal oscillators applied to the precision measurement of the velocity of sound in air and CO_2 at high frequencies," Proc. Am. Acad. Arts Sci. **60**, 271–302 (1925).
[3] M. A. Breazeale, J. H. Cantrell, Jr., and J. S. Heyman, "Ultrasonic wave velocity and attenuation measurements," in *Methods of Experimental Physics, Vol. 19, Ultrasonics*, edited by P. D. Edmonds (Academic, New York, 1981), Chap. 2, pp. 67–135.
[4] J. Kushibiki, N. Akashi, T. Sannomiya, N. Chubachi, and F. Dunn, "VHF/UHF range bioultrasonic spectroscopy system and method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1028–1039 (1995).
[5] Y. Deblock, P. Campistron, M. Lippert, and C. Bruneel, "Electrical characterization of plate piezoelectric transducer bonded to a finite substrate," J. Acoust. Soc. Am. **101**, 2681–2685 (2002).
[6] J. M. M. Pinkerton, "The absorption of ultrasonic waves in liquids and its relation to molecular constitution," Proc. Phys. Soc. London, Sect. B **62**, 129–141 (1949).
[7] W. Kroebel and K. H. Mahrt, "Recent result on absolute sound velocity measurements in pure water and sea water at atmospheric pressure," Acustica **35**, 154–164 (1976).
[8] L. R. O. Storey, "Ultrasonic absorption in mixtures of ethyl alcohol and water," Proc. Phys. Soc. London, Sect. B **65**, 943–950 (1952).

Polydisperse particle size characterization by ultrasonic attenuation spectroscopy for systems of diverse acoustic contrast in the large particle limit

Andreas Richter,[a] Frank Babick, and Siegfried Ripperger
Institut für Verfahrenstechnik und Umwelttechnik, TU Dresden, 01062 Dresden, Germany

Ultrasonic attenuation spectroscopy offers advantages over other spectrometric methods for the determination of suspension and emulsion size distributions. One advantage is the possibility of conducting measurements in highly concentrated or optically opaque dispersions. Models useable for the inversion of measured attenuation spectra to calculate particle size distributions are available and widely used, but their applied forms only allow for the characterization of particles smaller than 10 μm. In this paper, a methodology using the Faran model for elastic scatterers is examined in the region of micrometer-sized particles with respect to its suitability for the prediction of measured attenuation spectra and size distributions of various material combinations. All selected particle fractions and fluid materials were characterized independently from ultrasonic attenuation spectroscopy with respect to their material properties, size distributions, and shape. A comparison of measured and calculated attenuation spectra shows an acceptable agreement. The chosen methodology for particle sizing applications is further confirmed if a fit of the model to the measurement data is performed. In this approach, the solids volume fraction is treated as the only fit parameter. The findings indicate that the methodology is suitable for polydisperse particle size characterization for a wide range of acoustic contrast. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000773]

I. INTRODUCTION

Ultrasonic attenuation spectroscopy can be used to quantify dispersions with respect to particle or droplet size and concentration. It offers several advantages over other methods. First, because of favorable signal-to-noise ratios it can be applied to systems having high concentrations of dispersed material. Second, a wide range of particle or droplet sizes can be characterized, since several mechanisms contribute to the attenuation of sound. Some of the mechanisms are only valid in the region of nanometer-sized particles, while others dominate sound attenuation by suspensions with particles larger than 10 μm. This presence of more than one attenuation mechanism supports the distinction between attenuation spectra with respect to particle size. Finally, since the method is able to handle field applications where sample extraction or preparation (i.e., dilution) should be omitted, ultrasonic attenuation spectroscopy can be applied to online measurements.

Measurement devices and mathematical frameworks for the analysis of measured attenuation spectra to assess particle size information have been available for several years. Models for performing these analyses are well tested and commonly applied, although most of the established approaches are limited to the region of particle and droplet diameters below 10 μm, which is identified with the long wave regime

(for frequencies up to 100 MHz). The wave regime is determined by the wave number ka given by

$$ka = \frac{\pi d}{c_F}f = \frac{2\pi a}{c_F}f, \qquad (1)$$

where d is the particle diameter (or a the radius), c_F is the velocity of sound in the continuous phase, and f is the frequency of the sound wave. The long wavelength regime is characterized by $ka \ll 1$. Both of the models currently used for the analysis of attenuation results, the ECAH theory[1,2] and the coupled phase/cell models,[3,4] are only applied in the limit of this wave regime. Outside the long wavelength limit, other models for the mathematical description of sound attenuation exist.[5,6] One of these models, the Faran approach,[5] was developed to allow for the calculation of elastic scattering in the intermediate wave regime ($ka \approx 1$) for elastic immovable spheres, and attenuation measurements have been shown to be in agreement with the model for $ka \gtrsim 1$ in polystyrene-water suspensions with narrow distributions of particle size.[7,8] In these studies, model predictions were calculated under the implicit assumption that polydispersity can be expressed by linear superposition of the attenuation contribution of each size class. The observed good agreement between predictions and measured size and frequency dependency of attenuation support the validity of the assumption. This is a prerequisite for particle sizing applications, because it implies that the contribution to attenuation from different

[a]Electronic mail: ar5@rcs.urz.tu-dresden.de

size fractions could potentially be identified and separated. This study tries to present a combined approach for the calculation of size distributions from measured attenuation data in the large particle limit and relies on the superposition principle implicitly assumed by other works.

The purpose of this paper is twofold. First, the model framework for particle size characterization by ultrasonic attenuation spectroscopy in the intermediate wave regime is detailed. Aspects relevant to particle size extraction are outlined. Second, the validity of the combined approach is examined by comparing measured results obtained with an unaltered commercially available attenuation spectrometer to model predictions. The measurements were carried out for a wide range of material combinations to detail the influence of material properties on the validity of the combined approach. The discussion of the approach covers the agreement of measured attenuation spectra with prediction and the similarity between the polydisperse size distributions calculated from the very same measured ultrasonic attenuation and results obtained with an independent measurement technique.

II. ULTRASONIC ATTENUATION MODELING FOR PARTICLE SIZE CHARACTERIZATION

The theory developed by Epstein and Carhart,[1] which was later expanded by Allegra and Hawley[2] (ECAH theory), is based upon the description of the scattered thermal and hydrodynamic fields within and surrounding a single spherical particle. The model development is carried out by defining an infinite sum of scattering coefficients. These coefficients are defined by the interface between particulate and continuous phase. Boundary conditions at this interface are balances of impulse and heat exchange and introduce material properties to the model. The solution is given as a location-dependent pressure amplitude at large distances from the particle, from which expressions of attenuation in the far field can be derived.

In its original form, the ECAH theory is universal with respect to particle size and wavelength. Explicit formulations for the attenuation coefficient can be stated for the long wavelength or small particle size limit. Since the ECAH theory only considers single particles, the model is valid for low concentrations, for which the interaction of hydrodynamic, thermal, and elastic scattering fields between neighboring particles can be neglected (single scattering limit).

In contrast to the ECAH theory, the coupled-phase models[3,4] are generally able to correctly incorporate the interaction of thermal and hydrodynamic fields, but do not account for elastic scattering. Models of this type are based on formally combining particulate matter into a single phase which interacts with the continuous phase. For both of these abstract phases, balances of continuity, energy, and force can be stated and combined into a linear set of equations, under the assumption that time-dependent influences are negligible. The connection between the formal treatment of two separate phases and true dispersive state is usually achieved by the so-called "cell models." These models consider each particle and the surrounding medium as a single cell, with volume

fractions of dispersed and continuous phase chosen accord <>ing to the dispersion volume distribution. Consequently, the evaluation of acoustic field disturbances caused by the dispersion is carried out for a single average cell. Interactions between cells are incorporated by selecting suitable boundary conditions. Despite this seemingly more abstract approach, the coupled phase models are applied to the particle size characterization of nanosized suspensions.[4,9,10] The model is often preferred because of the more accurate reflection of the particle concentration influence even outside the single-scattering limit. However, because of the underlying abstract model formalism, only thermal and viscoinertial effects along with absorption within the phases are considered in coupled phase/cell models, which limits their applicability to the long wavelength region ($ka \ll 1$).

As stated above, the ECAH theory is based on the interaction of scattered, viscoinertial, and thermal fields with the surrounding medium. In contrast, if only elastic scattering (diffraction, refraction, and reflection) is assumed to govern the wave propagation, another type of model can be derived. This approach is relevant, if elastic scattering dominates the propagation of sound, which is the case in the intermediate wave regime ($ka \approx 1$). Models describing elastic scattering were derived by Faran[5] and Anderson.[6] Again, a set of equations derived from boundary conditions (differing in their nature from those in the ECAH theory) leads to the definition of scattering coefficients A_n. Depending on the properties of the particle-fluid combination, different expressions for A_n can be developed. The approach proposed by Faran[5] is suitable for the case of solid, elastic, immovable spheres. The scattering coefficients are expressed as[11]

$$A_n = - \sin \eta_n \cdot e^{-i\eta_n} = \frac{-i \tan(\eta_n)}{1 + i \tan(\eta_n)} \tag{2}$$

with

$$\tan \eta_n = \tan \delta_n(ka) \frac{\tan \alpha_n(ka) - (\rho_F/\rho_P)\tan \zeta_n(k_L a, k_T a)}{\tan \beta_n(ka) - (\rho_F/\rho_P)\tan \zeta_n(k_L a, k_T a)}, \tag{3}$$

$$\tan \delta_n(ka) = -j_n(ka)/n_n(ka), \tag{4}$$

$$\tan \alpha_n(ka) = -ka \cdot j'_n(ka)/j_n(ka), \tag{5}$$

$$\tan \beta_n(ka) = -ka \cdot n'_n(ka)/n_n(ka), \tag{6}$$

$$\tan \zeta_n(k_L a, k_T a) = -\frac{(k_T a)^2}{2}\left[\frac{\tan \alpha_n(k_L a)}{1 + \tan \alpha_n(k_L a)} - \frac{n^2 + n}{n^2 + n - 1 - (k_T a)^2/2 + \tan \alpha_n(k_T a)}\right]\Bigg/\left[\frac{n^2 + n - (k_T a)^2/2 + 2\tan \alpha_n(k_L a)}{1 + \tan \alpha_n(k_L a)}\right.$$

$$\left.-\frac{(n^2 + n)(1 + \tan \alpha_n(k_T a))}{n^2 + n - 1 - (k_T a)^2/2 + \tan \alpha_n(k_T a)}\right], \tag{7}$$

where the expressions of the wave numbers $k_L a$, $k_T a$, and ka are given by $k_L a = \pi d f / c_L$, $k_T a = \pi d f / c_T$, and $ka = \pi d f / c_F$ with c_L, c_T, and c_F denoting the velocity of the compressional (longitudinal) and shear (transversal) wave in the particle, and the compressional wave in the continuous phase, respectively. The quantities ρ_P and ρ_F are the densities of particle and continuous phase. The series j_n, n_n and j'_n, n'_n represent spherical Bessel and Neumann functions and their first-order derivatives. It is possible to deduce the sound velocities from the elastic properties of the medium, namely the modulus of elasticity E and Poisson number ν:

$$c_L = \sqrt{\frac{E(1 - \nu)}{\rho(1 + \nu)(1 - 2\nu)}}, \tag{8}$$

$$c_T = \sqrt{\frac{E}{2\rho(1 + \nu)}}. \tag{9}$$

Thus, the influence of material on the scattering coefficient and the scattering behavior for a known frequency and particle size can be expressed by the densities and the elastic properties. It implies that the knowledge of these properties is sufficient to determine the acoustical attenuation behavior of any particle-fluid system in the region of $ka \approx 1$.

If the dispersed system is not limited to wave numbers $ka \approx 1$ or above, viscoinertial and thermal attenuation need to be considered. Hay and Mercer[11] expanded the Faran approach by including viscoinertial effects in the same way as they appear in the ECAH theory, but the resulting scattering coefficients are significantly more complex.

In analogy to the ECAH theory, a term for the attenuation coefficient α can be stated which relates the scattering coefficients A_n to an amplitude or intensity decrease of a wave passing through a dispersed medium. Hay and Mercer[11] used the total absorption cross section, i.e., the sum of elastic scattering and intrinsic absorption by particles, defined by Morse and Ingard[12] in combination with Faran's coefficients A_n and state the following expression for the total attenuation coefficient for the dispersed phase:

$$\alpha = -\frac{3\phi}{2(ka)^2 a}\sum_{n=0}^{\infty}(2n + 1) \cdot \mathrm{Re}(A_n). \tag{10}$$

In Eq. (10) ϕ denotes the solids volume fraction.

III. ELASTIC SCATTERING BY POLYDISPERSE PARTICLE SYSTEMS

The expression following from the Faran model [Eq. (10)] describes the dependency of attenuation by elastic scattering on the wave number ka. This implies a dependency on both the particle diameter d (or radius a) and the frequency f. In its original form, Eq. (10) considers a monosized particle collective. However, in most cases of particle characterization, a polydisperse state of the system must be assumed.

Polydispersity is defined as the presence of more than one particle size.[13,14] This includes both monomodal and multimodal particle systems. The polydisperse state has been included in the discussion of the scattering formulation before.[7,8,15,16] In these studies the size distribution is treated as a number density or histogram function, which is a representation of the number fraction or particle concentration in each particle size class. For the following approach, the results are presented in a manner not fundamentally different from previous work, but in a way that separates the size distribution function from the particle volume concentration ϕ by using a particle volume based size distribution function.[14] This is an advisable form if size distribution and particle concentration should be treated separately.

Generally, a polydisperse state can be described by a vector of discrete particle sizes. The term "size" can be thought to be a more general description of a property of the particles, that is, a length measure. In the case of irregularly shaped particles, an average or equivalent diameter parameter is often used (i.e., diameter of the sphere having the same volume). Since this study primarily deals with spherical particles, the following derivation uses the sphere diameter as a measure of the particle size. Consider the vector $\mathbf{d} = (d_1, d_2, \ldots, d_v)$, where d_j represents a class-averaged particle diameter. Then $\mathbf{\Delta Q_3} = (\Delta Q_{3_1}, \Delta Q_{3_2}, \ldots, \Delta Q_{3_v})$ can be taken to be the volume-based size distribution vector describing the fraction of particles within each class.[14] Each value ΔQ_{3_j} defines the fraction of the overall particle volume belonging to the size class with diameter d_j. The associated requirement $\Sigma_v \Delta Q_{3_j} = 1$ emphasizes that the size distribution vector comprises a complete set of classes for a given polydisperse system. The index 3 denotes the fact that this is a volume-based treatment, as opposed to particle number-, length-, or area-based distributions (index 0, 1, 2).[14] $Q_3(d_j)$ is the cumulative distribution function. Its values $Q_3(d_j) = Q_{3_j}$ define the fraction of particles having a diameter smaller than d_j. Based on this notation the definition of the median particle diameter d_{50} can be expressed as $Q_3(d_{50}) = 0.5$, that is, 50% (by volume) of the particles in the distribution have a diameter smaller than d_{50}. The particle volume concentration in each size fraction j is $\phi \Delta Q_{3_j}$. The sum over these values again gives the overall volume concentration ϕ.

This work is based on the single scattering theory, where it is implicitly defined that the volume concentration and the attenuation are linearly related. Consequently, the attenuation coefficient α is found through superposition of the attenuation contribution by all size classes, for a continuous size distribution function and a discrete size distribution function according to Eqs. (11) and (12), respectively:

$$\alpha = -\frac{3\phi}{2}\int_0^{\infty}\frac{1}{(ka)^2 a}\sum_{n=0}^{\infty}(2n + 1)\mathrm{Re}(A_n)dQ_3, \tag{11}$$

TABLE I. Material properties of the studied particulate substances at 25 °C.

Material	E (GPa)	ν	ρ (g/cm^3)	c_L (m/s)	c_T (m/s)	Source
Glass	68.9	0.2[a]	2.51	5523	3382	Manufacturer
Aluminum	71	0.34	2.7	6362	3132	Ref. 24
Tin	41	0.33	7.3	2884.7	1453	Ref. 25
Polystyrene	3.95	0.34	1.06	2380	1180	Ref. 8

[a]Reference 26.

$$\alpha = -\frac{3\phi}{2}\sum_{j=1}^{\nu}\frac{1}{(ka_j)^2 a_j}\sum_{n=0}^{\infty}(2n+1)\mathrm{Re}(A_n)\Delta Q_{3_j}. \qquad (12)$$

Hall[7] and Mobley[8] used a similar polydisperse model development for demonstrating that measurements in the ka range up to 7 can be described by Faran's approach to a high degree of accuracy. The authors derived the single scattering model from the Waterman-Truell multiple scattering formulation[17] in the domain of weak scattering density, which resulted in an expression equivalent to Eq. (12).

For this study, another consideration regarding the representation of the attenuation spectrum is applied. The quantity α is interpreted as a vectorial representation of attenuation. Each element of α corresponds to one of the values in a vector of discrete frequencies $\mathbf{f} = (f_1, f_2, \ldots, f_u)^T$. This assumption implicitly used by other works[7,8,15,16] is based on the application of ultrasonics techniques. It makes use of the fact that measurements obtained from digitized broadband signals or by a step-through process using narrow-band signals result in values at discrete frequencies with finite spacing. This consideration is advisable if model prediction spectra are to be compared to measurements.

Based upon the two considerations implicitly made by other authors (superposition due to polydispersity and vectorial representation of the attenuation spectrum), a matrix version of Eq. (10) can be stated. Attenuation by elastic scattering for a polydisperse particle collective and a vector of frequencies is expressed as

$$\boldsymbol{\alpha} = \phi \cdot \mathbf{K} \cdot \Delta \mathbf{Q}_3^T \qquad (13)$$

where $\mathbf{K}$ is a kernel matrix with elements given by

$$K_{i,j} = \frac{3}{2(ka_{i,j})^2 a_j}\cdot\sum_{n=1}^{n_{max}}(2n+1)\mathrm{Re}(A_{n_{i,j}}). \qquad (14)$$

The matrix $\mathbf{K}$ depends on the elements of the wave number matrix $\mathbf{ka}$, the vector of particle diameters $\mathbf{a} = \mathbf{d}/2$, and the matrix of scattering coefficients $\mathbf{A_n}$. The matrix $\mathbf{ka}$ is defined by $\mathbf{d}$ and $\mathbf{f}$ as

$$\mathbf{ka} = \mathbf{f}\cdot\mathbf{d}\frac{\pi}{c_F} = \begin{bmatrix} f_1 d_1 & \cdots & f_1 d_\nu \\ \vdots & & \vdots \\ f_u d_1 & \cdots & f_u d_\nu \end{bmatrix}\frac{\pi}{c_F} \qquad (15)$$

and thus the scattering coefficients A_n can be expressed as

$$\mathbf{A_n} = \begin{bmatrix} A_{n_{1,1}} & \cdots & A_{n_{1,\nu}} \\ \vdots & & \vdots \\ A_{n_{u,1}} & \cdots & A_{n_{u,\nu}} \end{bmatrix}. \qquad (16)$$

Each of the values $A_{n_{i,j}} = A_n(ka_{i,j})$ can be calculated according to Eqs. (2)–(7).

The formation of the matrix is the central principle for determining polydisperse particle size information from attenuation measurements.[9,18] This approach offers the possibility of obtaining a particle size distribution vector $\Delta \mathbf{Q}_3$ from a measured attenuation spectrum $\boldsymbol{\alpha}$ as the solution to the linear system of equations set up by Eq. (13). Besides numerical inversion of $\mathbf{K}$, one possible strategy for finding such a solution vector consists of estimating a best-fit representation $\Delta \mathbf{Q}_{3_{fit}}$, under the assumption that $\Delta \mathbf{Q}_{3_{fit}}$ conforms to a predetermined function type. Naturally, the selected function type greatly impacts the fit quality. In the case of monomodal continuous size distributions, a common choice is a log-normal size distribution.[19–21] The density distribution function $q_3 = dQ_3/dd$ (or $q_3 = \Delta Q_3/\Delta d$ in discretized form) for this function type is

$$q_3(d) = \frac{1}{\sigma_{ln}\sqrt{2\pi}d}\exp\left[-\frac{1}{2}\left(\frac{\ln(d/d_{50})}{\sigma_{ln}}\right)^2\right]. \qquad (17)$$

In this case, the task of determining the best-fit representation $\Delta \mathbf{Q}_{3_{fit}}$ (based on a log-normal size distribution) is reduced to estimating the defining parameters, namely median diameter d_{50} and standard deviation σ_{ln}.

If measured attenuation spectra $\boldsymbol{\alpha}_{\mathbf{meas}}$ of suspensions are to be compared to model predictions $\boldsymbol{\alpha}$, it is necessary to account for the absorption of sound within the continuous phase. When assuming linear superposition, the total suspension attenuation $\boldsymbol{\alpha}_{\mathbf{tot}}$ can be calculated by adding the attenuation in the pure continuous phase $\boldsymbol{\alpha}_{\mathbf{F}}$, weighted according to the solids volume fraction, to the model predictions for the dispersed phase $\boldsymbol{\alpha}$:

$$\boldsymbol{\alpha}_{\mathbf{tot}} = \boldsymbol{\alpha} + (1-\phi)\cdot\boldsymbol{\alpha}_{\mathbf{F}}. \qquad (18)$$

IV. EXPERIMENTAL MATERIAL AND SETUP

Based upon the material properties defining elastic scattering according to Eqs. (2)–(9) (density, modulus of elasticity, and Poisson number), several particle substances were selected. Among the materials chosen, glass represents a class of materials having properties similar to those of minerals. Two fractions of glass particles manufactured by Potters-Ballotini GmbH Kirchheimbolanden, one with a mean diameter of 60 μm and another with a mean of 200 μm, were available for analysis.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Richter et al.: Large particle ultrasonic attenuation spectroscopy   1397
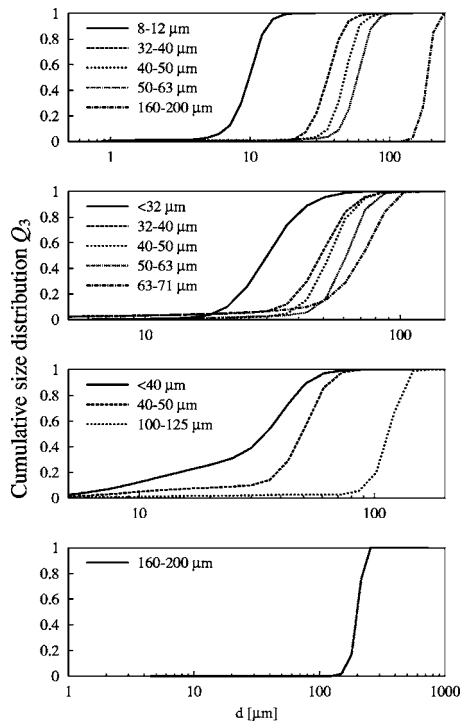
FIG. 1. Cumulative size distributions Q_3 as functions of particle diameter d of particulate materials used in the attenuation and particle size comparison. From top to bottom, the size distributions show glass, aluminum, tin, and polystyrene.

Aluminum and tin were selected as examples of metallic particulate matter and obtained from Nanoval GmbH, Berlin, and TU Bremen, respectively. Finally, a sample of polystyrene particles (Duke Scientific) was chosen to represent the class of polymer particles. Table I lists material properties of all particulate materials that underwent analysis.

All particle samples were split into fractions of narrow width by a combination of sieving and air separation. The aim of this separation step was to create well-defined particle size distributions and to cover a broad range of values for ka. The applied sieve mesh sizes define and denote the diameter fractions (Table IV). To obtain independent particle size distributions of the resulting fractions, a laser diffraction spectrometer (HELOS, Sympatec GmbH) was utilized. Reference size distributions are detailed in Fig. 1. Furthermore, each fraction was evaluated by scanning electron (SEM) and light microscopy to confirm particle shape and surface properties (i.e., roughness). Although materials with spherical shape were selected, up to 10% of all fractions contained aspherically shaped particles.

Most particle fractions were examined in more than one continuous phase fluid. During this study, water, ethanol, and a NaCl solution (20% NaCl by weight) were used in the preparation of the suspensions. Using more than one continuous phase fluid further diversifies the material combinations and allows for an evaluation of the material property sensitivity. Additionally, this approach allows for conclusions to be drawn regarding the impact of unfavorable properties such as surface roughness and particle shape. These effects should contribute to the same extent in different particle-fluid combinations. Table II lists elastic properties of the continuous phase materials used in this study.

TABLE II. Material properties of the studied fluids at 25 °C.

Material	ρ (g/cm^3)	c_F (m/s)	Source
Water	0.997[a]	1497	Ref. 27
Ethanol	0.79	1150	[a]
NaCl solution (20% b.w.)	1.15	1676	[a]

[a]Own measurement.

An acoustic spectrometer model DT-1200 (Dispersion Technology[22]) was utilized for the measurement of experimental attenuation spectra of the material combinations. This device allows the measurement of attenuation at up to 18 specified discrete frequency values in the range from 3 to 99 MHz. Both the number and the spacing (i.e., linear, logarithmic) of the frequencies can be adjusted. For the present work, the full range and number of frequencies was used and a logarithmic distribution selected [$\mathbf{f} = (3, 3.69, 4.53, \ldots, 99)^T$ MHz, $\ln(f_{i+1}/f_i) = 0.2057$]. The measurement chamber of the device contains a fixed and a movable broad-band transducer, allowing for settings of different gap widths between sound emitter and receiver. For each of the frequencies selected for the measurement, the device cycles through a defined number of gap widths between the emitter and the receiver. At each setting (frequency and gap width), a pulsed signal is generated. First, a frequency synthesizer generates a continuous wave signal. A field-programmable gate array (FPGA) controlled by a host computer then generates a transmit gate. This transmit gate is used by a gated amplifier to form a pulsed RF signal from the continuous wave signal. A power amplifier increases the peak power and the resulting signal is passed to the transducer.

Signal reception is carried out by measuring wave intensity during time frames corresponding to the mean sound travel time. This is done in the following manner: The detected RF input signal is passed to a wideband RF amplifier having a programmable gain set by the FPGA, which allows only pulses arriving within a predetermined time interval to pass. The resulting signal is connected to a quadrature mixer which demodulates the received signal, providing analog pulses proportional to the in-phase and quadrature components of the received RF pulse, respectively. After a filter step and A/D conversion, the digitized peak amplitudes of the quadrature signals upon an A/D strobe command are passed through a digital signal processor that accumulates

TABLE III. Tabulated reference and measured attenuation data α/f^2 for pure liquids having units of 10^{-3} dB/(MHz^2·cm), at 25 °C. Values were obtained at a frequency of 99 MHz and list average and standard deviation of repeat measurements.

Material	Reference value[26]	Measured value
Distilled water	1.87	1.82±0.05
Ethanol	4.43	4.43±0.04
Methanol	2.78	2.50±0.12
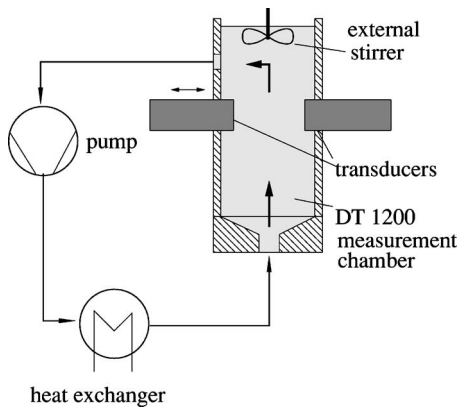
FIG. 2. Experimental setup with the acoustic attenuation spectrometer DT-1200. Arrows indicate flow direction.

data from a large number of pulses and generates statistical data on a given pulse set, and finally the data are routed to the computer bus.

The intermediate result of the measurement at different gap widths and frequencies is a matrix of signal intensity and its fluctuation. Statistical analysis of the signal fluctuation at each measurement point (frequency, gap width) and an averaging procedure are applied to extract the signal strength values. These values form an attenuation vector which corresponds to the discrete frequencies. The resulting attenuation values are divided by the frequency at which they were obtained and stored in a database ($\alpha_{\mathrm{meas}}/\mathbf{f}$ having units of dB/cm/MHz). These database values were employed in the subsequent study. During the measurement, the factory settings of the device remained unchanged. Neither signal conditioning nor our own processing of sound signal data was applied.

To confirm that attenuation measurements in the suspensions are reliable and reproducible, measurements of attenuation in pure liquids were performed and compared to established tabulated data. Table III shows good agreement between the tabulated and the measured data and can be seen as evidence that the device is capable of reproducing known results and is an appropriate tool for this study. Nevertheless, device specifics such as bandwidths of the individual signals or signal wave forms were not accessible during the study and must be seen as a source of influence when discussing attenuation results.

Because settling velocities in suspensions of particles larger than 10 μm are comparably high, precautions had to be taken to prevent sedimentation and to ensure constant concentration levels in the measurement chamber. This was achieved by continuously circulating the suspensions upwards through the measurement chamber past the sensor. Additionally, since material properties such as density and modulus of elasticity are temperature-dependent values, a heat exchanger was installed. Using this device, the suspension temperature was kept at a constant value of 25 °C. A schematic of the setup is depicted in Fig. 2.

The attenuation spectrum of the pure continuous phase was measured at the beginning of each measurement campaign. To this same sample of the continuous phase, the particle system was added. Solids volume fractions ϕ in the

measurement compartment were adjusted by adding small amounts of particle material until the desired concentration was obtained. The particle concentration was measured by extracting samples of 1 ml via pipette, which were analyzed for their solids weight fraction using a thermobalance (Mettler Toledo HG53 Halogen Moisture Analyzer). After the solids fraction had been established at the desired value by repeated addition of particle material and subsequent concentration measurement, the suspension attenuation measurement runs were started. Each examination of material combinations with each of the diameter fractions was repeated at least four times, in order to evaluate variations. For a comparison of the resulting attenuation spectra, mean and standard deviation of the measurement repetition were determined. At each of these four repeat measurements one concentration sample was taken. Although the concentration was carefully adjusted, fluctuations of the resulting solids volume concentrations ϕ occurred in all cases.

V. COMPARISON OF ATTENUATION SPECTRA MEASUREMENT TO PREDICTION

Measured attenuation spectra α_{meas} were compared with the corresponding calculations. Only the mean and standard deviation of the attenuation measurements along with the mean nominal concentration values were used for the evaluation of the model framework. Predictions of the sound attenuation spectrum were obtained by calculating α_{tot} [Eq. (18)]. The fraction of the attenuation caused by elastic scattering α was computed according to Eq. (13), using particle size distributions ΔQ_3 determined from laser diffraction spectroscopy (Fig. 1) and material properties listed in Tables I and II. These calculations rely on the scattering coefficient range from A_1 to A_{40}, which was found to be an optimal compromise between calculation time and exactness. Higher coefficient orders did not result in an accuracy increase greater than the round-off errors introduced by the numerical computation.

Figure 3 compares the measured attenuation spectra obtained for glass particles in water at volume fractions of 1% to the corresponding model predictions. All graphs display attenuation as $\alpha/\mathbf{f}$ versus ka. For these and all following graphical comparisons of attenuation spectra, the distribution median diameter d_{50} obtained from the laser diffraction measurements [the value of d for which $Q_3(d)=0.5$, see Fig. 1] was used in the calculation of the abscissa data. The division by $\mathbf{f}$ accentuates the oscillations in the region of $ka>3$.

When comparing attenuation spectra for different fractions of the same material (Fig. 3), it appears that function extremum locations coincide, and that the shape of the curve is reproduced. The presentation of attenuation as a function of ka emphasizes the normalization inherent in the use of the wave number. Deviations should be attributed to the different relative size distribution span. For size distributions comprised of a broader diameter range, attenuation spectra appear smoother due to superposition of attenuation by different portions of the particle size spectrum.

Judging by the attenuation functions, the agreement between model and measured results is adequate. For all glass particle fractions, the shape of the measured curve is re-
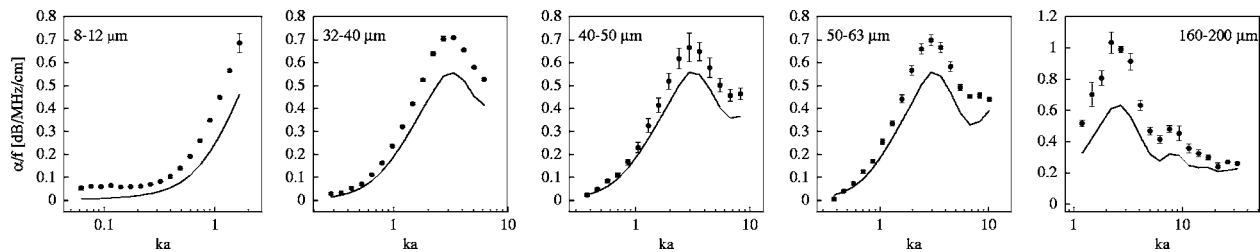
FIG. 3. Measured (dots) and calculated (lines) ultrasonic attenuation spectra of glass fractions in water, $\phi=1$ Vol-%. The wave number ka was calculated using the mean diameter d_{50} of the laser diffraction results and the frequency vector **f**.

flected by the calculations (see Fig. 3). Similar results for all other examined material combinations are shown in Fig. 4. All graphs emphasize the effect of normalization introduced by the wave-number-dependent function (shape and magnitude of attenuation spectra of different fractions match regardless of material combinations).

Despite this good agreement, size distribution functions of the smaller fractions measured by laser diffraction show the presence of particles with diameters smaller than 5 μm. This size region exhibits attenuation by viscoinertial interac-

tion between continuous and dispersed phase. This can be observed in the corresponding attenuation spectra by declining slopes with increasing values of ka in the region of $ka=0.1$. This effect is apparent for glass particles in ethanol or in NaCl solution [Figs. 4(a) and 4(b)]. The prediction curves do not exhibit this increase, since the mechanism of viscoinertial attenuation is not included in the calculations.

Two device-dependent factors may have had an influence on the measured results. Both relate to the discrete step-through-frequency method of acquiring the data. First, the
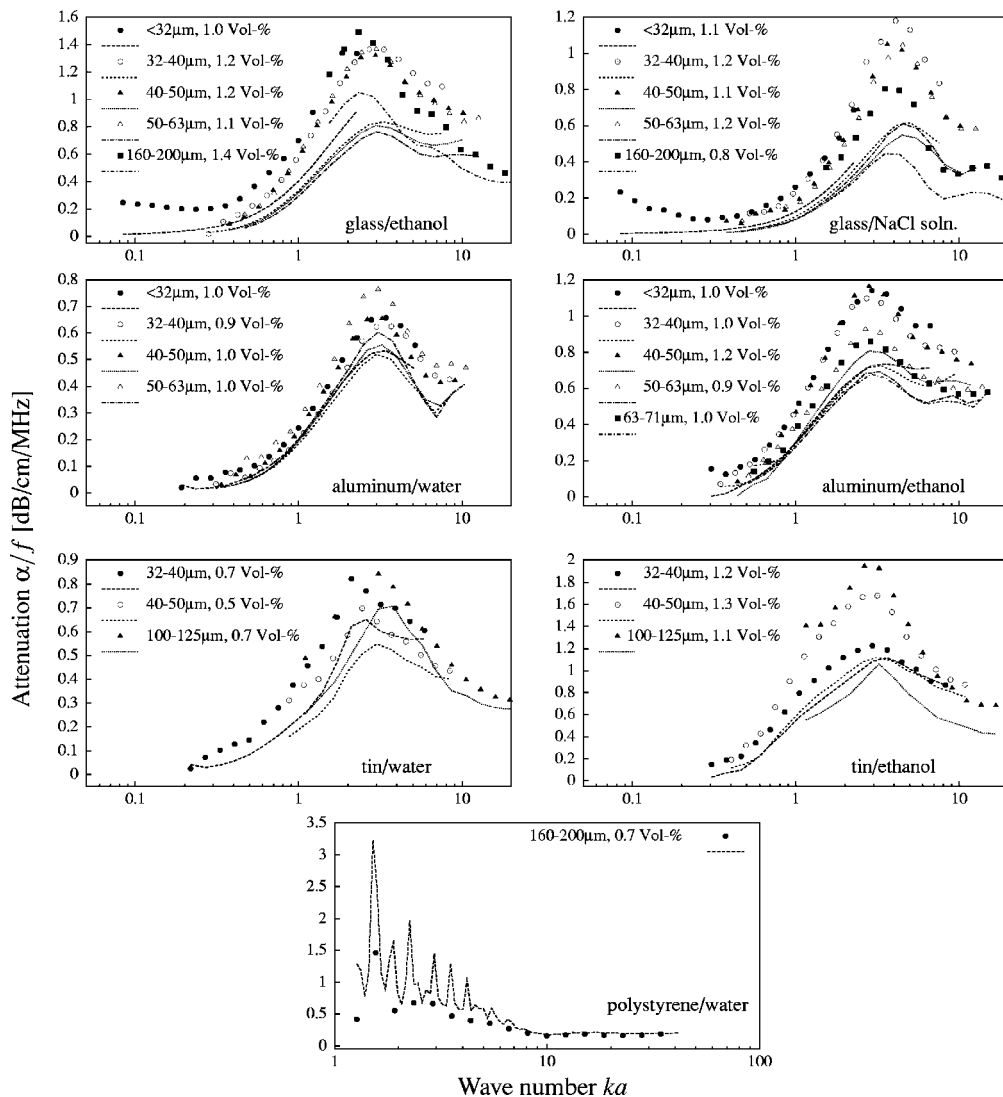
FIG. 4. Measured (dots) and calculated (lines) ultrasonic attenuation spectra of various particle-fluid combinations, $\phi=1$ Vol-%. The wave number ka was calculated using the mean diameter d_{50} of the laser diffraction results and the frequency vector **f**.

Richter *et al.*: Large particle ultrasonic attenuation spectroscopy

TABLE IV. The results of the best-fit approximation of the measured attenuation to the polydisperse attenuation model for all examined material combinations. σ^2 and s^2 are the coefficients of variation of the measurement and of measurement and fit, respectively.

Material combination	Diameter fraction (μm)	ϕ nominal (vol-%)	ϕ best fit (vol-%)	σ^2 (dB/cm/MHz)^2	s^2 (dB/cm/MHz)^2
Glass-ethanol	8–12	1.02	1.39	1.57×10^{-3}	1.92×10^{-2}
	32–40	1.23	1.72	7.05×10^{-4}	1.17×10^{-2}
	40–50	1.2	1.69	4.72×10^{-4}	6.95×10^{-3}
	50–63	1.12	1.77	1.12×10^{-4}	8.17×10^{-3}
	160–200	1.42	1.96	1.54×10^{-3}	6.65×10^{-4}
Glass-NaCl solution	8–12	1.09	1.48	1.94×10^{-4}	9.21×10^{-3}
	32–40	1.23	1.95	2.75×10^{-4}	1.44×10^{-2}
	40–50	1.06	1.7	3.76×10^{-4}	6.96×10^{-3}
	50–63	1.16	1.68	3.21×10^{-4}	1.05×10^{-2}
	160–200	0.75	1.34	5.43×10^{-4}	8.66×10^{-4}
Aluminum-water	<32	0.98	0.94	4.61×10^{-5}	5.36×10^{-3}
	32–40	0.94	0.92	5.35×10^{-5}	9.44×10^{-3}
	40–50	101	0.99	4.75×10^{-5}	4.07×10^{-3}
	50–63	1.01	1.18	1.62×10^{-4}	8.59×10^{-3}
	63–71	2.04	3.34	1.22×10^{-3}	5.56×10^{-2}
Aluminum-ethanol	<32	1.03	1.34	3.14×10^{-4}	7.03×10^{-3}
	32–40	1.03	1.35	3.73×10^{-4}	3.84×10^{-3}
	40–50	1.22	1.45	1.37×10^{-4}	5.81×10^{-3}
	50–63	0.93	1.12	1.14×10^{-4}	3.85×10^{-3}
	63–71	1.01	1.06	1.14×10^{-4}	5.02×10^{-3}
Tin-water	32–40	0.69	0.83	1.94×10^{-4}	9.18×10^{-3}
	40–40	0.55	0.45	7.67×10^{-4}	4.12×10^{-2}
	100–125	0.71	0.9	6.63×10^{-4}	1.41×10^{-2}
Tin-ethanol	32–40	1.19	1.35	5.41×10^{-4}	7.25×10^{-3}
	40–50	1.3	1.75	1.52×10^{-3}	3.65×10^{-2}
	100–125	1.06	1.83	2.77×10^{-3}	1.36×10^{-1}
Polystyrene-water	160–200	0.74	0.29	4.46×10^{-4}	2.71×10^{-2}

spectral coverage may have been too wide to reproduce attenuation peaks which are narrow in frequency, leading to peaks being missed. Furthermore, the width of the generated pulse in time may have been too large. If the frequency bandwidth of the ultrasonic pulse is as wide or wider than the frequency width of the peak, it may not be properly resolved.

The result obtained with the polystyrene-water system exposes some limits of the measurement, even though the general shape of the curve is in agreement with the prediction. For this sample, the calculated and the measured spectra have the same order of magnitude. For higher wave numbers these spectra even coincide. However, a number of predicted oscillations in the low ka range are not reproduced by the measurement results. The possible reasons for this result are threefold. First, a detailed inspection of the calculations reveals that the frequency spacing employed in the study does not allow for a finer resolution of the attenuation function. Second, size distributions based on laser diffraction can be expected to span a greater width than actually present. An indication of this (algorithm-induced) result can be found

from a comparison of the sieve mesh sizes used for fractioning (Table IV, column "diameter fraction") to the laser diffraction size distributions (Fig. 1). The laser diffraction results extend to diameters well beyond the sieve mesh size. Finally, the tone burst created by the device at each frequency may be too wide to resolve the peaks even though its center frequency is in line with the peak. This can be problematic when the peaks of the attenuation curve are narrow, which is the case for the polystyrene calculations.

VI. DISCREPANCY BETWEEN MODEL AND MEASUREMENT

As stated above, the agreement between model prediction and measurement is good for all material combinations. Nonetheless, there appears to be a systematic trend to higher measured attenuation values compared with the model predictions. When examining Figs. 3 and 4 it appears that a simple scaling (such as a multiplication with a constant factor for all attenuation values of a spectrum) could lead to an even better agreement between model and prediction. On
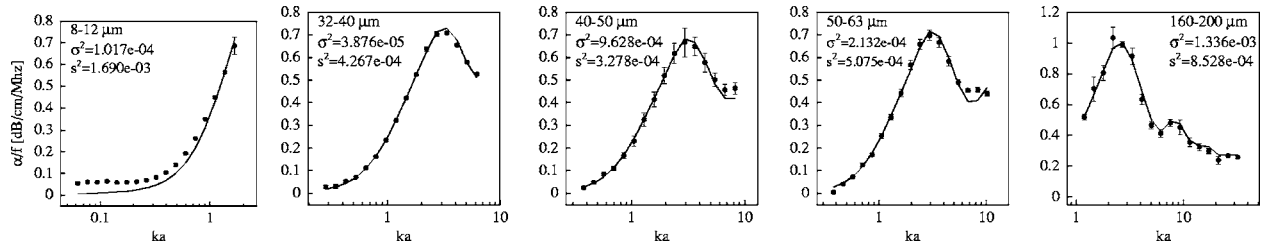
FIG. 5. Measured (dots) and calculated (lines) ultrasonic attenuation spectra of glass fractions in water. Calculations were carried out by using best-fit solids volume concentration values. Compare to Fig. 3. The wave number ka was calculated using the mean diameter d_{50} of the laser diffraction results and the frequency vector $\mathbf{f}$.

examination of the model [Eq. (13)] it can be concluded that such a scaling parameter could be expressed in the form of a modified solids volume concentration. Equation (10) expresses attenuation as linearly dependent in the factor ϕ.

To investigate the hypothesis of a scaling-dependent influence, the model [Eq. (13)] was subjected to a parameter estimation procedure. During this step the solids volume fraction ϕ was treated as a variable while all other parameters (material properties, size distribution) remained at the previously defined values. The graphs in Fig. 5 were obtained by a least squared error fit of Eq. (13) in combination with Eq. (18) to the measured attenuation data, with ϕ being the only fit parameter. Figure 5 shows that with best-fit concentration values, better agreement between measured and predicted attenuation functions can be achieved.

To quantify the agreement between measured and predicted best-fit curves, the coefficients of variation can be examined. Therefore, the coefficient of variation of measurement and fit s^2 was calculated according to

$$s^2 = \frac{\sum_{i=1}^{n} (\alpha_{tot_i} - \alpha_{meas_i})^2}{n-1}.$$ (19)

This coefficient was compared to the average measurement variance σ^2 as presented in Eq. (20):

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} \sigma_i^2.$$ (20)

Each value σ_i corresponds to the variance from the repeated measurements m_i at the frequency f_i, as shown in Eq. (21):

$$\sigma_i^2 = \frac{\sum_{k=1}^{m_i} (\alpha_{meas_{i,k}} - \bar{\alpha}_{meas_i})^2}{m_i - 1}.$$ (21)

Table IV lists values of σ^2 and s^2 along with nominal and approximated volume fractions for the studied material combinations. For all examined cases, σ^2 and s^2 have a comparable order of magnitude, which implies that the agreement of the estimation (using the best fit concentration values) is not significantly different from the random error obtained during the repeated attenuation measurements. For every examined case, the solids volume fraction fit caused an improvement of the agreement between prediction and measured result. The details of the curves are well reproduced (i.e., locations as well as magnitudes of the extrema, for glass-water 160–200 μm, see Fig. 5).

It is noteworthy that the best-fit volume fractions are larger by a factor of 1 to 1.4 in comparison with the measured values (see Table IV). One explanation for the observed deviation could be a method-dependent influence. In fact, the solids volume fraction showed variations between consecutive repeat measurements, even though great effort had been made to maintain constant solids volume fractions in the measurement chamber. It is possible that the method used for sampling the suspension for concentration analysis leads to lower values than the actual concentration. It can be speculated that gravitational settling removed particles from the pipette in the very moment the sampling was carried out, resulting in a sampling volume with a smaller particle fraction compared to the measurement chamber. These types of systematic procedural errors cannot be ruled out.

A second interpretation relies on the assumption that the data acquisition by the spectrometer results in generally higher attenuation values. However, measurements for pure liquids (see Table III) do not indicate that there is a general trend to higher attenuation values. Furthermore, the measuring device was specifically developed for particle sizing applications and its capabilities have been repeatedly confirmed.[4,9,18,23] It is not possible to provide a conclusive evaluation of device influence in this experiment, since the data used for the study was accessed as provided by the spectrometer, that is, without further consideration of the signal processing. As pointed out before, the spectral coverage and frequency resolution may have influenced the results.

Finally, in the event that device- or method-dependent effects are not responsible for the observed deviation, it is possible that effects of multiple scattering may be present in the measured attenuation spectra. However, Hall[7] and Mobley[8] stated that measurements can be considered to be in the single scattering region if the following condition applies:

$$c_{MH} = \frac{\phi 12}{k^2 d^3} |f(\theta)| \ll 1 \quad (\theta = 0 \text{ or } \pi).$$ (22)

In Eq. (22) the scattering length f is defined by

$$f = \frac{i}{k} \sum_{n=0}^{\infty} (2n + 1) A_n P_n(\cos \theta).$$ (23)

For all of the measurements in the present study this criterion is satisfied, since the values c_{MH} do not exceed 10^{-5}. Regardless of its origin, the discrepancy between model and measurement does not invalidate the approach using the
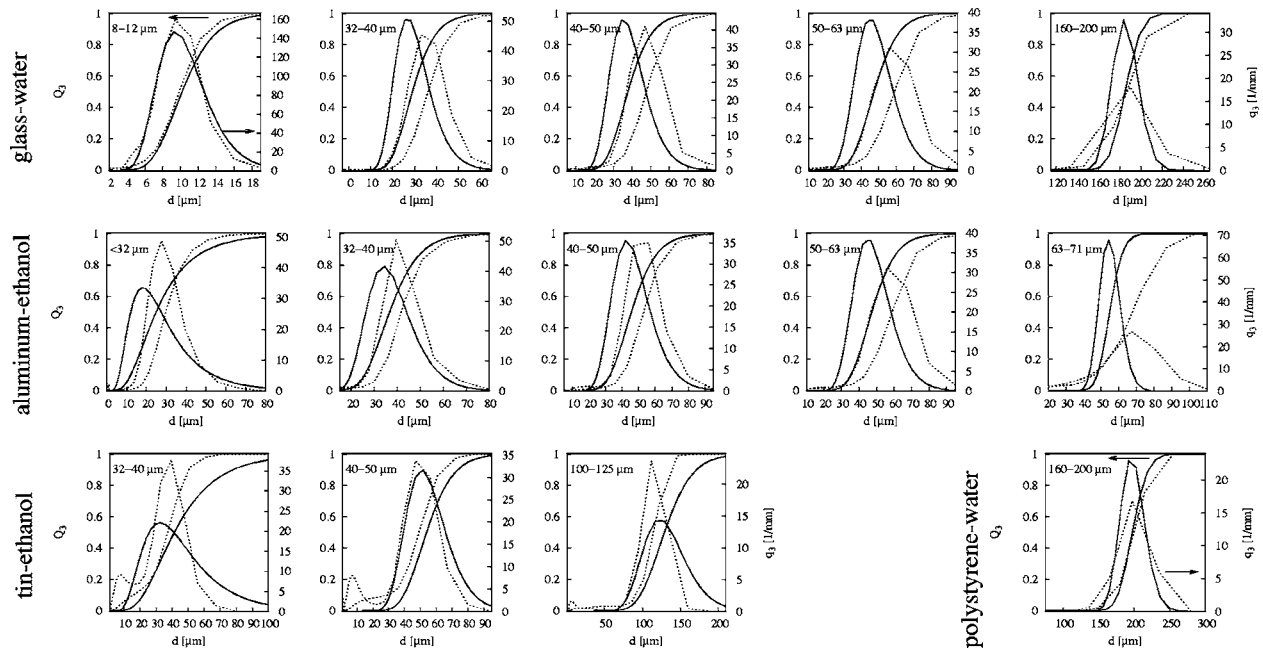
FIG. 6. Cumulative (Q_3) and density size distributions (q_3) as functions of the particle diameter d, obtained by laser diffraction spectrometry (dashed line) and from measured attenuation data (solid line) for the studied material combinations, $\phi=1$ Vol-%. The values based on the attenuation data were calculated by best-fit approximation based on a long-normal size distribution function.

Faran model, because the difference appears to be remediable by a simple scaling of the data.

VII. PARTICLE SIZE DISTRIBUTIONS

The validity of the approach for polydisperse particle sizing applications in the intermediate wave regime was evaluated by estimating particle size distributions by performing a best-fit of the parameters of the log-normal size distribution function q_3 [Eq. (17)]. Since the selected value of the solids volume fraction ϕ was found to influence the agreement between model and measured results, this parameter was considered as a third variable of the fit process (besides the mean diameter d_{50} and the standard deviation σ_{ln}). An implementation of the Levenberg-Marquardt approximation algorithm was applied to determine the best-fit values.

Figure 6 depicts the resulting log-normal size distributions along with their counterparts measured by laser diffraction spectrometry. The results are shown for selected material combinations of all examined particle size fractions. Each graph contains both the cumulative size distribution Q_3 and the density distribution q_3 as functions of particle diameter d. A compilation of best-fit values for all measured attenuation spectra is listed in Table V. To quantify the goodness of fit of the parameter estimation results, the variance between measured attenuation spectra and the estimated spectra based on the best-fit size distributions was computed. Variance values are listed in Table V besides fit results and sieving mesh sizes.

Judging by the agreement between the graphical representation of the best-fit size distributions, it appears that the estimates are comparable to the laser diffraction results. This is the case for both the mean values and distribution widths. In every case, the order of the size fraction is correctly re-

produced, since the median diameter d_{50} is comparable to sieving mesh sizes (Table V, column "diameter fraction") and laser diffraction (Fig. 6). Distribution widths of ultrasonic attenuation and laser diffraction results are similar, even if mean values do not agree (Fig. 6, glass-water $32-40$ μm, $40-50$ μm, $50-63$ μm; aluminum-ethanol <32 μm, $32-40$ μm, $40-50$ μm, $50-63$ μm).

For some cases, mean values of the best-fit are smaller, although of the same order of magnitude as the measured values. In select cases, the widths of the calculated distribution do not match the measured widths (glass-water $160-200$ μm, aluminum-ethanol $63-71$ μm, tin-ethanol $100-125$ μm, polystyrene-water). In most of these cases, the σ_{ln} values result in narrower calculated distributions. As stated above, this is mainly caused by shortcomings of the laser diffraction result interpretations.

In many of the observed measurements, the fit variance values listed in Table V are of comparable order of magnitude as the average measurement variance, indicating good fit results of the calculations. For some results, the success of the approximation procedure as expressed by the variance was found to be greatly influenced by the best-fit starting values. For one measurement, glass $8-12$ μm in ethanol, the algorithm did not converge regardless of the starting values. It is possible that the squared error function (to be minimized by the Levenberg-Marquardt algorithm) did not expose a distinct global minimum because of the monotonous character of the attenuation function. The presented approach for the polydisperse case is monotonous for the smaller size fractions examined in this work [i.e., compare Figs. 3(a), 4(a), and 4(b), fraction with diameters $8-12$ μm].

Additionally, a general feature of the fit approach needs to be considered when evaluating the fit results. The log-normal size distribution excludes a wide range of size distri-

TABLE V. The numerical results of the approximation of a log-normal size distribution to the measured attenuation data. The value d_{50} describes the average particle diameter, while σ_{ln} is the standard deviation of the distribution. σ^2 and s^2 are the coefficients of variation of the measurement and of measurement and fit, respectively.

Material combination	Diameter fraction (μm)	d_{50} (μm)	σ_{ln}	ϕ (Vol-%)	σ^2 (dB/cm/MHz)^2	s^2 (dB/cm/MHz)^2
Glass-ethanol	8–12	$\cdots$	$\cdots$	$\cdots$	1.57×10^{-3}	
	32–40	29.3	0.35	1.8	7.05×10^{-4}	6.05×10^{-4}
	40–50	40.5	0.26	1.7	4.72×10^{-4}	1.76×10^{-4}
	50–63	50.1	0.25	1.8	1.12×10^{-4}	3.15×10^{-4}
	160–200	188.4	0.13	2.0	1.54×10^{-3}	3.18×10^{-4}
Glass-water	8–12	10.4	0.28	1.1	1.02×10^{-4}	1.10×10^{-3}
	32–40	29.7	0.28	1.1	3.87×10^{-5}	8.53×10^{-5}
	40–50	38.8	0.25	1.1	9.63×10^{-4}	9.04×10^{-5}
	50–63	48.6	0.22	1.1	2.13×10^{-4}	2.57×10^{-4}
	160–200	189.9	0.065	1.6	1.33×10^{-3}	5.83×10^{-4}
Glass-NaCl solution	8–12	9.9	0.67	1.6	1.94×10^{-4}	8.91×10^{-3}
	32–40	30.9	0.32	2.1	2.75×10^{-4}	7.3×10^{-4}
	40–50	41.3	0.29	1.8	3.76×10^{-4}	9.6×10^{-4}
	50–63	48.8	0.26	1.8	3.21×10^{-4}	5.60×10^{-4}
	160–200	184.7	0.16	1.3	5.43×10^{-4}	3.8×10^{-4}
Aluminum-water	<32	24.5	0.42	1.0	4.61×10^{-5}	6.00×10^{-4}
	32–40	36.8	0.34	1.0	5.35×10^{-5}	3.3×10^{-4}
	40–50	44.0	0.28	1.0	4.75×10^{-5}	1.50×10^{-4}
	50–63	50.9	0.36	1.3	1.62×10^{-4}	6.96×10^{-4}
	63–71	59.3	0.32	3.5	1.22×10^{-3}	2.3×10^{-3}
Aluminum-ethanol	<32	26.4	0.51	1.4	3.14×10^{-4}	3.64×10^{-4}
	32–40	37.6	0.27	1.4	3.73×10^{-4}	5.01×10^{-4}
	40–50	46.8	0.26	1.5	1.37×10^{-4}	7.26×10^{-4}
	50–63	54.5	0.20	1.1	1.14×10^{-4}	4.60×10^{-4}
	63–71	56.2	0.11	1.1	1.14×10^{-4}	5.72×10^{-4}
Tin-water	32–40	41.8	0.39	0.8	1.94×10^{-4}	7.02×10^{-4}
	40–50	57.6	0.44	0.7	7.67×10^{-4}	9.06×10^{-4}
	100–125	134.0	0.50	1.0	6.63×10^{-4}	3.55×10^{-4}
Tin-ethanol	32–40	43.1	0.49	1.3	5.41×10^{-4}	6.24×10^{-4}
	40–50	55.3	0.24	1.8	1.52×10^{-3}	1.7×10^{-3}
	100–125	132.5	0.22	2.2	2.77×10^{-3}	1.4×10^{-2}
Polystyrene-water	160–200	202.3	0.87	0.4	4.46×10^{-4}	1.05×10^{-2}

bution types (i.e., multimodal distributions). In cases where the log-normal distribution is inappropriate, flawed size distributions are calculated if this function type is applied anyway. This may have been the case with some fractions, especially for those where a significant small particle fraction (diameter $d < 5$ μm) is present (Fig. 1, tin 32–40 μm).

Finally, the parameter estimation is a second indication in favor of the assumption of a scaling-dependent discrepancy between attenuation model and measured result. In some cases, the best-fit concentrations are twice as large (see Table V). When evaluating this result it needs to be considered that the parameter fit process used to find these concentration values involves three parameters. Otimization processes depending on more parameters are characterized by more degrees of freedom. For ill-posed fit problems, this results in a greater uncertainty for each of the fit values when compared to a fit of a single value, as carried out for the comparison of the attenuation spectra (Sec. V). Thus, the fit values reported in Table V can be expected to show a larger deviation from the nominal values. Nevertheless, any deviation between nominal measurements and estimate values indicates either a problem with measurement setup and method, or an incorrectly modeled influence of the concentration. Despite these shortcomings, the results are an indication that the size distributions calculated from measured attenuation spectra are appropriate representations of the dispersed systems.

VIII. CONCLUSION

In this work, a possible approach for ultrasonic attenuation spectroscopy for particle size characterization in the intermediate wave regime was outlined and evaluated. A methodology relying on the Faran elastic scattering model was cast for the polydisperse case and a vector of frequencies and compared to measured data. Various combinations of particulate and continuous phase were experimentally studied with respect to their sound attenuating behavior in the region of elastic scattering. A commercially available attenuation spectrometer designed for particle characterization was used for the measurement of attenuation. The measured polydisperse attenuation spectra for a vector of discrete frequencies were found to correlate with model predictions for the polydisperse case. By studying materials with diverse acoustic contrast, the influence of material properties on the validity of the approach was examined. The only material properties necessary for a complete description of the suspension were the densities and elastic properties of its components.

The measurements resulted in attenuation values higher than those predicted by theory. The deviation was found across the whole frequency region and for all material combinations. To evaluate the nature of the deviation, the solids volume concentration was treated as a variable and subjected to a fit process. As a result of the best-fit procedure, agreement between model and measurement result was further increased. By evaluating the measurement results, it can be concluded that the measurement setup or the experimental procedure are likely sources of the deviation.

Particle size distributions associated with the measured attenuation spectra were compared with those obtained by an independent method (laser diffraction) for additional validation of the approach. For all measured attenuation spectra, size distributions were calculated by a best-fit approximation. The resulting distributions were found to agree with the laser diffraction size distributions in the limit of shortcomings of each method. Deviations present in the distribution results must be addressed and discussed in the light of the applied measurement technology. Furthermore, the validity of the methodology should be confirmed through the application of additional polydisperse size distribution types, i.e., bi-modal suspensions. Nevertheless, the findings indicate that a methodology using the Faran model for elastic scattering is applicable to particle size measurements in the intermediate wave regime, and that size distributions can be inferred from measured attenuation spectra in the large particle limit.

ACKNOWLEDGMENTS

[1] P. S. Epstein and R. R. Carhart, "The absorption of sound in suspensions and emulsions. I. Water fog in air," J. Acoust. Soc. Am. **25**, 553–565 (1953).

[2] J. R. Allegra and S. A. Hawley, "Attenuation of sound in suspensions and emulsions: Theory and experiments," J. Acoust. Soc. Am. **51**, 1545–1564 (1972).

[3] D. J. McClements, Y. Hemar, and N. Herrmann, "Incorporation of thermal overlap effects into multiple scattering theory," J. Acoust. Soc. Am. **105**, 915–918 (1999).

[4] A. Dukhin and P. Goetz, "Acoustic spectroscopy for concentrated polydisperse colloids with high density contrast," Langmuir **12**, 4987–4997 (1996).

[5] J. J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).

[6] V. C. Anderson, "Sound scattering from a fluid sphere," J. Acoust. Soc. Am. **22**, 426–431 (1950).

[7] C. S. Hall, J. N. Marsh, M. S. Hughes, J. Mobley, K. D. Wallace, J. G. Miller, and G. H. Brandenburger, "Broadband measurements of the attenuation coefficient and backscatter coefficient for suspensions: A potential calibration tool," J. Acoust. Soc. Am. **101**, 1162–1171 (1997).

[8] J. Mobley, K. R. Waters, C. S. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "Measurements and predictions of the phase velocity and attenuation coefficient in suspensions of elastic microspheres," J. Acoust. Soc. Am. **106**, 652–659 (1999).

[9] F. Babick, F. Hinze, M. Stintz, and S. Ripperger, "Ultrasonic spectrometry for particle size analysis in dense submicron suspensions," Part. Part. Syst. Charact. **15**, 230–236 (1998).

[10] J. Evans and K. Attenborough, "Coupled phase theory for sound propagation in emulsions," J. Acoust. Soc. Am. **102**, 278–282 (1997).

[11] A. Hay and D. Mercer, "On the theory of sound scattering and viscous absorption in aqueous suspensions at medium and short wavelengths," J. Acoust. Soc. Am. **78**, 1761–1771 (1985).

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

[13] R. J. Hunter, *Foundations of Colloid Science*, 1st ed. (Clarendon, Oxford, 1986).

[14] ISO, "Representation of results of particle size analysis," ISO 9276 (International Organization for Standardization, Geneva, Switzerland, 1998).

[15] R. Challis, J. Tebbutt, and A. Holmes, "Equivalence between three scattering formulations for ultrasonic wave propagation in particulate mixtures," J. Phys. D **31**, 3481–3497 (1998).

[16] F. Peters and L. Petit, "Propagation of ultrasound waves in concentrated suspensions," Acust. Acta Acust. **86**, 838–846 (2000).

[17] P. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).

[18] F. Babick and S. Ripperger, "Information content of acoustic attenuation spectra," Part. Part. Syst. Charact. **19**, 176–185 (2002).

[19] A. Hipp, G. Storti, and M. Morbidelli, "Acoustic characterization of concentrated suspensions and emulsions. II. Experimental validation," Langmuir **18**, 405–412 (2002).

[20] D. J. McClements, "Principles of ultrasonic droplet size determination in emulsions," Langmuir **12**, 3462–3467 (1996).

[21] P. Mougin, D. Wilkinson, K. J. Roberts, and R. Tweedie, "Characterization of particle size and its distribution during the crystallization of organic fine chemical products as measured in situ using ultrasonic attenuation spectroscopy," J. Acoust. Soc. Am. **109**, 274–282 (2001).

[22] A. Dukhin and P. Goetz, "US Patent 6,109,098: Method and device for characterizing particle size distribution and zeta potential in concentrated systems by means of acoustic and electroacoustic spectroscopy" (2000).

[23] A. S. Dukhin and P. J. Goetz, "Characterization of chemical polishing materials (monomodal and bimodal) by means of acoustic spectroscopy," Colloids Surf. A **158**, 343–354 (1999).

[24] H. Kuchling, *Taschenbuch der Physik* (Hanser, München, 2001).

[25] D. R. Lide, *CRC Handbook of Chemistry and Physics*, 80th ed. (CRC, Cleveland, OH, 1999).

[26] G. W. Kaye and T. H. Laby, *Tables of Physical and Chemical Constants*, 16th ed. (Longmans, London, 1995).

[27] L. W. Anson and R. C. Chivers, "Thermal effects in the attenuation of ultrasound in dilute suspensions for low values of acoustic radius," Ultrasonics **28**, 16–25 (1990).

Miniature, high performance, low-cost fiber optic microphone

Joseph A. Bucaro,[a)] Nicholas Lagakos, and Brian H. Houston
Naval Research Laboratory, 4555 Overlook Avenue SW, Washington, DC 20375-5320

Jacek Jarzynski and Maxim Zalalutdinov
SFA, Inc., 2200 Defense Highway, Suite 405, Crofton, Maryland 21114

A small, high performance fiber optic microphone has been designed, fabricated, and tested. The device builds on a previous design utilizing a thin, seven-fiber optical probe, but now adds a micromachined 1.5 μm thick silicon diaphragm active element. The resulting sensor head is thin (several millimeters) and light, and the overall microphone system is less expensive than conventional microphones with comparable performance. Measurements in the laboratory using a standard free-field technique at high frequencies, an enclosed calibrator at lower frequencies, and pseudostatic pressure changes demonstrate uniform broadband response from near dc (0.01 Hz) up to near 20 kHz. The measured microphone internal noise is nearly flat over this band and does not exhibit noticeable levels of $1/f$ noise. Over the audible portion of this band, the minimum detectable pressure is determined to be 680 μPa per root Hz with further reductions possible using lower noise/higher power light sources and/or improvements in the diaphragm. In contrast to conventional high-performance microphones, there is no need for preamplifier packages close to the relatively small sensor head resulting in much lower acoustic scattering cross sections. This attribute, together with high performance, low cost, and immunity to emi, makes the microphone ideal for multielement array applications. [DOI: 10.1121/1.2000749]

I. INTRODUCTION

Since the initial surge of interest in fiber optic sensors some 20 years ago, the development of fiber optic microphones has made slow but steady progress. This interest has been connected with one or more of the advantages that such optically based sensors are perceived to possess over conventional capacitive technology. These include immunity to electromagnetic interference, safety in hazardous, explosive environments, material component compatibility with hostile environments, and the absence of troublesome capacitive effects at low frequencies requiring expensive, bulky preamps at the sensor head. Up to the present time, only interferometric fiber optic microphones have been demonstrated having threshold detectabilities $P_{\min}$ comparable to those of the best electrical microphones. While interferometric sensors can be made to achieve very low threshold levels, they are somewhat complex and costly. Our interest here is in achieving a noninterferometric, optical intensity-based microphone with a very low sound detection threshold. He and Cuomo (1992) have presented analysis indicating that optical microphones based on reflection from a membrane should be possible having $P_{\min}$ levels below 100 μPa/$\sqrt{\text{Hz}}$. This paper reports on the development of a noninterferometric fiber optic microphone based on their concept which now approaches the state-of-the-art threshold performance of many commercially available conventional microphones while delivering all the usual advantages connected with optical versus electrical

based sensing. In addition, these optical microphones are considerably smaller, lighter, and cheaper than the conventional high performance devices.

The starting point for this new fiber optic microphone is a small, light, low-cost device reported by Bucaro and Lagakos (2001) which was designed for acoustic active control systems, applications which typically do not require sensors with low threshold detectability. This device in turn was inspired by the much earlier work of He and Cuomo (1991), who introduced the first multifiber probe lever microphone, Hu *et al.* (1992), who developed the first experimental fiber optic microphone based on this idea, and Zuckerwar *et al.* (1995), who in reporting a high-temperature fiber-optic microphone presented a thorough analysis of the various components of the device. Each of these devices, including that reported here, is based on the principle illustrated in Fig. 1. Here an optical fiber probe carries light to a reflective diaphragm and some of the reflected light is recaptured by an adjacent fiber which transmits this optical signal to a photodetector. Motion of the reflector normal to the fiber's end in response to an acoustic pressure linearly modulates the light level recaptured by the collection fiber so that the photodetector current contains a signal proportional to the varying acoustic pressure disturbance at the diaphragm. The fiber optic microphone reported by Bucaro and Lagakos (2001) utilized a bundle of seven fibers similar to that first reported by He and Cuomo (1991) with the central fiber transmitting light to the reflective diaphragm and six outer fibers hexagonally arranged about the central fiber collecting the reflected light. This sensor was small (1.5 cm³), light (1.3 g), cheap ($\sim \$ 25$), and achieved a modest threshold detectability of 16 mPa/$\sqrt{\text{Hz}}$. This latter number is to be contrasted with that

---

[a)]Electronic mail: joseph.bucaro@nrl.navy.mil
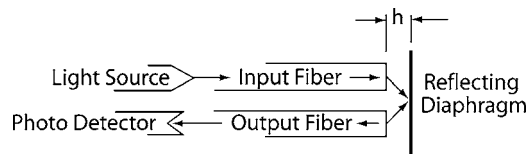
FIG. 1. Optical microphone principle.

FIG. 2. Detailed design of the probe component of the fiber optic microphone.

achieved by (low noise) broadband conventional microphones, one example of which is the Bruel & Kjaer (BK) 4180 capacitive device which has a threshold detectability of 158 $\mu Pa/\sqrt{Hz}$. The goal of this effort was to realize a noninterferometric fiber optic microphone having the overall characteristics of that reported by Bucaro and Lagakos (2001) but with a significant decrease (~two orders of magnitude) in the minimum detectable acoustic level so as to approach that of conventional microphones such as the above-mentioned BK device.

In the following sections, we present a short theoretical background which informs how we have chosen to design the new fiber optic microphone, give fabrication details, provide measurement results establishing the microphone's performance, and conclude with remarks indicating the impact this new technology should have.

II. THEORETICAL BACKGROUND

Treatment of the response of the fiber optic microphones of interest here naturally splits into two parts—that associated with the fiber bundle probe and that with the mechanical design and attachment of the reflective diaphragm.

A. Fiber probe

Considerable analysis has been presented in each of the above-quoted references regarding the optical fiber probe. The work includes one, two, and seven fiber bundle designs. In each case, one attempts to maximize the light level modulation for a given diaphragm displacement by the choice of optical fiber bundle parameters as well as the probe-diaphragm separation distance. Generally speaking, as discussed by He and Cuomo (1991) increasing the number of fibers in the bundle results in designs with increased performance. On the one hand, a seven-fiber probe design is preferred over one with fewer fibers because of its superior sensitivity. On the other hand, compared to higher fiber count designs, its natural hexagonal arrangement of outer fibers with one at the center leads to a compact, simple, relatively thin and cost-effective probe. The parameters of an individual fiber can be optimized using the theoretical expressions also given by He and Cuomo (1991). Briefly, their results call for favoring larger fiber core radii, larger fiber numerical apertures, and smaller fiber claddings. Ultimately, a trade-off is made among sensitivity, complexity, and cost—the last of which is moderated by choosing from commercially available fibers. The probe used by Bucaro and Lagakos (2001) and described in detail by Lagakos and Bucaro (2001), which uses a bundle of seven fibers each having a 0.37 numerical aperture, a 15 μm cladding, and a 200 μm core, was designed making these tradeoffs in the context of a
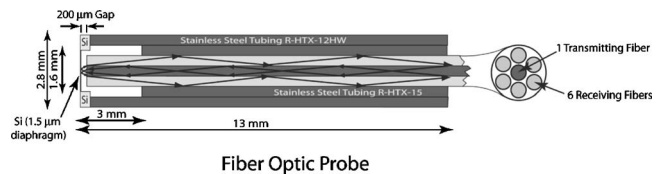
light-weight, miniature, inexpensive microphone. Accordingly, we consider this fiber bundle probe to be near optimum for our purpose and use this design (shown in Fig. 2) for our new device.

B. Diaphragm reflector

In designing a diaphragm consistent with low threshold detectability, it is useful to consider similar diaphragms used in current commercial low noise electrical microphones. In these devices, the diaphragm is a thin metal or metallized plastic foil which is set under significant tension. The purpose of the tension, T, is to increase the microphone bandwidth and to maintain a flat diaphragm surface. The diaphragm response is dominated by this tension, and the diaphragm is modeled as a membrane. The static (low frequency) displacement, X, at the center of the diaphragm produced by a uniform applied pressure P is

$$X = \frac{Pr^2}{4T},\qquad(1)$$

where r is the diaphragm radius.

The application of MEMS technology to production of microphone diaphragms offers a much wider choice of parameters. In particular, a wide bandwidth can be achieved by making r small. Also, thin, flat silicon or silicon nitride diaphragms can be made with minimum tension. When T is small the diaphragm is modeled as a clamped plate, as described in the following. The heart of the major breakthrough reported here in the fiber optic microphone is associated with the mechanical design and details of such a low tension small, thin reflective diaphragm. Consider the response of a thin circular disc of thickness t, radius r, Young's modulus E, and Poisson ratio σ to a dynamic pressure P. For various mounting conditions the center displacement X is given as

$$X = \frac{3Pr^4 H}{16Et^3},\qquad(2)$$

where for example H is given by $1 - \sigma^2$ and $4(1 - \sigma^2)$ for the clamped and freely supported attachment conditions respectively.

The minimum detectable pressure, defined to be that for a signal to noise ratio S/N=1 in a band Δf, can be estimated [He and Cuomo (1992)] by realizing that the signal S is given in terms of the displacement by $X(\partial W_{op}/\partial X)D$, where W_{op} is the recaptured reflected light intensity and D is the photodetector sensitivity:

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Bucaro *et al.*: Miniature fiber optic microphone   1407

$$P_{\min} = \frac{16Et^3N\sqrt{\Delta f}}{3r^4(1-\sigma^2)D\left(\frac{\partial W_{op}}{\partial X}\right)}. \qquad (3)$$

Here N is the rms noise, including shot noise, thermally driven diaphragm vibration, detector and LED noise, etc. Clearly, a low detection threshold requires sufficiently low N; and what is possible practically will be discussed later. However, as can be seen in Eq. (3), significant reduction in $P_{\min}$ should be possible simply through tailoring the dimensions and mechanical properties of the diaphragm. While attempting to achieve a significant reduction through a corresponding reduction in diaphragm stiffness through E is not practical, capitalizing on the strong dependence on diaphragm dimensions can be very effective. The fourth power dependence of $P_{\min}$ on r is attractive, and recognizing that the fiber optic microphone reported by Bucaro and Lagakos (2001) utilized a 6-mm-diam diaphragm, we project that $P_{\min}$ can be reduced by a factor of 100 by increasing the diaphragm diameter to 19 mm. This, however, results in a microphone large by today's standards (although still smaller than some high performance conventional microphones) with significant disadvantages for most applications, particularly for those involving arrays of microphones. On the other hand, $P_{\min}$ also has a strong dependence on the thickness t, and the required reduction in this parameter is attractive in that it can also move the device toward smaller size.

One obvious way in which to achieve a significant reduction in diaphragm thickness is through the use of micromachining techniques. Interestingly, these fabrication methods have been exploited for some time in conventional, capacitive microphones. Scheeper *et al.* (1994) have presented an excellent treatment of this subject and have demonstrated significant improvements in capacitive microphone performance by decreasing the thickness of the silicon nitride diaphragm using micromachining.

With reference to Fig. 2, for the new fiber optic microphone design we estimate the potential drop in $P_{\min}$ associated with a practically realizable decrease in the diaphragm thickness. We take the diaphragm diameter to be 1.6 mm, roughly the diameter of the seven-fiber probe. Our experience suggests that a practical diaphragm thickness is in the neighborhood of 1 to 2 μm. For 1.5 μm thickness and the above diameter, the ratio r^4/t^3 is 2×10^8 compared to 4×10^6 in the original fiber optic microphone of Bucaro and Lagakos (2001). This and Eq. (3) imply a reduction of $P_{\min}$ by a factor of 50 for the new design which would achieve our goal of ~two orders of magnitude reduction in the threshold detectability.

III. FINAL DESIGN, FABRICATION, AND ASSEMBLY

The overall fiber optic microphone design is shown in Fig. 3. Light of wavelength 850 nm from a light emitting diode (LED model Optek OPF370A) is coupled into the central fiber of a seven multimode fiber bundle and propagates to the well-polished end of this input fiber. Upon exiting the fiber, the light propagates a short distance in air
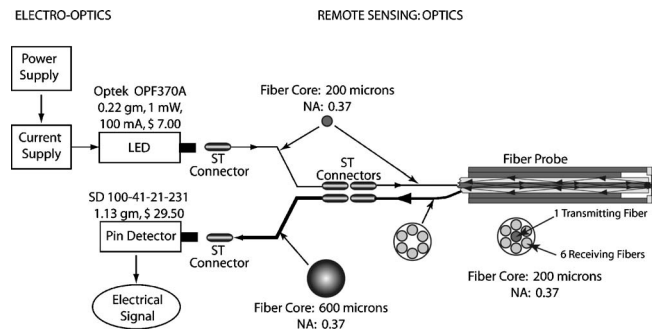
FIG. 3. Overall fiber optic microphone schematic.

(~200 μm) and is reflected by the diaphragm back into the bundle where six receiving fibers surrounding the transmitting fiber carry the recaptured light to a *PIN* photodiode (SD 100-41-21-231). The LED is driven by a current supply (100 mA current and 1 V) which requires about 100 mW of power. The transmitting and receiving multimode fibers are identical commercially available fibers with a 200 μm glass core, a 230 μm plastic clad, a 500 μm Tefzel plastic coating, and a 0.37 numerical aperture. The LED is mounted on a commercially available ST connector (slightly modified) which is then coupled to the ST connector on which the transmitting fiber is mounted. At the output end of the bundle, the ends of the six receiving fibers are coupled to a single 600 μm core fiber using another standard ST connector. This fiber is finally connected to a *PIN* silicon photodiode, mounted on a modified ST connector similar to that used for the LED.

The diaphragm was fabricated by utilizing bulk micromachining techniques as illustrated in Fig. 4. Silicon-on-insulator (SOI)$\langle100\rangle$ wafers with 1.5 μm device layer thickness and 1 μm buffered oxide (BOX) thickness were used as a starting material. Prior to micromachining, a protective SiO_2 layer (1 μm) was deposited on the front side of the device layer by plasma-enhanced chemical vapor deposition (PECVD). The same PECVD technique was then used to build a 4-μm-thick silicon oxide layer on the back side of the wafer. This relatively thick oxide was covered with a photoresist layer (~5 μm) and patterned using contact optical lithography and a subsequent CF_4-based reactive ion etch. As a result of the patterning step, circular openings (with diameter 1.6 mm, approximately the outer diameter of the optical fiber bundle) were created in the back side SiO_2 layer to allow exposure of the silicon wafer to a subsequent deep reactive ion etch (DRIE). The remaining oxide film formed a high-resistance mask which served to protect those parts of the wafer intended to become the frame of the diaphragm (see Fig. 2). A Bosch DRIE process featuring high etch rate, very high anisotropy (nearly vertical walls), and high selectivity (ratio of Si etch rate to that for SiO_2) was utilized to etch cylindrical channels originating at the circular openings in the back side oxide and penetrating through the entire thickness of the Si wafer, stopping at the bottom surface of the BOX layer. At the completion of the DRIE step, a suspended circular diaphragm was created comprised of the silicon device layer sandwiched between the BOX oxide at the bottom and the protecting PECVD oxide on the top. A short
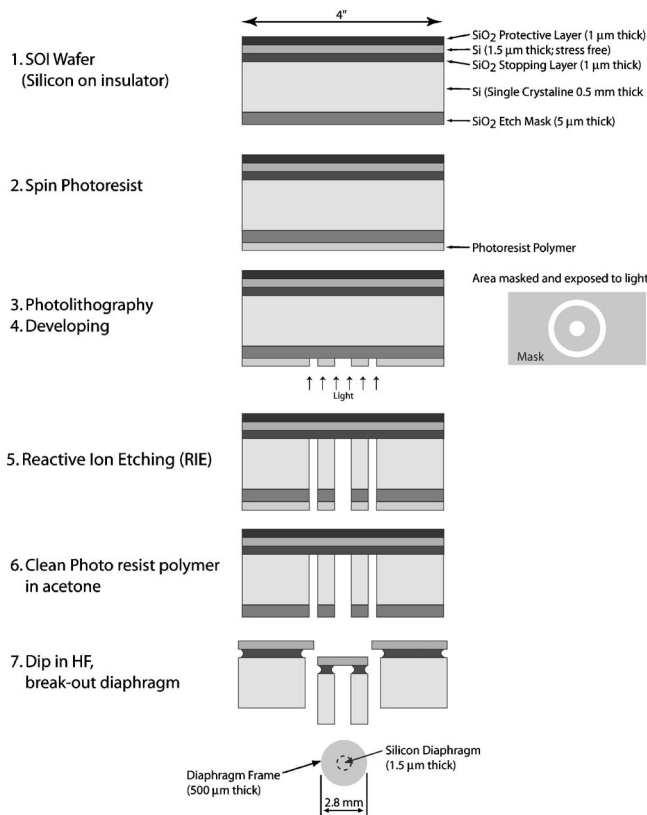
FIG. 4. Step by step procedure used to micro-machine the thin silicon diaphragm.

In the figure labels:

1. SOI Wafer (Silicon on insulator)
   - SiO₂ Protective Layer (1 µm thick)
   - Si (1.5 µm thick; stress free)
   - SiO₂ Stopping Layer (1 µm thick)
   - Si (Single Crystaline 0.5 mm thick
   - SiO₂ Etch Mask (5 µm thick)
2. Spin Photoresist — Photoresist Polymer
3. Photolithography
4. Developing — Area masked and exposed to light / Mask / Light
5. Reactive Ion Etching (RIE)
6. Clean Photo resist polymer in acetone
7. Dip in HF, break-out diaphragm
   Silicon Diaphragm (1.5 µm thick)
   Diaphragm Frame (500 µm thick)
   2.8 mm

wet etch in hydrofluoric acid (HF) was used to dissolve both oxide layers thus yielding a relatively low stress single-crystal Si diaphragm.

It is well known that in thin films residual stress can play a large role in the dynamic mechanical response. If our circular diaphragm is under tension T, the displacement response predicted by Eq. (2) is altered. For low tension values, the response is modified in the following way (Bickley, 1933):

$$X = \frac{3Pr^4H}{16Et^3}\left[1 - \frac{5T(1-\sigma^2)r^2}{6Et^3}\right]. \qquad (4)$$

It is interesting to ask at what diaphragm tension would the displacement be reduced by say 10% because of the second term in Eq. (4). Setting the second term in brackets equal to 0.1 and solving for T yields a tension of 0.13 N/m. Accordingly, as a rough rule we attempt to achieve micromachined diaphragms in which the residual tension is not significantly larger than this value.

We estimated the level of residual stress present in the finished diaphragms by the presence and number of optical fringes observed in the image of the diaphragm in white light. Typical diaphragms we fabricated exhibited less than one optical fringe across their diameter. Considering the dimensions of our diaphragm, we estimate this to be consistent with a tension of 0.04 N/m, a value lower than the 0.13 N/m goal derived earlier.

As illustrated in Fig. 4, the finished 1.5-µm-thick silicon diaphragm together with its 500-µm-thick frame to which it is rigidly attached is easily broken from the overall wafer by

the application of some weak forces. Under observation with a microscope, this 2.8 mm structure is then placed over the sensor end of the fiber probe and attached to the end of the stainless steel tubing using a two-component, hard epoxy adhesive applied to the end of the tubing.

In the final structure, the actual distance between the diaphragm and the fiber probe end face is an important parameter in determining the final sensitivity. Before diaphragm attachment, this distance was predetermined and set using the procedure outlined by Bucaro and Lagakos (2001) wherein a controlled micrometer translation stage is used to vary the distance between the fiber bundle (within the stainless steel tubing) and a fixed mirror mounted on a piezoelectric transducer at the end of the outer stainless steel tubing. Both the ac and dc displacement sensitivity ($\partial W_{op}/\partial X$) in Eq. (3) can be measured in this way versus the gap separation allowing the fiber bundle to be fixed with respect to the supporting tubing so that the sensitivity is maximum.

IV. MEASUREMENTS

Measurements to determine the dynamic, broadband response of the fiber optic microphone were carried out using two systems, a dynamic pressure calibrator over the low frequency portion of the band and a standard open-air, speaker technique over the high frequency segment. We arranged the measurements so that there was a small region of overlap for the two techniques. In addition, we measured the pseudo-static response (change in optical signal over ~100 s period) to small step changes in pressure (water height changes) using a 6.4-mm-i.d. U-shaped plastic tubing to which the microphone diaphragm end was attached.

The low frequency measurements were made over the band 30–1000 Hz by placing the sensor in a Bruel & Kjaer (BK) pressure calibrator, type 4221, together with a standard calibration microphone. The latter was a BK 4938 sensor with a 2669 BK preamplifier and a 2690 BK amplifier. A broadband chirp pulse covering the band was applied to the calibrator, and the output signals from the fiber optic and calibration microphones were recorded and stored in a Macintosh computer using a ML 750/M Power Lab recorder.

A speaker setup was used to measure the fiber optic microphone over the higher frequency band, 900 Hz–25 kHz. This technique used a free-field configuration in which the sound source (a high-fidelity loudspeaker) was mounted on a pedestal with the fiber optic microphone attached to a vertical supporting rod. Both the source and the microphone were positioned so that the nearest reflecting surfaces in the room were at least 1.5 m away. The loudspeaker was driven with a gated sinusoidal signal, whose duration was sufficiently short so that the direct sound signal from the speaker to microphone and echoes from surrounding reflecting surfaces could be separated in time. The fiber optic microphone was removed after recording the response signals and replaced by the BK calibration microphone at the identical location to provide the absolute pressure levels required to calibrate the fiber optic microphone response. At the highest frequency measured (25 kHz), the acoustic wavelength is
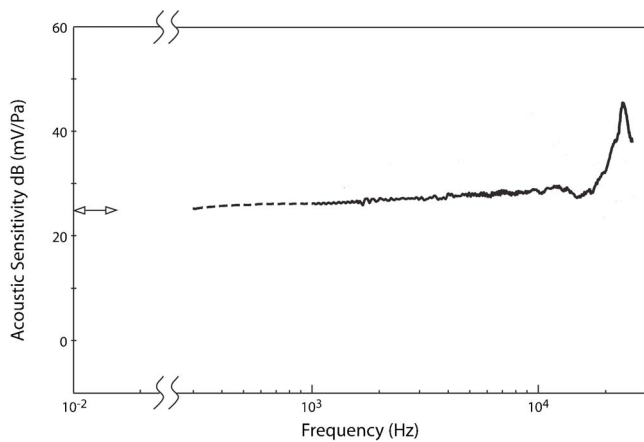
FIG. 5. Laboratory measured fiber optic microphone response vs frequency (—) cw free-field; (- - -) BK 4221 pressure calibrator; and (↔) pseudo-static measurements.

still much larger than the microphone diameter, and we assume the measured free-field response to be the same as the pressure response.

The fiber optic microphone response measured over the entire band is shown in Fig. 5 in units of mV/Pa which have been converted directly from W/Pa upon multiplication by the photo-detector sensitivity $D = 1.1 \times 10^6$ V/W for our detector package. We also recorded the intrinsic noise level of the microphone in a quiet room (ambient acoustic levels $\ll 100$ μPa), and this is shown in Fig. 6. (This can be converted to Pa/$\sqrt{\text{Hz}}$ upon division by the microphone sensitivity shown in Fig. 5.)

V. DISCUSSION

As can be seen (Fig. 5), the microphone exhibits a reasonably flat, broadband response to frequencies approaching 20 kHz, above which a large resonance dominates. We also notice that the pseudostatic response is virtually identical to that measured dynamically below 1 kHz. The small fluctuations observed in the frequency response above 1 kHz (measured in an open laboratory) are, for the most part, related to ambient noise in the laboratory. This is contrasted to the virtually noise-free measurements made at lower frequencies within the shielded calibrator. One can notice above this noise level, however, a small, broad peak just above 10 kHz. This is in fact associated with the first resonance of the disk-
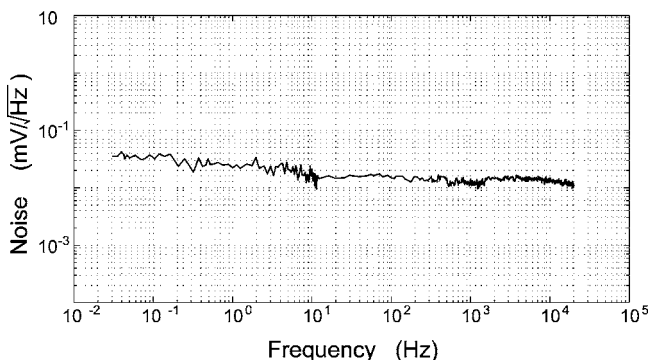
FIG. 6. Measured microphone intrinsic noise levels vs frequency.

like diaphragm. The resonance frequencies and associated mode shapes of circular disks, which depend on E, t, r, and σ, are easily predicted [see, for example, Soedel (1981)]. For our diaphragm parameters, the first resonance is predicted to occur at 10.1 kHz and involves radially symmetric vibration. This resonance response peak is in fact just visible in the response curve of Fig. 5. The next resonant response predicted is that for the so-called 0,1 mode which is an antisymmetric vibration whose nodal line is located along the disc diameter. The resonance frequency is predicted to be 21.2 kHz, which coincides almost exactly with the frequency of the large response peak seen in Fig. 5. A relatively large sensor response (compared for example to that from the first vibrational mode) is to be expected for the 0,1 mode as can be understood from the following consideration. The deflection in this mode results in a tilting of the diaphragm plane about its diameter (along which most of the probing light is incident), and this motion is considerably more effective at changing the light level captured by the outer ring of receiving fibers than motion of the diaphragm normal to its plane.

We would like to determine to what extent the diaphragm fabrication and attachment procedures have produced a sensitivity as predicted by Eq. (2) In particular, from Eq. (2) one sees that the diaphragm response to pressure P is given by $(\partial X/\partial P) = 3 \ r^4(1 - \sigma^2)/(16 \ Et^3)$. For our intended diaphragm dimensions ($r = 8 \times 10^{-4}$ m and $t = 1.5$ μm) and using the available elastic parameters for silicon ($E = 14.5 \times 10^{10}$ Pa and $\sigma = 0.245$), this predicts $(\partial X/\partial P) = 1.47 \times 10^{-7}$ m/Pa. Since the microphone sensitivity in W/Pa is given by $(\partial X/\partial P) \times (\partial W_{\text{op}}/\partial X)$, $(\partial X/\partial P)$ can be extracted from the measurements by using the value for sensitivity shown in Fig. 5 and dividing by $(\partial W_{\text{op}}/\partial X)$ measured for the probe. This latter parameter was determined by mounting the probe on a micrometer translation stage and placing it in close proximity to a mirror mounted on a piezoelectric transducer as described by Bucaro and Lagakos (2001). The displacement amplitude of the vibrating mirror was obtained from the output of a small reference accelerometer (Endevco 2250A) mounted close to the mirror. For our gap separation, $\partial W_{\text{op}}/\partial X$ is found to be 6.35×10^{-1} W/m. This gives $(\partial X/\partial P) = 3.46 \times 10^{-8}$ m/Pa. Thus the experimental value for $(\partial X/\partial P)$ is about four times lower than that predicted by Eq. (2).

At the present time, the exact source(s) for this reduced diaphragm sensitivity is not certain. One could surmise that it may involve a combination of effects. These would include: (1) diaphragm thickness variation or uncertainty; (2) silicon material parameter variations; (3) increased tension in the attached diaphragm; (4) a decreased effective diaphragm diameter resulting from the flow of adhesive onto the perimeter of the diaphragm; (5) the neglect of effects associated with the entrapped air in the cavity between the diaphragm and fiber bundle end; and (6) neglect of the crystal diaphragm anisotropy.

Unlike the case for the first four issues above, the quantitative impact of the entrapped air on the sensitivity of our actual microphone can be accurately determined since we know the required parameters. This air is confined between the diaphragm and the end surface of the optical fiber bundle,

and it can be modeled as an isothermal film squeezed by the motion of the diaphragm. This model was originally developed by Crandall (1918) and recently applied by Hansen *et al.* (2000) to micromachined transducers. An approximate solution of the squeeze-film equations is obtained assuming piston motion of the diaphragm and a free boundary with the atmosphere at the edge of the fiber bundle. The force of the air on the diaphragm consists of two components: (1) a component in phase with velocity which dissipates energy like a resistance, R_{air}; (2) a component in phase with the diaphragm displacement that stores energy through compression of the gas, like a spring with a stiffness constant, S_{air}. An important parameter of the air film is $\beta = (12\,\mu\omega/P_0)^{1/2}(r/h)$, where ω is the angular frequency of vibration, μ is the viscosity of the air, P_0 is atmospheric pressure, r is the diaphragm radius, and h is the gap between the diaphragm and fiber optic bundle. When $\beta \ll 1$ the expressions for the equivalent resistance and stiffness of the air film are

$$R_{air} = \frac{3\pi\mu r^4}{2d^3} \quad S_{air} = \frac{3\pi\mu^2 r^6}{P_0 d^5}. \tag{5}$$

An exact but more complicated solution for the air motion was developed by Zuckerwar (1978). However, Eq. (5) is sufficient for our estimates, presented in the following, for the effect of the air on the diaphragm response.

The above-noted expressions are used to calculate R_{air} and S_{air} at a frequency of 10 kHz, to determine the effect of the air on the fundamental resonance of the diaphragm (resonance frequency $\omega_0 \cong 2\pi \times 10^4$ rad/s). For the fiber optic microphone in this study the parameters in Eq. (5) are: $R_{air} = 4.3 \times 10^{-6}$ Ns/m, and $S_{air} = 9.66 \times 10^{-5}$ N/m. The air stiffness, S_{air}, can be compared with the stiffness of the diaphragm, defined as S_m= applied pressure/average diaphragm deflection = $192\pi D_b/r^2$, where D_b is the bending stiffness of the diaphragm. The calculated $S_m = 52$ N/m, and it is seen that $S_{air} \ll S_m$. Apparently, because the gap d between the diaphragm and the end of the fiber bundle is relatively large (200 μm), the air in this gap can move freely and does not impede the motion of the diaphragm. We conclude that the trapped air is not responsible for the drop in sensitivity. We also point out that the present design does not include a capillary vent for pressure equilibration on the two sides of the diaphragm, a technique often used in microphone designs. An estimate of the backpressure resulting from motion of the diaphragm to pressure P for our cavity size gives $P \times 10^{-4}$ and suggests that equilibration is not necessary.

Having estimated R_{air}, we can also determine whether the breadth of the first resonance peak can be attributed to the trapped air. The quality factor for the fundamental resonance, Q_{air}, corresponding to the above-noted air resistance can be estimated from $Q_{air} = m\omega_0/R_{air}$ where m is (approximately) the mass of the diaphragm. From the above, with $m = 6.9 \times 10^{-9}$ kg, the calculated $Q_{air} \sim 10^2$. This relatively high value of the quality factor shows that the air resistance is small and cannot account for the observed width $Q^{-1}\omega_0/2\pi$ of the fundamental resonance. Apparently, the

mechanical properties of the diaphragm itself and its attachment material are responsible for the relatively high damping of the fundamental.

At the present time, we do not know accurately the parameters required to determine the contributions of effects numbered (1)–(4) to the decrease in sensitivity. Regarding (1) and (2), diaphragm thickness and material property variation or uncertainty, our experience indicates that the micromachining process and the high quality of the silicon wafers utilized result in thickness variations of no more than ±0.3% and uncertainties no more than ±3%; Young's modulus variations of no more than ±10%; and Poisson ratio uncertainties of less than ±10%. Given the dependence of sensitivity on these parameters indicated in Eq. (2), these uncertainties could lead to sensitivity differences of 9%, 10%, and 1% for thickness, Young's modulus, and Poisson ratio, respectively. Taking the largest of each of these would account for only about 20 % of the sensitivity drop.

We can obtain a very rough estimate of what diaphragm tension would lead to a factor 4 drop in the sensitivity using Eq. (4). [The estimate is rough since Eq. (4) is accurate only for small tensions.] Setting the bracketed term equal to 1/4 gives a value of 0.72 N/m for the tension, a value some 20 times higher than our estimate based on the number of optical fringes seen from white light reflected from the unattached diaphragms. While the attachment process itself can lead to increased tension, we do not think an increase by a factor of 20 is likely, nor do we see any significant change in the fringe pattern associated with the attached diaphragm.

Next, we consider the possibility of decreased effective diaphragm diameter resulting from the flow of adhesive onto its perimeter. Given the strong fourth power dependence of the sensitivity on diaphragm radius, it would only take enough material sufficient to constrain an outer annulus of width of only 29% the original radius to produce the entire effect. Microscopic observation of the external diaphragm surface of the microphone showed no such adhesive flow. However, we are not able to observe the diaphragm underside in order to draw any conclusion.

Finally, regarding crystal anisotropy, we have not yet developed an expression equivalent to Eq. (2) for this case. We are thus unable at this time to calculate the magnitude of this effect. We note, however, that the variation in E and σ for $\langle 100 \rangle$ and $\langle 111 \rangle$ silicon wafers as measured by Franca and Blouin (2004) is $\sim$36% and 25%, respectively; and since the sensitivity is proportional to E^{-1} and to $(1-\sigma^2)$, it does not seem likely that anisotropic elastic properties will account for the observed sensitivity decrease.

Further work will be carried out in order to determine to what extent each of these effects [numbered (1)–(4) and (6)] is in fact contributing to the factor of 4 decrease of the microphone dynamic sensitivity.

Next we discuss the minimum detectable pressure P_{min} defined to be the acoustic level at which the S/N per $\sqrt{Hz}$ is one. The intrinsic noise level, which is determined for our sensor by the LED light and power source (with 150 μW on the detector), can be seen in Fig. 6 to be $\sim$17 pW/$\sqrt{Hz}$ above 30 Hz and is nearly independent of frequency. From Fig. 5 we can see that at 5 kHz the acoustic sensitivity is

measured to be −152 dB re: W/Pa. This gives a minimum detectable pressure of −215.4 dB re:W/$\sqrt{\text{Hz}}$+152 dB re: W/Pa=−63.4 dB re: Pa/$\sqrt{\text{Hz}}$ or 680 μPa/$\sqrt{\text{Hz}}$. (We also note that this level is an order of magnitude higher that the ambient acoustic levels in the quiet room.)

The fact that the intrinsic noise level of the microphone is determined by the LED/power source was established by noting that the same noise level was observed when the LED/detector system was monitored before integration into the microphone (using the same 150 μW optical power level) and that the detector specified noise level is much lower. The current microphone uses a relatively inexpensive LED, and we note that more expensive light sources are available with significantly lower noise levels. How far below the 680 μPa/$\sqrt{\text{Hz}}$ threshold level one could achieve with this replacement depends upon the other noise mechanisms at play in the microphone [see, for example, Zuckerwar *et al.* (2003) and He and Cuomo (1992)]. These noise sources include detector noise, mechanical noise associated with thermally generated diaphragm vibration, and photon noise.

This last noise source, which is the fundamental limit regarding the optical components of our microphone, can be estimated straightforwardly. In this case, over the band Δf the mean square noise current $\langle i^2 \rangle$ is given by $2eD\,W_{\text{op}}\Delta f$, with e the electronic charge. We then have

$$P_{\min} = \sqrt{2e\Delta f \frac{W_{\text{op}}}{D}} \bigg/ \left(\frac{\partial W_{\text{op}}}{\partial P} \right), \qquad (6)$$

which for our parameters gives $P_{\min}=360$ μPa/$\sqrt{\text{Hz}}$.

As can be seen, the limiting LED noise is not far above the photon noise limit for our power level on the detector, i.e., 150 μW. However, were we able to attain the theoretical diaphragm sensitivity (with the additional factor of 4), the minimum detectable pressure would decrease to 90 μPa. In addition, higher power and lower noise LEDs are available, and with a source having 10 times the power (and noise ≤ shot noise limit), $P_{\min}$ could reach 28 μPa.

Ultimately, background diaphragm motion will determine whether these levels can be attained. We have estimated the levels of thermally generated diaphragm noise associated with the moderate amount of internal damping of the diaphragm and its attachments as manifested in the Q of the first resonant peak. This is simply estimated using Nyquist's relation [Callen and Welton (1951)] as discussed by Gabrielson (1993) which gives $A^{-1}\sqrt{4k_B T\omega_0 m\Delta f/Q}$ for the noise equivalent pressure, where A and m are the diaphragm area and mass, k_B is Boltzmann's constant, T the temperature, and ω_0 the resonance frequency. For our parameters, we find that this yields noise pressures less than 1 μPa/$\sqrt{\text{Hz}}$, which are well below the noise levels discussed earlier.

The current design demonstrates an inexpensive, miniature fiber optic microphone with a threshold detectability 24 times lower and a diameter a factor of 2 smaller than its progenitor device reported by Bucaro and Lagakos (2001). Its limiting noise is that associated with the LED source, whose power spectrum is nearly flat (versus $1/f$) down to very low frequencies (∼0.01 Hz). This leads to an inexpensive, broadband, sensitive microphone covering frequencies from about 20 kHz down below subhertz. We also note that if one could achieve the theoretical diaphragm sensitivity and employ higher power/lower noise LEDs, threshold detectabilities would approach within an order of magnitude that of the best conventional device we could identify, namely the 25-mm-diam Bruel & Kjaer #4146 [Gabrielson (1993)].

The fiber optic microphone threshold detectabilities identified earlier—680 μPa/$\sqrt{\text{Hz}}$ achieved in the present design, 90 μPa/$\sqrt{\text{Hz}}$ were we to achieve the diaphragm sensitivity predicted by Eq. (2), and 28 μPa/$\sqrt{\text{Hz}}$ for the higher power/low noise source and optimized diaphragm—approach and attain, respectively, those levels of performance achieved by some of the best broadband conventional microphones. However, unlike these conventional high performance counterparts, large expensive preamps connected directly to the fiber optic microphone head are not required. In addition, the current fiber optic microphone (2.8 mm diameter and 13 mm length) is significantly smaller than most of the other high performance devices, and it uses inexpensive optical and electrical components whose total costs can be well below $100.

The high performance fiber optic microphone reported here can be expected to find application in many areas. However, this is especially true where a nonelectrical, miniature, light-weight, and/or inexpensive sensor head is required. In this regard, perhaps the highest payoff of this technology will be in the design and implementation of microphone arrays where sensor head and cabling size, cost, and acoustic cross section are most important. We also note that in such an array a single light source can be shared by a number of microphones, N_{mic}, without drastically impacting the threshold sensitivity achieved in the present device. In this case, the minimum detectable pressure for each microphone would be increased by the factor $\sqrt{N_{\text{mic}}}$.

ACKNOWLEDGMENTS

Bickley, W. G. (**1933**). "Deflections and vibrations of a circular elastic plate under tensions," Philos. Mag. **15**, 776–797.

Bucaro, J. A., and Lagakos, N. (**2001**). "Lightweight fiber optic microphones and accelerometers," Rev. Sci. Instrum. **72**, 2816–2821.

Callen, H., and Welton, T. (**1951**). "Irreversibility and generalized noise," Phys. Rev. **83**, 34–40.

Crandall, I. B. (**1918**). "The air damped vibrating system: Theoretical calibration of condenser transmitter," Phys. Rev. **11**, 449–460.

Franca, D. R., and Blouin, A. (**2004**). "All optical measurement of in-plane and out-of-plane Young's modulus and Poisson's ratio in silicon wafers by means of vibration modes," Meas. Sci. Technol. **15**, 859–868.

Gabrielson, T. B. (**1993**). "Mechanical thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**, 903–909.

Hansen, S. T., Turo, A., Degertekin, F. L., and Khuri-Yakib, B. T. (**2000**). "Characterization of capacitive micromachined ultrasonic transducers in air using optical measurements," *2000 IEEE Ultrasonics Symposium*, pp.

1412   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Bucaro *et al.*: Miniature fiber optic microphone

947–950.

He, G., and Cuomo, F. W. (**1991**). "Displacement response, detection limit, and dynamic range of fiber-optic lever sensors," J. Lightwave Technol. **9**, 1618–1625.

He, G., and Cuomo, F. W. (**1992**). "The analysis of noises in a fiber optic microphone," J. Acoust. Soc. Am. **92**, 2521–2526.

Hu, A, Cuomo, F. W., and Zuckerwar, A. J. (**1992**). "Theoretical and experimental study of a fiber optic microphone," J. Acoust. Soc. Am. **91**, 3049–3056.

Lagakos, N., and Bucaro, J. A. (**2001**). "Fiber optic pressure and acceleration sensors." *Proceedings of the 47th International Instr. Symposium Denver*, CO, 6–10 May.

Scheeper, P. R., Olthius, W., and Bergveld, P. (**1994**). "Improvement of the performance of microphones with a silicon nitride diaphragm and backplate," Sens. Actuators, A **40**, 179–186.

Soedel, W. (**1981**). *Vibrations of Shells and Plates*, (Dekker, New York), pp 99–104.

Zuckerwar, J. (**1978**). "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285.

Zuckerwar, A. J., Cuomo, F. W., Nguyen, T. D., Rizzi, S. A., and Clevenson, S. A. (**1995**). "High temperature fiber-optic lever microphone," J. Acoust. Soc. Am. **97**, 3605–3616.

Zuckerwar, A. J., Kahn. T. R., and Serbyn, R. M. (**2003**). "Background noise in piezoelectric condenser, and ceramic microphone," J. Acoust. Soc. Am. **113**, 3179–3187.

Properties of transducers and substrates for high frequency resonators and sensors

R. H. Wittstruck, N. W. Emanetoglu, and Y. Lu
Department of Electrical and Computer Engineering, Rutgers University, Piscataway, New Jersey 08854

Sally Laffey
General Technical Services Corporation, Wall, New Jersey 07719

Arthur Ballato[a]
US Army Communications-Electronics Command, Fort Monmouth, New Jersey 07703

Properties of transducers and substrates for bulk acoustic wave resonators and sensors are described. These resonators utilize one-dimensional thickness vibrations of structures consisting of a low-loss substrate crystal surmounted by a thin active piezoelectric film that drives the composite in resonant modes to achieve gigahertz frequencies. The structures considered include oblique orientations of the substrate, leading to generation of coupled elastic modes in the composite. A modified Christoffel-Bechmann (CB) formalism is presented to calculate acoustic wave speeds and displacements in the piezoelectric film transducer and the substrate. The CB method also yields the piezoelectric coupling coefficients of arbitrarily oriented piezofilms, for electric fields impressed either along the thickness or laterally. The calculations apply generally to transducer and substrate crystals of any symmetry class. The piezoelectric portion is then made specific for films of class 6 mm (wurtzite structure) with arbitrary orientation on the substrate, and the substrate calculations are specified for class $\bar{3}m$ materials, but apply also to any substrate with $3m$ or 32 symmetry. Zinc oxide and sapphire are used in an example of the acoustic resonator structure. [DOI: 10.1121/1.1985076]

I. INTRODUCTION

Acoustic resonators, filters, and actuators are key enabling components which find numerous applications in communications and in sensors. These frequency control and selection components are members of a class of generic network elements that possess the superiority of low acoustic loss, small size, and attractive thermal characteristics.

The quartz resonator, introduced in the 1920s, is the prototypic element of this class. With perhaps the single exception of the 32 kHz wristwatch crystal, applications have steadily pushed upper frequency limits, primarily to achieve higher system bandwidths. In the 1940s, 30 MHz was considered the industrial limit. Today, applications demand gigahertz capabilities. To achieve these frequencies, acoustic resonances utilize thickness vibrations of thin films and plates. These bulk acoustic wave (BAW) devices are capable of handling relatively high levels of power, with low noise performance. At the upper frequency ranges, composite structures are commonly used. These consist of a thin active piezoelectric layer deposited on a robust, high acoustic quality factor (Q) substrate, and are often referred to as HBAWs (high frequency BAWs),[1] or thin-film resonators (TFRs).[2]

TFR substrate candidates include silicon, diamond, quartz, lithium niobate, gallium arsenide, and sapphire (corundum). Each of these materials possesses beneficial characteristics. Among them sapphire has extremely low acoustic loss, high thermal conductivity, and is readily available commercially.

The focus of this paper is the development of a modified Christoffel-Bechmann (CB) formalism to calculate acoustic wave velocities, displacements, and piezoelectric coupling of arbitrarily oriented composite transducer structures, composed of piezoelectric thin films on a substrate. ZnO on sapphire structures are used as the example, however, the calculations apply generally to transducer and substrate crystals of any symmetry class.

II. CHRISTOFFEL-BECHMANN METHOD

A. Overview

We treat the case of simple thickness modes of homogeneous, but arbitrarily anisotropic thin plates, films, and composites driven piezoelectrically. This is a one-dimensional approximation, with no lateral variations in motion. It has proven to be adequate in many practical situations, and permits most salient features to be simply described. General background is found in Ref. 3.

There are two methods in general usage for treating the case of plane acoustic wave propagation in crystals, from which the solution to this problem is composed. In the plate-coordinate method, the relevant elastic, piezoelectric, and dielectric material coefficients are transformed to the rotated coordinate system specifying the orientation of the film or plate. In the CB method, the transformation involves instead

---

[a]Corresponding author; electronic mail: arthur.ballato@us.army.mil

Plane-Coordinate Method

Christoffel-Bechmann Method

plane wave phase front
(specified by φ,θ)
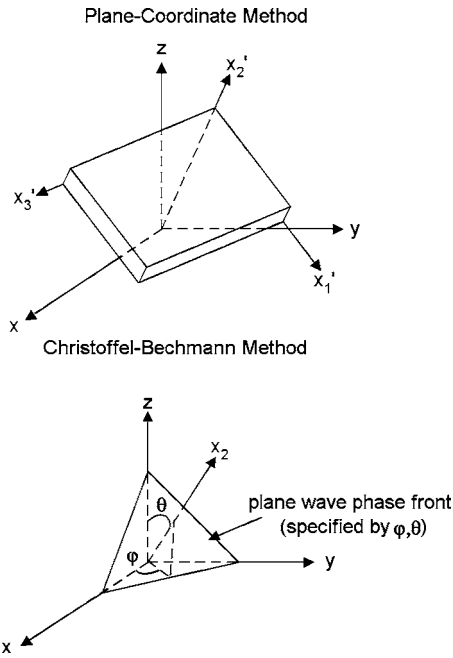
FIG. 1. Schematic representations of the rotations used for the plate-coordinate and Christoffel-Bechmann methods. In the plate-coordinate method (x,y,z) are the crystallographic axes, (x'_1,x'_2,x'_3) are the plate axes, and the (x'_1,x'_3) plane coincides with the phase front. In the Christoffel-Bechmann method, (x,y,z) are the crystallographic axes, x_2 is the propagation direction, and the plane wave phase front is specified by φ and θ.

only the coordinate representing the direction of propagation of the plane waves with respect to the crystallographic axes. It is therefore simpler than the first method. Figure 1 is a schematic representation of the rotations involved in these two methods. Either method yields the wave velocities that determine the resonance frequencies, the directions of particle motion for each mode, and the usual thickness excitation piezoelectric coupling coefficients. When lateral excitation is used, the electric field is imposed along a transverse axis, and additional coefficients are required to be calculated. The procedure for doing this has been known using the plate-coordinate method. Here we provide the computationally simpler CB method to encompass lateral excitation, and obtain therefrom the lateral piezocoupling coefficients for oblique orientations of films and plates. These will be specialized in subsequent sections for particular crystal classes to classify the acoustic behavior of materials for the purposes of analyzing device performance.

B. Christoffel-Bechmann elastic wave formalism

George Green, in an 1839 paper, derived equations containing 21 coefficients that today we recognize as those governing acoustic plane wave propagation in crystals. Christoffel, in 1877, published a long article in which he cast the equations in terms of strain instead of stress, and gave them a form that is still used. Cady introduced piezoelectric effects,[4] and Bechmann used them to augment Christoffel's method.[5] Here we sketch the CB method, to include the case where the electric field driving the motion is lateral to the

direction of the advancing phase front. We then apply the extended method to the simple thickness modes of films with arbitrary piezoelectric symmetry.

The method starts by assuming that acoustic plane waves are propagating in a specified direction with respect to the crystallographic axes of the material. If such waves are subsequently confined as standing waves between the major surfaces of a thin plate or film, one refers to the ensuing motion as simple thickness modes. We adopt the IEEE Standards[6,7] defining both the crystallographic axial set, and the notation $(YXwlt)\varphi/\theta/\psi$ specifying the orientation of a film or plate aligned with a rotated coordinate set of axes. Thickness is taken along the rotated X_2 axis, which is the direction of plane wave progression. The angle set φ, θ, and ψ specifying the three rotations leads to a rotation matrix (α). Normally, the film is piezoelectrically driven by electrodes placed on the major surfaces, creating an electric field in the thickness direction; this is called thickness excitation (TE). When the exciting field lies in the plane of the film, the situation is referred to as lateral excitation (LE). We assume that the lateral electric field is impressed along the rotated X_1 axis.

With the specification of φ, θ, and ψ, elements of the matrix (α) are determined as

$$\alpha_{11} = [\cos\varphi\cos\psi - \sin\varphi\sin\theta\sin\psi], \tag{1}$$

$$\alpha_{12} = [\sin\varphi\cos\psi + \cos\varphi\sin\theta\sin\psi], \tag{2}$$

$$\alpha_{13} = [-\cos\theta\sin\psi], \tag{3}$$

$$\alpha_{21} = [-\sin\varphi\cos\theta], \tag{4}$$

$$\alpha_{22} = [\cos\varphi\cos\theta], \tag{5}$$

$$\alpha_{23} = [\sin\theta], \tag{6}$$

$$\alpha_{31} = [\cos\varphi\sin\psi + \sin\varphi\sin\theta\cos\psi], \tag{7}$$

$$\alpha_{32} = [\sin\varphi\sin\psi - \cos\varphi\sin\theta\cos\psi], \tag{8}$$

$$\alpha_{33} = [\cos\theta\cos\psi]. \tag{9}$$

The CB method takes $\alpha_n = \alpha_{2n}$, the elements of the second row of (α), and $\zeta_n = \alpha_{1n}$, the elements of the first row of (α). Then, using the full tensor forms of the elastic, piezoelectric, and dielectric constants, the following quantities are formed:

$$\Gamma_{jk} = c_{ijkp}\alpha_i\alpha_p, \tag{10}$$

$$\Xi_k = e_{nkp}\alpha_n\alpha_p, \tag{11}$$

$$\varepsilon'_{22} = \varepsilon_{in}\alpha_i\alpha_n, \tag{12}$$

$$\Gamma^*_{jk} = \Gamma_{jk} + \Xi_j\Xi_k/\varepsilon'_{22}. \tag{13}$$

The second term in Eq. (13) is the piezoelectric stiffening term. The quantities appropriate to LE are now formed:

$$\Lambda_k = e_{nkp}\zeta_n\alpha_p, \tag{14}$$

$$\varepsilon'_{12} = \varepsilon_{in}\zeta_i\alpha_n, \tag{15}$$

$$\varepsilon'_{11} = \varepsilon_{in}\zeta_i\zeta_n. \tag{16}$$

The stiffnesses c_{ijkp} are those at constant electric field, c^E_{ijkp}, and the dielectric permittivities ε_{ij} are those at constant elastic strain, ε^S_{ij}. Where these quantities appear in the following, the standard Voigt contraction for the indices has been applied.

Next, the eigenvalue equation

$$|\Gamma^*_{jk} - c\,\delta_{jk}| = 0 \tag{17}$$

is solved. It determines three piezoelectrically stiffened eigenstiffnesses, c_m, yielding the elastic velocities $v_m = (c_m/\rho)^{1/2}$, and corresponding eigenvectors $\gamma^{(m)}$, specifying the directions of particle displacement for each mode m, with respect to the crystallographic axes. By tradition, $m=a,b,c$, corresponds, respectively, to the quasi-extensional, the fast-, and slow-quasi-shear modes. Apart from the wave speeds, v_m, that determine the resonance frequencies of the film, and the angles

$$\delta_m = \cos^{-1}(\gamma^{(m)}_k\alpha_k) \tag{18}$$

that describe the deviations of the particle motions with respect to the propagation axis, the most important practical quantities associated with the vibrational modes are the values of the piezoelectric coupling factors, $k(\text{TE})_m$, and $k(\text{LE})_m$. These are obtained as follows. For the TE case:

$$e^0_m(\text{TE}) = \gamma^{(m)}_n\Xi_n, \tag{19}$$

$$k_m(\text{TE}) = |e^0_m(\text{TE})|/(\varepsilon'_{22}c_m)^{1/2}. \tag{20}$$

For the LE case:[2]

$$e^0_m(\text{LE}) = \gamma^{(m)}_n\Lambda_n, \tag{21}$$

$$e^{00}_m(\text{LE}) = e^0_m(\text{LE}) - \{\varepsilon'_{12}/\varepsilon'_{22}\}e^0_m(\text{TE}) \tag{22}$$

$$= \gamma^{(m)}_n\{\Lambda_n - (\varepsilon'_{12}/\varepsilon'_{22})\Xi_n\}, \tag{23}$$

$$\varepsilon^0_{11} = [\varepsilon'_{11} - (\varepsilon'_{12})^2/\varepsilon'_{22}], \tag{24}$$

$$k_m(\text{LE}) = |e^{00}_m(\text{LE})|/(\varepsilon^0_{11}c_m)^{1/2}. \tag{25}$$

III. MATERIALS IN CLASS 6mm

A. General

For the piezoelectric film component of high frequency bulk acoustic wave resonator (HBARs), materials of symmetry group 6mm represent attractive choices, because of their strong piezo-coupling factors and low cost. Materials belonging to symmetry class 6mm (wurtzite crystal structure) include zinc oxide (ZnO) and members of the lead zirconate titanate (PZT) family of polarized ferroelectric ceramics. Other representatives include barium titanate, and some binary semiconductors, e.g., AlN, GaN, and InN in the III-V system, and BeO, α-ZnS, CdS, and CdSe in the II-VI system.[8] The Van Dyke matrix[9] for class 6mm is given in Table I. For this class, the angle φ is irrelevant, so the X cut

TABLE I. Van Dyke matrix for symmetry class 6mm. ($c_{12}=c_{11}-2c_{66}$; $[c]$ is $[c^E]$; $[\varepsilon]$ is $[\varepsilon^S]$.)

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 & 0 & 0 & e_{31} \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 & 0 & 0 & e_{31} \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 & 0 & 0 & e_{33} \\ 0 & 0 & 0 & c_{44} & 0 & 0 & 0 & e_{15} & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 & e_{15} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & e_{15} & 0 & \varepsilon_{11} & 0 & 0 \\ 0 & 0 & 0 & e_{15} & 0 & 0 & 0 & \varepsilon_{11} & 0 \\ e_{31} & e_{31} & e_{33} & 0 & 0 & 0 & 0 & 0 & \varepsilon_{33} \end{bmatrix}$$

(film normal along X_1) is equivalent to the Y cut (film normal along X_2), and in Eqs. (14)–(16), the ζ_k are, in general:

$$\zeta_1 = \cos\psi, \tag{26}$$

$$\zeta_2 = \sin\theta\sin\psi, \tag{27}$$

$$\zeta_3 = -\cos\theta\sin\psi. \tag{28}$$

Thin-film resonators are most often fashioned with the polar direction along the film normal, producing a thickness-stretch (extensional) motion. Sometimes the polar axis is made to lie in the plane of the film, resulting in a thickness-shear mode of vibration. In both cases, the driving voltage is impressed across two electrodes placed on the major surfaces. This situation can be generalized in two simple ways, by permitting both the polar axis and the exciting electric field directions to be non-collinear, and to be oblique to the thickness. There have been few reports of transducers with oblique orientation because of perceived difficulties in manufacture. New materials and technologies can be expected to ease this impediment in the future. Alternative electroding arrangements can provide oblique driving fields.

The point group symmetry 6mm results in several simplifications of the general case treated in the last section, even when the polar axis is oblique. First, the gamut of allowable orientations is limited. For the most general type of crystal, the one-dimensional vibrational characteristics follow from specification of the two angles that determine a plane in space. For 6mm symmetry, however, the basal plane is isotropic, so the angle of longitude is irrelevant, and only the single angle of latitude is meaningful. ("Latitude" in this usage refers to the angle between the film normal and the direction of the polar axis.) Second, one of the shear modes always remains a pure mode, regardless of angle, that is, it is not coupled to any other motion; the other two modes are generally coupled together.

Regarding the direction of the electric field driving the motion, the traditional electroding scheme, where the electrodes are placed on the film major surfaces, is referred to as "thickness excitation" because the resulting field points along the film thickness. One may alternatively configure the electrode system to the sides, producing a lateral field in the plane of the film. Lateral excitation is more commonly used for driving g bars than plates or films. By a combination of

TABLE II. Effective elastic stiffnesses for thickness modes for crystals of symmetry class 6mm.

Mode	Y cut	Rotated Y cut	Z cut
	All modes pure	Pure mode assumed b	All modes pure
a	c_{11}^E	$\bar{c}_a$	$\bar{c}_{33}=c_{33}^E+(e_{33})^2/\varepsilon_{33}$
b	c_{66}^E	$c_b^E=c_{66}^E\cos^2\theta+c_{44}^E\sin^2\theta$	c_{44}^E degenerate mode
c	$\bar{c}_{44}=c_{44}^E+(e_{15})^2/\varepsilon_{11}$	$\bar{c}_c$	c_{44}^E degenerate mode

TABLE III. Piezocoupling factors for thickness modes of Y-cut crystals of symmetry class 6 mm.

Mode	TE	LE(E_1)	LE(E_3)	LE(ψ)
a	0	0	$|e_{31}|\,(\varepsilon_{33}c_{11}^E)^{-1/2}$	$|e_{31}\sin\psi|\,(\varepsilon_{11}^o c_{11}^E)^{-1/2}$
b	0	0	0	0
c	$|e_{15}|\,(\varepsilon_{11}\bar{c}_{44})^{-1/2}$	0	0	0

electroding arrangements, it is possible to produce a composite driving electric field that is neither parallel nor perpendicular to the film normal. The analysis, however, shows that the problem must be broken down into the two canonical forms of TE and LE, and each is treated separately, because the effective material coefficients involved are somewhat different for each form.

B. Method of CB applied to the wurtzite structure

Many III-V and II-VI crystals have point group symmetry 6mm, the wurtzite structure. The piezoelectric matrix contains three independent coefficients, e_{15}, e_{31}, and e_{33}. Orientations in this symmetry class are limited in their variety to a single rotation about any axis normal to the polar axis. A synopsis of the behavior of films or plates of class 6 mm material piezoelectrically driven in simple thickness modes is given in the following. The polar axis is denoted P; P' is the direction of projection of P onto the plane of the film, and $\tilde{n}$ is the direction of the film normal.

For P parallel to the film normal (Z cut), all modes are pure, with the shears degenerate. TE drives only the extensional mode. LE drives only the shear mode, and displacement is along the LE field direction.

For P in the plane of the film (Y cut), all three modes are pure, with one shear mode having displacement along P, and the other having displacement in the plane, but normal to P. TE drives only the shear mode having motion along P; its effective stiffness is $\bar{c}_{44}=c_{44}^E+e_{15}^2/\varepsilon_{11}$. LE drives only the extensional mode, providing there is a component of E_{lateral} along P; its effective stiffness is c_{11}^E.

When the unique axis P is neither parallel nor perpendicular to the film normal, one shear mode remains pure, and inert to TE (polarized perpendicular to both P and to the film normal). It may be either the "b" or "c" mode, depending on the material and angle. The other two modes are coupled, and TE drives both. All three modes, the pure shear and the two coupled modes, are driven by LE if E_{lateral} is neither normal nor parallel to P'. If E_{lateral} is normal to P', only the pure shear mode, polarized normal to P' (and therefore along E_{lateral}), is driven by LE. If E_{lateral} is parallel to P', only the two coupled modes are driven by LE (since they both have components of displacement along E_{lateral}), and the motions are normal to that of the inert pure shear mode, i.e., in the plane containing P' and the film normal.

For a rotated cut, with P oblique to the film normal, the pure shear perpendicular to both P' and $\tilde{n}$ is driven by LE (for a rotated Y cut, the required E is E_1); although k(LE) goes to zero at the Y cut. At the Y cut ($\theta=0°$) the stiffness is

c_{66}^E; for the Z cut, it is c_{44}^E. For θ between 0° and 90°, the stiffness for this pure mode transforms as $(c_{66}^E\cos^2\theta+c_{44}^E\sin^2\theta)$.

The coupled shear and extensional modes (in the plane of P' and $\tilde{n}$), for a rotated cut, are driven by LE (for a rotated Y cut, the required E is E_3'). At the Y cut, the mode coupling ceases, and k(LE) drives the a mode only. At the Z cut, the mode coupling likewise ceases, but now k(LE) drives the degenerate shear polarized normal to P with isotropic azimuthal piezocoupling; the displacement is along the applied field. At the Y cut, the a mode stiffness is c_{11}^E; for the Z cut, it is $\bar{c}_{33}=c_{33}^E+e_{33}^2/\varepsilon_{33}$. For θ between 0° and 90°, the stiffness for this mode transforms in a somewhat complicated manner.

The above-mentioned results are detailed in the following tables. Table II gives the effective elastic stiffnesses for the three modes of Y-cut, rotated-Y-cut, and Z-cut, class 6 mm piezoelectrics. Tables III–V give the piezocoupling factors for the three modes of Y-cut, Z-cut, and rotated-Y-cut, class 6 mm piezoelectrics, respectively. In Tables III and V, the additional abbreviations and relations hold:

$$e_a^0(\text{TE}) = [\gamma_2^{(a)}\Xi_2 + \gamma_3^{(a)}\Xi_3], \tag{29}$$

$$e_a^0(\text{LE}) = [\gamma_2^{(a)}\Lambda_2 + \gamma_3^{(a)}\Lambda_3], \tag{30}$$

$$e_c^0(\text{TE}) = [\gamma_2^{(c)}\Xi_2 + \gamma_3^{(c)}\Xi_3], \tag{31}$$

$$e_c^0(\text{LE}) = [\gamma_2^{(c)}\Lambda_2 + \gamma_3^{(c)}\Lambda_3], \tag{32}$$

where

$$\Xi_2 = (e_{15} + e_{31})\cos\theta\sin\theta, \tag{33}$$

$$\Xi_3 = [e_{15}\cos^2\theta + e_{33}\sin^2\theta], \tag{34}$$

$$\Lambda_2 = [e_{15}\sin^2\theta - e_{31}\cos^2\theta]\sin\psi, \tag{35}$$

$$\Lambda_3 = (e_{15} - e_{33})\cos\theta\sin\theta\sin\psi, \tag{36}$$

$$|e_a^{00}(\text{LE})| = |e_a^0(\text{LE}) - (\varepsilon_{12}'/\varepsilon_{22}')e_a^0(\text{TE})|$$
$$= |(\gamma_2^{(a)}\Lambda_2 + \gamma_3^{(a)}\Lambda_3) - (\varepsilon_{12}'/\varepsilon_{22}')(\gamma_2^{(a)}\Xi_2 + \gamma_3^{(a)}\Xi_3)|, \tag{37}$$

TABLE IV. Piezocoupling factors for thickness modes of Z-cut crystals of symmetry class 6mm.

Mode	TE	LE(E_1)	LE(E_3)	LE(ψ)
a	$|e_{33}|\,(\varepsilon_{33}\bar{c}_{33})^{-1/2}$	0	0	0
b, c	0	$|e_{15}|\,(\varepsilon_{11}c_{44}^E)^{-1/2}$	$|e_{15}|\,(\varepsilon_{11}c_{44}^E)^{-1/2}$	$|e_{15}|\,(\varepsilon_{11}c_{44}^E)^{-1/2}$

TABLE V. Piezocoupling factors for thickness modes of rotated-Y-cut crystals of symmetry class $6mm$.

Mode	TE	LE(E_1)	LE(E_3')	LE(ψ)
a	$\lvert e_a^0 \rvert \, (\varepsilon_{22}' \bar{c}_a)^{-1/2}$	0	$\lvert e_a^{oo} \rvert \, (\varepsilon_{11}^o \bar{c}_a)^{-1/2}$	$\lvert e_a^{oo} \rvert \, (\varepsilon_{11}^o \bar{c}_a)^{-1/2}$
b	0	$\lvert e_{15} \sin\theta \rvert \, (\varepsilon_{11} c_b^E)^{-1/2}$	0	$\lvert e_{15} \sin\theta \cos\Psi \rvert \, (\varepsilon_{11}^o c_b^E)^{-1/2}$
c	$\lvert e_c^0 \rvert \, (\varepsilon_{22}' \bar{c}_c)^{-1/2}$	0	$\lvert e_c^{oo} \rvert \, (\varepsilon_{11}^o \bar{c}_c)^{-1/2}$	$\lvert e_c^{oo} \rvert \, (\varepsilon_{11}^o \bar{c}_c)^{-1/2}$

$$\lvert e_c^{00}(\mathrm{LE}) \rvert = \lvert e_c^0(\mathrm{LE}) - (\varepsilon_{12}'/\varepsilon_{22}') e_c^0(\mathrm{TE}) \rvert$$
$$= \lvert (\gamma_2^{(c)}\Lambda_2 + \gamma_3^{(c)}\Lambda_3) - (\varepsilon_{12}'/\varepsilon_{22}')(\gamma_2^{(c)}\Xi_2 + \gamma_3^{(c)}\Xi_3) \rvert. \tag{38}$$

For the Y cut, $\gamma^{(a)} = [0,1,0]$ and $\gamma^{(c)} = [0,0,1]$, and for the Z cut, $\gamma^{(a)} = [0,0,1]$ and $\gamma^{(c)} = [0,1,0]$, but for rotated orientations, a two-dimensional eigenvalue problem must be solved. The matrix consists of the elements Γ_{22}^*, Γ_{23}^*, and Γ_{33}^* where

$$\Gamma_{22}^* = [c_{11}\cos^2\theta + c_{44}\sin^2\theta$$
$$+ (e_{15}+e_{31})^2 \cos^2\theta \sin^2\theta/\varepsilon_{22}'], \tag{39}$$

$$\Gamma_{33}^* = [c_{44}\cos^2\theta + c_{33}\sin^2\theta$$
$$+ (e_{15}\cos^2\theta + e_{33}\sin^2\theta)^2/\varepsilon_{22}'], \tag{40}$$

$$\Gamma_{23}^* = [\cos\theta \sin\theta][(c_{13}+c_{44}) + (e_{15}+e_{31})[e_{15}\cos^2\theta$$
$$+ e_{33}\sin^2\theta]/\varepsilon_{22}']. \tag{41}$$

The eigenvalues are $\bar{c}_a$ and $\bar{c}_b$ and

$$\gamma^{(a)} = [0, \gamma_2^{(a)}, \gamma_3^{(a)}], \tag{42}$$

$$\gamma^{(c)} = [0, \gamma_2^{(c)}, \gamma_3^{(c)}] \tag{43}$$

are the eigenvectors.

The rotated permittivities are

$$\varepsilon_{11}' = \varepsilon_{11} + \Delta \cos^2\theta \sin^2\psi, \tag{44}$$

$$\varepsilon_{12}' = -\Delta \cos\theta \sin\theta \sin\psi, \tag{45}$$

$$\varepsilon_{22}' = \varepsilon_{11}\cos^2\theta + \varepsilon_{33}\sin^2\theta, \tag{46}$$

$$\varepsilon_{11}^0 = \varepsilon_{11}' - (\varepsilon_{12}')^2/\varepsilon_{22}'$$
$$= \varepsilon_{11} + \Delta\cos^2\theta\sin^2\psi(1 - \Delta\sin^2\theta/\varepsilon_{22}'), \tag{47}$$

where

$$\Delta = (\varepsilon_{33} - \varepsilon_{11}). \tag{48}$$

For $\varepsilon_{33} \neq \varepsilon_{11}$, ε_{11}^0 is a function of ψ (albeit usually a weak one), so k_m (LE) is not strictly proportional to $\lvert\cos(\psi)\rvert$. If $(\varepsilon_{11}-\varepsilon_{33})=0$, then it is cosinusoidal in ψ; as the difference widens, the departure grows. The ε_{11}^0 variation is independent of considerations of piezoelectricity, and applies to all uniaxial crystals, for rotations in the basal plane.

The considerations given here for crystal class $6mm$ apply equally to members of crystal class $4mm$, which includes some of the newer single-crystal ferroelectrics. Both classes possess identical Van Dyke matrices, with the exception that in $4mm$ c_{12} is an independent constant.

C. Piezoelectric ZnO thin films

Zinc oxide (ZnO) is a wide band gap (3.37 eV) II-VI semiconductor with strong piezocoupling. Through proper doping to achieve high resistivity, it can be made a piezoelectric material with strong piezoelectric coupling coefficients. It can also be made highly conductive to serve as an electrode in certain HBAR configurations.[10] It is capable of being deposited on technologically important substrates as lattice-matched, oriented thin films by a variety of techniques, such as molecular-beam epitaxy, metalorganic chemical vapor deposition, metalorganic vapor phase epitaxy, and various types of sputtering. Measured elastic, piezoelectric, and dielectric values are given in Tables VI(a) and VI(b) for film and bulk growth; Ref. 13 contains ZnO temperature coefficients.

Of particular interest is the capability of growing oriented films of ZnO and its alloys on various facets of crystal sapphire. Depending on the particular application, it may be advantageous to orient the transducer polar axis either normal to the film thickness (appropriate to TE), or in the plane of the film (for LE). For measured orientations of ZnO grown on various planes of sapphire, see Ref. 14, and references therein, as well as Ref. 15.

We have found[10,16] that the alloy $Mg_xZn_{1-x}O$ retains the $6mm$ piezoelectric character for $x < 0.33$ to 0.35, and, for

TABLE VI. (a) Elastic stiffnesses of film and bulk zinc oxide and (b) piezoelectric and dielectric coefficients of film and bulk zinc oxide.

(a)	c_{11}^E (GPa)	c_{33}^E (GPa)	c_{44}^E (GPa)	c_{66}^E (GPa)	c_{13}^E (GPa)	(c_{12}^E) (GPa)	ρ (kg/m^3)	Type	Ref.
	157	208	38	34	83	89	5720	Film	11
	207.0	209.5	44.8	44.65	106.1	117.7	5665	Bulk	12

(b)	e_{15} (C/m^3)	e_{31} (C/m^3)	e_{33} (C/m^3)		ε_{11}^S (pF/m)	ε_{33}^S (pF/m)	Type	Ref.
	-0.45	-0.51	1.22	$\varepsilon^S/\varepsilon_0=$	73.75	78.00	Film	11
	-0.48	-0.57	1.321		8.33	8.81	Bulk	12

TABLE VII. Van Dyke matrix for symmetry class $3m$. ($c_{12}=c_{11}-2c_{66}$; $[c]$ is $[c^E]$; $[\varepsilon]$ is $[\varepsilon^S]$).

$$
\begin{bmatrix}
c_{11} & c_{12} & c_{13} & c_{14} & 0 & 0 & 0 & -e_{22} & e_{31} \\
c_{12} & c_{11} & c_{13} & -c_{14} & 0 & 0 & 0 & e_{22} & e_{31} \\
c_{13} & c_{13} & c_{33} & 0 & 0 & 0 & 0 & 0 & e_{33} \\
c_{14} & -c_{14} & 0 & c_{44} & 0 & 0 & 0 & e_{15} & 0 \\
0 & 0 & 0 & 0 & c_{44} & c_{14} & e_{15} & 0 & 0 \\
0 & 0 & 0 & 0 & c_{14} & c_{66} & -e_{22} & 0 & 0 \\
0 & 0 & 0 & 0 & e_{15} & -e_{22} & \varepsilon_{11} & 0 & 0 \\
-e_{22} & e_{22} & 0 & e_{15} & 0 & 0 & 0 & \varepsilon_{11} & 0 \\
e_{31} & e_{31} & e_{33} & 0 & 0 & 0 & 0 & 0 & \varepsilon_{33}
\end{bmatrix}
$$

greater x values, transforms to the nonpiezoelectric cubic rocksalt point group $m3m$ characteristic of silicon and diamond. Moreover, depending on the facet upon which the film is grown, this new ternary composition allows either TE and LE to be utilized, with higher acoustic velocities, and only a slight diminution of the piezocoupling, from the pure ZnO value. Material constants for $Mg_xZn_{1-x}O$ may be adequately calculated by transforming the constants of MgO from the rocksalt to the wurtzite phase, and using Vegard's law (parameters of substitutional solid solutions vary linearly between the values for the components) to combine the wurtzite ZnO and MgO constants.[10] In addition to ZnO and $Mg_xZn_{1-x}O$ compositions, the $6mm$ nitrides, such as GaN[17] and AlN, show considerable future promise as piezoelectric transducers for HBAR/TFR devices.

IV. MATERIALS IN CLASS $3m$

In this section we address some acoustic properties of materials in symmetry class $3m$, and apply the CB formalism to Y cuts of the strongly piezoelectric refractory oxides lithium niobate and lithium tantalate. The Van Dyke matrix for class $3m$ is shown in Table VII. Material values for $LiNbO_3$ and $LiTaO_3$ are found in Ref. 18; pertinent subsets are listed in Tables VIII(a) and VIII(b).

According to the IEEE orientational definitions, a Y cut is specified by $(YX1)\theta=0$, and $\alpha_2=1$. With the matrix elements from Table VII inserted in Eqs. (10)–(13), along with the $LiNbO_3$ values from Tables VIII(a) and VIII(b), the piezoelectrically stiffened gammas are

$$\Gamma_{11}^* = \Gamma_{11} + (\Xi_1)^2/\varepsilon_{11}^S = c_{66}^E = 72.09 \text{ GPa}, \qquad (49)$$

$$\Gamma_{22}^* = \Gamma_{22} + (\Xi_2)^2/\varepsilon_{11}^S = c_{11}^E + (e_{22})^2/\varepsilon_{11}^S = 213.43 \text{ GPa}, \qquad (50)$$

$$\Gamma_{33}^* = \Gamma_{33} + (\Xi_3)^2/\varepsilon_{11}^S = c_{44}^E + (e_{15})^2/\varepsilon_{11}^S = 93.45 \text{ GPa}, \qquad (51)$$

$$\Gamma_{23}^* = \Gamma_{23} + (\Xi_2\Xi_3)/\varepsilon_{11}^S = -c_{14}^E + (e_{22}e_{15})/\varepsilon_{11}^S$$
$$= 14.30 \text{ GPa}, \qquad (52)$$

$$\Gamma_{13}^* = \Gamma_{13} + (\Xi_1\Xi_3)/\varepsilon_{11}^S = 0, \qquad (53)$$

$$\Gamma_{12}^* = \Gamma_{12} + (\Xi_2\Xi_1)/\varepsilon_{11}^S = 0. \qquad (54)$$

Setting the Eq. (17) determinant

$$|\Gamma_{jk}^* - c\delta_{jk}| = \begin{vmatrix} (\Gamma_{11}^* - c) & 0 & 0 \\ 0 & (\Gamma_{22}^* - c) & \Gamma_{23}^* \\ 0 & \Gamma_{23}^* & (\Gamma_{33}^* - c) \end{vmatrix} \qquad (55)$$

equal to zero yields the stiffness eigenvalues. By inspection, one root is $c=c_{66}^E$; this is the elastic stiffness of the pure shear mode. The corresponding velocity is

$$v_{\text{pure-shear}} = (c_{\text{pure-shear}}/\rho)^{1/2} = (c_{66}^E/\rho)^{1/2} = 3940.6 \text{ m/s}. \qquad (56)$$

The rest of the determinant gives the quadratic for the effective stiffnesses of the quasi-shear and quasi-extensional modes:

$$(\Gamma_{22}^* - c)(\Gamma_{33}^* - c) - (\Gamma_{23}^*)^2 = 0, \qquad (57)$$

yielding

$$c_{\text{quasi-ext}} = 215.11 \text{ GPa}, \qquad (58)$$

$$c_{\text{quasi-shear}} = 91.77 \text{ GPa}, \qquad (59)$$

and

$$v_{\text{quasi-ext}} = (c_{\text{quasi-ext}}/\rho)^{1/2} = 6806.8 \text{ m/s}, \qquad (60)$$

$$v_{\text{quasi-shear}} = (c_{\text{quasi-shear}}/\rho)^{1/2} = 4446.0 \text{ m/s}. \qquad (61)$$

The piezocoupling calculations follow from Eqs. (18) to (25), and are listed in Table IX for both TE and LE excitations of Y-cut lithium niobate and lithium tantalate.

TABLE VIII. (a) Elastic stiffnesses of lithium niobate and lithium tantalate and (b) piezoelectric and dielectric coefficients of lithium niobate and lithium tantalate (Ref. 18).

(a)	c_{11}^E (GPa)	c_{33}^E (GPa)	c_{44}^E (GPa)	c_{66}^E (GPa)	c_{13}^E (GPa)	c_{14}^E (GPa)	(c_{12}^E) (GPa)	ρ (kg/m³)
$LiNbO_3$	198.86	234.18	59.85	72.09	67.99	7.83	54.68	4642.8
$LiTaO_3$	233.05	275.22	95.26	93.31	83.46	−10.75	46.43	7460.4

(b)	e_{15} (C/m²)	e_{22} (C/m²)	e_{31} (C/m²)	e_{33} (C/m²)	ε_{11}^S (pF/m)	ε_{33}^S (pF/m)
$LiNbO_3$	3.655	2.407	0.328	1.894	397.6	236.4
$LiTaO_3$	2.628	1.831	−0.145	1.849	371.0	370.1

$(YXl)\,\theta=0°$	Lithium niobate			Lithium tantalate		
v_a (quasi-longitudinal) (m/s)	6806.8			5746.2		
v_b (quasi-shear) (m/s)	4446.0			3833.4		
v_c (pure shear) (m/s)	3940.6			3536.5		
Piezoelectric coupling (%)	TE	LE		TE	LE	
		E_1	E_3		E_1	E_3
k_a	30.5	0	4.6	23.7	0	1.5
k_b	55.4	0	0.8	35.5	0	0.4
k_c	0	45.0	0	0	31.1	0

FIG. 2. Growth habits of crystalline sapphire (symmetry class $\bar{3}m$), showing the natural *a*, *r*, and *n* facets, and basal plane *c*. *X* is the digonal axis, *Z* the trigonal axis, and *Y-Z* is the mirror plane.

V. MATERIALS IN CLASS $\bar{3}m$

Sapphire (α-Al_2O_3, corundum, alumina, ruby, emery) belongs to symmetry point group $\bar{3}m$; other members of this class include hematite (Fe_2O_3), and calcite ($CaCO_3$). Some salient physical characteristics are listed in Table X. Sapphire is a technologically popular and important crystal, particularly it is a primary substrate for growing the wide band gap semiconductor GaN. It has the same elastic stiffness matrix as lithium niobate (class $3m$) and quartz (class 32). Its Van Dyke matrix is that of Table VII with the piezoelectric $[e]$ entries set to zero. It possesses a trigonal rotatory inversion axis (Z) with mirror plane. Three equivalent secondary axes, lying 120° apart, are chosen parallel to the diad axes. *X* coincides with any one of the secondary axes, and *Y* is perpendicular to *Z* and *X* to form a right-hand coordinate system.[6,7] Growth habits of crystalline sapphire, showing the natural *a*, *r*, and *n* facets, and basal plane *c* are seen in Fig. 2. In the crystallographic literature, an *a* face is referred to as a "second-order prism." Similarly, *r* and *R* faces are referred to as "positive and negative rhombohedrals," respectively.

Three mirrors lie in the planes the normals to the *a* faces make with the *Z* axis, and divide adjacent *n* facets. A mirror plane is designated *m*, and called a "first-order prism." The normal to an *r* face lies in the (*Y-Z*) mirror plane. The *r* and *R* planes in sapphire are akin to the *BT* and *AT* (rotated-*Y*-cuts) of quartz;[19] wave propagation is perpendicular to the digonal axis, and at an angle to the trigonal axis.

TABLE X. Room temperature physical characteristics of sapphire.

Quantity	Unit	Value
Mass density	kg/m^3	3982
Molecular weight		101.96
Debye temperature	K	1033.9
Melting temperature	K	2325±5°
Specific heat capacity	J/(K mol)	79.45
Thermal conductivity	W/(m K)	K_1=25.2
		K_3=23.1
Lattice parameters	nm	a=0.475 70
		c=1.298 77
Morphological axial ratio	$\rho=c/2a$	1.365 11
Mohs hardness		9
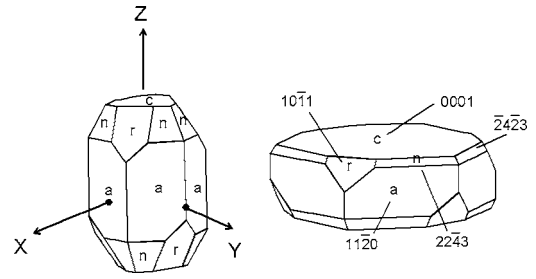
The Miller indices used here for sapphire are pertinent to the morphological axial ratio ($c/2a$). Some authors and vendors use the ratio (c/a), and this can be a source of confusion. For example, $(1,0,\bar{1},1)$ specifies an *r* plane (with an angle of 57.62° between its normal and the *c*-plane normal) if the ratio ($c/2a$) is used, but would specify an *R* plane (with an angle of 38.25° between its normal and the *c*-plane normal) if (c/a) were to be used. The presence of c_{14} in the Van Dyke matrix gives a polarity to the *X* axis that depends on the sign of c_{14}. We take c_{14} to be negative, which determines the sign of the angle θ for rotated-*Y*-cuts, $(YX1)\theta$. The *r* plane occurs for positive θ values, and the *R* plane for negative θ values with this convention. In the following labeling of the planes, which establishes a concordance between some of the planes in sapphire and the nomenclature used to describe quartz, the morphological axial ratio ($c/2a$) is used to identify the planes in accordance with Belyaev,[20]

c plane: $(0,0,0,1) \sim Z$ cut,

m planes: $(1,0,\bar{1},0);(0,1,\bar{1},0);(\bar{1},1,0,0);(\bar{1},0,1,0);$

$\times (0,\bar{1},1,0);(1,\bar{1},0,0) \sim Y$ cut,

r planes: $(1,0,\bar{1},1);(\bar{1},1,0,1);(0,\bar{1},1,1)$

$\sim BT$ cut(r closer to Z cut);rotated $-Y-$ cut,

R planes: $(\bar{1},0,1,2);(1,\bar{1},0,2);(0,1,\bar{1},2)$

$\sim AT$ cut(R closer to Y cut);rotated $-Y-$ cut,

a planes: $(2,\bar{1},\bar{1},0);(\bar{1},2,\bar{1},0);(\bar{1},\bar{1},2,0) \sim X$ cut,

n planes: $(2,2,\bar{4},3);(\bar{2},4,\bar{2},3);(\bar{4},2,2,3);(\bar{2},\bar{2},4,3);$

$\times (2,\bar{4},2,3);(4,\bar{2},\bar{2},3) \sim$ rotated $-X-$ cuts.

Much work has been reported on various features of sapphire and its characterization.[20,21] An extensive review of extant literature leads us to conclude that the best present-day values for the elastic stiffnesses of sapphire, and their temperature variation are those of Goto *et al.*;[22] these are given in Table XI(a) for room temperature. The thermoelastic values of White and Roberts[23] in Table XI(b) are similarly to

TABLE XI. (a) Sapphire elastic stiffnesses and temperature coefficients (RT) and (b) sapphire thermoelastic coefficients (RT).

(a)		c_{11} (GPa)	c_{33} (GPa)	c_{44} (GPa)	c_{66} (GPa)	c_{13} (GPa)	c_{14} (GPa)	(c_{12}) (GPa)	ρ (MPa/K)	Ref.
	c_{ij}	497.3	500.9	146.8	167.25	116.0	−21.90	162.8	3982	22
	$T_c{}^a$ (ppm/K)	−70.2	−68.9	−167.6	−108.8	−71.6	+267.1	+9.2		
	$\partial c/\partial T$ (MPa/K)	−34.9	−34.5	−24.6	−18.2	−8.3	−5.85	+1.5		
(b)		α_{11} (ppm/K)	α_{33} (ppm/K)	$(2\alpha_{11}+\alpha_{33})$ (ppm/K)	T_ρ (ppm/K)	Ref.				
		5.15	5.90	16.20	−16.20	23				

$^a T_c = (\partial c/\partial T)/c$.

be preferred. In general, the stiffnesses are about five times those of quartz,[19] but the velocity anisotropy is much less because $c_{11}=c_{22}\approx c_{33}$ and $c_{44}=c_{55}\approx c_{66}$.

Application of the CB formalism of Sec. II discloses the following behavior for different orientations:

X cut ($\alpha_1=-1, \alpha_2=0, \alpha_3=0$): Propagation along X; $\Gamma_{11}=c_{11}$ uncoupled longitudinal wave with motion along X axis; two coupled shear waves.

Y cut ($\alpha_1=0, \alpha_2=1, \alpha_3=0$): Propagation along Y; $\Gamma_{11}=c_{66}$ uncoupled shear wave, with motion along X axis; coupled quasi-shear and quasi-longitudinal waves.

Z cut ($\alpha_1=0, \alpha_2=0, \alpha_3=1$): Propagation along Z; three uncoupled waves, longitudinal motion along Z axis; two degenerate shear waves with motions perpendicular to Z.

Rotated-Y-cut ($\alpha_1=0, \alpha_2=\cos\theta; \alpha_3=\sin\theta$): Propagation perpendicular to X; Γ_{11} uncoupled shear wave, with motion along X axis; coupled quasi-shear and quasi-longitudinal waves.

Rotated X cut ($\alpha_1=-\cos\theta, \alpha_2=0, \alpha_3=\sin\theta$), or any arbitrarily doubly rotated cut ($\alpha_1=-\sin\varphi\cos\theta; \alpha_2=\cos\varphi\cos\theta; \alpha_3=\sin\theta$): Propagation perpendicular to Y; three coupled quasi-shear and quasi-longitudinal waves.

Table XII contains the acoustic velocities and direction cosines for plane wave propagation in various directions in sapphire. The angles θ for the r, R planes are the complements of the 57.62° and 38.25° values mentioned earlier. The columns headed (−r) and (+R) contain calculations for θ values that are the negatives of those for these planes; this is done to show how little acoustic anisotropy is present for the rotated-Y-cut values ($\varphi=0$) given; this lack is maintained for all general doubly rotated cuts ($\varphi\neq 0$) in sapphire. The direction cosines referred to the crystallographic coordinates (γ) are related to those referred to the plate coordinates (β) by the transformation $\beta_{ik}=\alpha_{ij}\gamma_{jk}$.

VI. COMPOSITE RESONATORS

A. Single-mode HBARs

The foregoing has considered the acoustic behavior of piezofilm transducers and substrates separately. In this sec-

TABLE XII. Acoustic velocities and direction cosines for various planes in sapphire; "p" indicates a pure (uncoupled) mode.

Plane $\theta° \rightarrow$	c ±90	−R −51.75	−45	(−r) −32.38	m 0	+r +32.38	+45	(+R) +51.75	c ±90
va km/s	11.216p	10.547	10.428	10.443	11.191	11.105	10.940	10.905	11.216p
vb	6.072p	6.761	6.925	6.926	6.481p	6.608	6.929	6.981	6.072p
vc	6.072p	6.647p	6.703p	6.746p	6.043	5.963p	5.825p	5.787p	6.072p
				Direction cosines—crystallographic coordinates					
γ_{qL1}	0	0	0	0	0	0	0	0	0
γ_{qL2}	0	−0.5460	−0.6716	−0.8743	+0.9981	+0.8747	+0.7316	+0.6227	0
γ_{qL3}	1	+0.8378	+0.7409	+0.4854	+0.0621	+0.4846	+0.6818	+0.7825	1
γ_{qS1}	0	0	0	0	0	0	0	0	0
γ_{qS2}	1	+0.8378	+0.7409	+0.4854	−0.0621	−0.4846	−0.6818	−0.7825	1
γ_{qS3}	−0	+0.5460	+0.6716	+0.8743	+0.9981	+0.8747	+0.7316	+0.6227	−0
γ_S	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]
				Direction cosines—plate coordinates					
β_{qL1}	0	0	0	0	0	0	0	0	0
β_{qL2}	1	−0.9960	−0.9988	−0.9983	+0.9981	+0.9982	+0.9994	+0.999+	1
β_{qL3}	0	+0.0898	+0.0490	−0.0583	+0.0621	−0.0592	−0.0352	−0.0046	0
β_{qS1}	0	0	0	0	0	0	0	0	0
β_{qS2}	0	+0.0898	+0.0490	−0.0583	−0.0621	+0.0592	+0.0352	+0.0046	0
β_{qS3}	−1	+0.9960	+0.9988	+0.9983	+0.9981	+0.9982	+0.9994	+0.999+	−1
β_S	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]	[1,0,0]

tion the mode structure of composites formed by joining the layers is briefly considered. Virtually all of the HBAR structures constructed to date utilize a single acoustic mode. In the standard ZnO/Al_2O_3 configuration, the zinc oxide is deposited on a sapphire face cut perpendicular to its trigonal axis (c plane); the ZnO is found to grow with its polar axis normal to this face. The trigonal axis of the sapphire and the hexagonal (polar) axis of the zinc oxide are therefore parallel. Piezoelectric excitation of the ZnO film by an electric field in its thickness direction (TE) drives the pure longitudinal mode in the ZnO, as discussed in Sec. III. This pure mode couples energy into the sapphire by virtue of the mechanical attachment of the film to the substrate. Within the sapphire substrate, it is found that a pure longitudinal plane wave is allowed. It propagates along the Z axis of the sapphire. Thus the single driven mode in the zinc oxide gives rise to only a single wave type in the sapphire. The two other pure shear waves in the ZnO and the two other pure shear waves in the sapphire are unexcited. If the sapphire substrate bearing the ZnO film is terminated at its other end by a plane parallel to the film, then the longitudinal wave will be reflected thereat, and the device is regarded as a resonator supporting motions of a single mode type.

Alternatively, using the same ZnO-sapphire configuration, piezoelectric excitation of the piezofilm by an X-directed electric field in the plane of the film-substrate interface (lateral excitation, LE), drives a pure shear mode in the ZnO, which couples to pure shear motion in the sapphire. The two other pure modes in the ZnO (one longitudinal, and one shear) and the two pure waves in the sapphire (again, one longitudinal, and one shear) are unexcited.

In either of these TE or LE cases, the undriven wave types are irrelevant to the electrical characteristics of the device measured between the two electrodes. The resulting situation is that of a single piezoelectrically driven mode in the film mechanically driving a single piezoelectrically inert mode in the substrate by the surface tractions at the film-substrate interface. These single-mode to single-mode resonator cases are of some complexity, but have been treated previously in the literature.[1]

However, for other orientations of the piezoelectric film and substrate, a plurality of modes will be excited in general, even in the one-dimensional simple thickness mode approximation. While the ensuing mode structure is rendered more complicated with these orientations, the opportunities for utilization of the resulting multi-mode spectrum is also enhanced. Not only may new types of multi-mode HBARs and TFRs be constructed, but one may also use the increased gamut of possibilities in reverse, to infer the elastic properties of newly developed but poorly characterized substrate materials such as langasite and its isomorphs.[24] Family members include: langasite ($La_3Ga_5SiO_{14}$); langanite ($La_3Ga_{11/2}Nb_{1/2}O_{14}$); and langatate ($La_3Ga_{11/2}Ta_{1/2}O_{14}$). These class 32 crystals offer the promise of exceedingly low acoustic loss, combined with excellent temperature behavior.

B. Multi-mode HBARs

The CB formulation of Sec. II discloses the possibilities arising in the more intricate and interesting case where a single mode is piezoelectrically driven in a film of ZnO deposited on a planar substrate of sapphire as before, but where four modes (two in each medium) are coupled by the mechanical boundary conditions at the interface. This situation arises when the ZnO/sapphire components assume different crystallographic orientations, and leads to enlarged prospects for producing novel and practical devices. In one such demonstrated device,[10] the structure consists of $Al/Mg_xZn_{1-x}O/n+\text{-}ZnO$ on an r-plane sapphire plate, where Al and the n-type doped ZnO thin film serve as top and bottom electrodes, respectively.

For $Mg_xZn_{1-x}O$ films grown on $r\text{-}Al_2O_3$ substrates, the orientational relationship is $(11\bar{2}0)Mg_xZn_{1-x}O\|(01\bar{1}2)Al_2O_3$ and $[0001]Mg_xZn_{1-x}O\|[01\bar{1}1]Al_2O_3$. Thus the $Mg_xZn_{1-x}O$ films have their c axis in the growth plane, and perpendicular to the $\bar{3}mX$ axis. As a consequence, piezoelectric excitation of the $Mg_xZn_{1-x}O$ film by an electric field in its thickness direction (TE), drives the pure shear mode in the film, which mechanically forces the two elastically coupled modes (quasi-longitudinal and quasi-shear) in the sapphire. This, in turn, leads to additional mechanical coupling with the pure longitudinal mode in the $Mg_xZn_{1-x}O$ via the mechanical boundary conditions at the interface.

Application of an electric field in the plane of the $Mg_xZn_{1-x}O$ film (LE) drives piezoelectrically the pure longitudinal mode in the film. The two elastically coupled modes (quasi-longitudinal and quasi-shear) in the sapphire substrate are thereby forced. This, in turn, leads to further coupling with one of the pure shear modes in the $Mg_xZn_{1-x}O$ via the mechanical boundary conditions at the interface.

Table XIII gives the excitation possibilities arising when ZnO or MgZnO is deposited on various facets of sapphire (or any class $\bar{3}m$, $3m$, or 32 material). It is formed using the results of the CB analysis of Sec. II. Although multi-mode HBARs may be fashioned from any oblique orientations of sapphire, the use of r- or R-plane (rotated-Y-cut) material appears at present to be most desirable because of commercial availability. Table XIV gives the excitation possibilities arising when a Y-cut crystal of class $3m$ is used as the piezoelectric excitation layer. Examples include a Y-cut $LiNbO_3$ crystal bonded to sapphire for resonant ultrasound spectroscopy, or a $LiNbO_3$ or $LiTaO_3$ thin film deposited on a sapphire substrate.

TABLE XIII. Acoustic modes excited in thin-film ZnO/Al_2O_3 composite.[a]

| Substrate Al_2O_3 Orientation | Thin film ZnO | Excitation method | | | |
| | | Thickness (TE) | | Lateral (LE) | |
		ZnO	Al_2O_3	MgZnO	Al_2O_3
c plane	c plane	p L	p L	p S	p S
r,R planes	a plane	p S	p S, q-L, q-S	p L	q-L, q-S
a plane	c plane	p L	p L	p S	$2S$[b]
n plane	a plane	p S	q-L, $2q$-S	p L	q-L, $2q$-S

[a]p=pure mode; q-S, q-L=quasi-shear, quasi-longitudinal modes.
[b]Two coupled shear modes.

1422 J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Wittstruck *et al.*: Resonator transducer and substrate properties

TABLE XIV. Motions driven in Al_2O_3 substrate by Y-cut crystal of class $3m$ such as lithium niobate.

Substrate	Transducer	Thickness excitation (TE)			
Al_2O_3	$LiNbO_3$	$\psi=0$[a]		$E_3\,\psi\neq0$[a]	
Orientation		$LiNbO_3$	Al_2O_3	$LiNbO_3$	Al_2O_3
c plane	Y cut	q-L, q-S	p S, p L	q-L, q-S	p S, p L
r,R planes	Y cut	q-L, q-S	q-L, q-S	q-L, q-S	p S, q-L, q-S
a plane	Y cut	q-L, q-S	p L, 2 S[b]	q-L, q-S	p L, 2 S[b]
n plane	Y cut	q-L, q-S	q-L, $2q$-S	q-L, q-S	q-L, $2q$-S

[a]Angle between $LiNbO_3$ X_1 axis and the X_1' axis of Al_2O_3 substrate.
[b]Two coupled shear modes.

VII. CONCLUSIONS

We have presented the Christoffel-Bechmann method for calculating acoustic wave speeds and coupling coefficients of arbitrarily oriented piezoelectric plates and films, for electric fields impressed either along the thickness or laterally. The formalism has been applied to characterize the generic behavior of plate composites consisting of a crystal substrate surmounted by a piezoelectric film that drives the structure in resonant modes as part of a high stability oscillator or low loss filter circuit. The calculations apply generally to crystals of any class. The piezoelectric portion is made specific for films of crystal class $6mm$ with arbitrary orientation on the substrate, and the substrate calculations are specified for class $\bar{3}m$ materials because these combinations appear to encompass the most desirable current material classes, although the $\bar{3}m$ results apply also to any substrate crystal in the $3m$ or 32 crystal classes. Zinc oxide and sapphire have been used in an example of the acoustic resonator structure.

[1]R. Weigel, D. P. Morgan, J. M. Owens, A. Ballato, K. M. Lakin, K. Hashimoto, and C. C. W. Ruppel, "Microwave acoustic materials, devices and applications," IEEE Trans. Microwave Theory Tech. **50**, 738–749 (2002).
[2]R. Naik, J. J. Lutsky, R. Reif, and C. Sodini, "Electromechanical coupling constant extraction of thin-film piezoelectric materials using a bulk acoustic wave resonator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 257–263 (1998).
[3]T. R. Meeker, "Piezoelectric resonators and applications," in *Encyclopedia of Applied Physics*, edited by G. Trigg (VCH, New York, 1996), Vol. **14**, pp. 147–169.
[4]W. G. Cady, "The piezoelectric resonator and the effect of electrode spacing upon frequency," Physics (N.Y.) **7**, 237–259 (1936).
[5]R. Bechmann, "Über Dickenschwingungen piezoelektrischer Kristallplatten," Arch. Elektr. Uebertrag. **6**, 361–368 (1952); Nachtrag **7**, 354–356 (1953). (present name: Archiv für Elektronik und Übertragungstechnik).
[6]"IRE Standards on Piezoelectric Crystals, 1949, " Proc. IRE **37**, 1378–1395 (1949). (IEEE Standard No. 176).
[7]"IEEE Standard on Piezoelectricity," ANSI/IEEE Standard 176-1987 (IEEE, New York, 1987).
[8]A. Ballato, "Piezoelectricity: Old effect, new thrusts," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 916–926 (1995).
[9]"IRE Standards on Piezoelectric Crystals: Determination of the elastic, piezoelectric, and dielectric constants—the electromechanical coupling factor, 1958," Proc. IRE **46**, 764–778 (1958). (IEEE Standard No. 178).
[10]R. H. Wittstruck, X. Tong, N. W. Emanetoglu, P. Wu, Y. Chen, J. Zhu, S. Muthukumar, Y. Li, and A. Ballato, "Characteristics of $Mg_xZn_{1-x}O$ thin film bulk acoustic wave devices," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 1272–1278 (2003).
[11]G. Carlotti, G. Socino, A. Petri, and E. Verona, "Elastic constants of sputtered ZnO films," IEEE International Ultrasonics Symposium Proceedings, 1987, pp. 295–299.
[12]E. F. Tokarev, I. B. Kobyakov, I. P. Kuzmina, A. N. Lobachev, and G. S. Pado, "Elastic, dielectric, and piezoelectric properties of zincite in the $4.2–800°$ K temperature range," Sov. Phys. Solid State **17**, 629–632 (1975).
[13]S. L. Pinkett, W. D. Hunt, B. P. Barber, and P. L. Gammel, "Determination of ZnO temperature coefficients using thin film bulk acoustic wave resonators," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1491–1496 (2002).
[14]S.-H. Lim, D. Shindo, H.-B. Kang, and K. Kakamura, "Study of defects and interfaces in epitaxial ZnO films on $(1,1,\bar{2},0)$ Al_2O_3 grown by electron cyclotron resonance-assisted molecular beam epitaxy," J. Cryst. Growth **225**, 202–207 (2001).
[15]C. R. Gorla, N. W. Emanetoglu, S. Liang, W. E. Mayo, Y. Lu, M. Wraback, and P. H. Shen, "Structural, optical, and surface acoustic wave properties of epitaxial ZnO films grown on $(01\bar{1}2)$ sapphire by metalorganic chemical vapor deposition," J. Appl. Phys. **85**, 2595–2602 (1999).
[16]S. Muthukumar, J. Zhong, Y. Chen, Y. Lu, and T. Siegrist, "Growth and structural analysis of metalorganic chemical vapor deposited $(11\bar{2}0)$ $Mg_xZn_{1-x}O$ $(0<x<0.33)$ films on $(01\bar{1}2)$ R-plane Al_2O_3 substrates," Appl. Phys. Lett. **82**, 742–744 (2003).
[17]Semiconductor epitaxial material and microwave devices employing GaN are available, e.g., from Osemi, 300 First Street NE, Rochester, MN 55908.
[18]J. Kushibiki, I. Takanaga, M. Arakawa, and T. Sannomiya, "Accurate measurements of the acoustical physical constants of $LiNbO_3$ and $LiTaO_3$ single crystals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1315–1323 (1999).
[19]A. Ballato, "Elastic properties of crystalline quartz," in *Handbook of Elastic Properties of Solids, Liquids, and Gases*, Vol. **2**, edited by M. Levy, H. E. Bass, and R. Stern (Academic, San Diego, 2001), pp. 257–279.
[20]*Ruby and Sapphire*, edited by L. M. Belyaev (Amerind, New Delhi, 1980). Translation of *Rubin i Sapfir* (Nauka, Moscow, 1974).
[21]D. G. Isaak, "Elastic properties of minerals and planetary objects," in *Handbook of Elastic Properties of Solids, Liquids, and Gases*, edited by M. Levy, H. E. Bass, and R. Stern, Elastic Properties of Solids: Biological and Organic Materials, Earth and Marine Sciences, Vol. **3** (Academic, San Diego, 2001), pp. 325–376.
[22]T. Goto, O. L. Anderson, I. Ohno, and S. Yamamoto, "Elastic constants of corundum up to 1825 K," J. Geophys. Res. **94B**, 7588–7602 (1989).
[23]G. K. White and R. B. Roberts, "Thermal expansion of reference materials: tungsten and α-Al_2O_3," High Temp. - High Press. **15**, 321–328 (1983).
[24]R. C. Smythe, "Material and resonator properties of langasite and langatate: A progress report," IEEE Intl. Freq. Control Symp. Proc., 1998, pp. 761–765.

A theoretical and numerical analysis of vibration-controlled modules for use in active segmented partitions

Timothy W. Leishman
Acoustics Research Group, Department of Physics and Astronomy, Brigham Young University, Eyring Science Center, Provo, Utah 84602

Jiri Tichy
Graduate Program in Acoustics, The Pennsylvania State University, Applied Science Building, University Park, Pennsylvania 16802

Many applications of active sound transmission control (ASTC) require lightweight partitions, high transmission loss over a broad frequency range, simple control strategies, and consistent performance for various source and receiving space conditions. In recent years, researchers have begun to investigate active segmented partitions (ASPs) because of their potential to meet such requirements. This paper provides a theoretical and numerical analysis of four ASP module configurations that are candidates for these applications. Analogous circuit methods are used to provide normal-incidence transmission loss and reflection coefficient estimates for their passive and active states. The active control objective for each configuration is to induce global vibration control of various transmitting surfaces through direct vibration control of a *principal* transmitting surface. Two characteristic single-composite-leaf (SCL) configurations are unable to use the strategy effectively. However, design adjustments are investigated to improve their performances. Two double-composite-leaf (DCL) configurations use the strategy much more effectively to produce efficient global control of transmitting surface vibrations and achieve high transmission loss over a broad frequency range. This is achieved through a minimum volume velocity condition on the *source* side of each module. One DCL configuration enhances module isolation in full ASP arrays while satisfying other design and performance criteria. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1992767]

LIST OF SYMBOLS

B = Magnetic flux density in magnet air gap of moving-coil actuator

c = Speed of sound in fluid medium

C_{Mn} = Effective mechanical compliance of nth module element

C_{Mmn} = Effective mechanical compliance coupling mth and nth module elements ($m \neq n$)

$\hat{e}_g$ = Complex open-circuit voltage amplitude of electrical control source (signal generator)

f = Frequency

G_A = Acoustic ground (ambient reference pressure)

G_M = Mechanical ground (zero reference velocity)

I = Roman numeral representing source side of module

II = Roman numeral representing transmitting side of module

j = $\sqrt{-1}$

k = Acoustic wave number, $= \omega/c$

$\tilde{k}$ = Complex acoustic wave number accounting for propagation losses, $= k - j\alpha$

L = Effective cavity length in double-composite-leaf modules

l = Effective length of voice coil conductor in magnet air gap of moving-coil actuator

L_E = Electrical inductance of voice coil in moving-coil actuator

m = Integer index representing given module element, = 1, 2, 3, or 4

M_{Mn} = Effective mechanical mass of nth module element

n = Integer index representing given module element, = 1, 2, 3, or 4

$\hat{p}_i$ = Complex amplitude of constant normally incident acoustic pressure[1]

$\hat{p}_r$ = Complex amplitude of normally reflected acoustic pressure, $= \hat{p}_i - (\rho_0 c/S)\hat{U}_I$

$\hat{p}_t$ = Complex amplitude of normally transmitted acoustic pressure, $= (\rho_0 c/S)\hat{U}_{II}$

R = Complex pressure-amplitude reflection coefficient, $= 1 - (\rho_0 c/S)(\hat{U}_I/\hat{p}_i)$

R_E = Electrical resistance of voice coil in moving-coil actuator

R_g = Output resistance of electrical control source (signal generator)

R_{Mn} = Effective mechanical resistance of nth module element

R_{Mmn} = Effective mechanical resistance coupling mth and nth module elements ($m \neq n$)

S = Total cross-sectional area of ASP modules and hypothetical source and receiving space plane-wave tubes, $= S_1 + S_2 = S_3 + S_4$

S_n = Cross-sectional area of nth module element

TL = Normal-incidence sound transmission loss, $= 10 \log(1/\tau)$

$\hat{u}_n$ = Complex normal velocity amplitude of nth module element

$\hat{U}_\mathrm{I}$ = Volume velocity on side I (source side) of module

$\hat{U}_\mathrm{II}$ = Volume velocity on side II (transmitting side) of module

Z_{A1} = Acoustic impedance substitution for waveguide two-port network, $= j(\rho_0 \omega / \tilde{k} S) \tan(\tilde{k} L/2) \approx j(\rho_0 c / S) \tan(k L/2)$

Z_{A2} = Acoustic impedance substitution for two-port waveguide network, $= -j(\rho_0 \omega / \tilde{k} S) \csc(\tilde{k} L) \approx -j(\rho_0 c / S) \csc(k L)$

Z_{A3} = Acoustic impedance substitution, $= Z_{A1} + Z_{A2} + (\rho_0 c / S)$

Z_{A4} = Acoustic impedance substitution, $= Z_{A3} - [Z_{A2}^2 (Z_{M3} S_4^2 + Z_{M4} S_3^2 + Z_{M34} S^2)] / [(Z_{M3} S_4^2 + Z_{M4} S_3^2 + Z_{M34} S^2) Z_{A3} + Z_{M3} Z_{M4} + (Z_{M3} + Z_{M4}) Z_{M34}]$

Z_E = Isolated electric impedance of moving-coil actuator and electrical control source output, $= R_g + R_E + j\omega L_E$

Z_{Mn} = Effective mechanical impedance of nth module element, $= \begin{cases} j\omega M_{Mn}; \ n=1, \text{for module configurations 1 and 4} \\ R_{Mn} + j[\omega M_{Mn} - (1/\omega C_{Mn})]; \quad \text{otherwise} \end{cases}$

Z_{Mmn} = Effective mechanical impedance coupling mth and nth module elements $(m \neq n)$, $= R_{Mmn} + (1/j\omega C_{Mmn})$

α = Total thermoviscous absorption coefficient for acoustic losses in module cavities[2]

δ = In-plane width of thin transmitting surface

ρ_0 = Ambient density of fluid medium

σ_{Mn} = Effective mass surface density of nth element, $= M_{Mn}/S_n$

τ = Sound power transmission coefficient

ω = Angular frequency, $= 2\pi f$

I. INTRODUCTION

An active segmented partition (ASP) is a contiguous array of interconnected modules or elements that are actively controlled to reduce sound transmission between a source space and a receiving space. While limited investigations of ASPs have been conducted over the past few years,[3–16] in-depth comparative analyses of individual ASP module configurations are still needed to better understand their behaviors and to establish those that produce the highest transmission loss using simple control strategies.

A discrete ASP module is a complete subsection of an array, consisting of an integrated arrangement of active and passive components that are often acoustically and structurally small. If a module is well designed and controlled, it interacts with adjacent modules and supporting structures to yield effective control of sound transmission through the entire partition. Moreover, because an ASP typically consists of several identical modules, individual module design governs overall partition performance, as does the method of control, the method of structural support, and the method of module interconnection. Inadequate attention to any of these details

can lead to serious performance limitations (at least over substantial bandwidths) or relegate the approach to use with impractical control schemes.

Much of the past research in active sound transmission control (ASTC) has focused on difficulties of simultaneously controlling distributed structural and fluid media, and has essentially dismissed active vibration control of partitions as an effective means of controlling sound transmission.[17–22] The research has instead emphasized restructuring of distributed partition modes (rather than suppression of modes) to produce normal surface velocity distributions that transmit inefficiently into receiving spaces.[23] Nevertheless, researchers have found that effective reduction of sound transmission is generally accomplished by both partition modal amplitude reduction and modal restructuring,[24] with the weight of each mechanism depending upon specific geometric and structural-acoustic parameters.[5,25–27] As a result, the notion that active suppression of partition vibration yields ineffective control of sound transmission should not necessarily be applied as a generalization to all types of partitions.

Many complexities of distributed partition ASTC that have discounted vibration control methods result from a common condition in which secondary control force or moment distributions fail to match corresponding distributions produced by primary incident pressure disturbances. However, local physical conditions differ significantly for partitions segmented into small, discretely controlled components. In the case of ASPs, primary and secondary force distributions may be constrained to match more closely on key module surfaces, permitting more efficient simultaneous control of partition vibration and sound transmission. Moreover, efficient global control of transmitting surface vibrations produces sound transmission control applicable to many types of source and receiving spaces—to both their near fields and far fields.

Past ASTC research has also emphasized the behavior and positional optimization of error sensors and actuators in conjunction with active partitions and their surroundings. However, transducer positional optimization schemes are often based on unreliable assumptions of system time invariance. While adaptive controllers are able to modify and optimize control filters for time-varying systems, *a priori* optimized transducer locations must necessarily remain fixed and nonadaptive. What is needed, therefore, is a physical ASTC strategy that performs well for a variety of source and receiving spaces and remains essentially unaffected by their time-varying physical changes.

In recent years, the authors have identified several research considerations to address such needs and to otherwise improve the practicality and performance of ASTC systems. Other efforts have provided tools to conceptualize, model, and evaluate ASP modules, and have led to the proposal of seven criteria for their design and performance.[7,11–13,16] While some research efforts have shown that ASPs may be used to control transmitted volume velocity, sound power, total potential energy, or other receiving space field quantities (typically through elaborate multichannel sensing and

control), the seven criteria suggest that a much simpler scheme based on localized module vibration control may be advantageous.

This paper provides a comparative theoretical and numerical investigation of four ASP modules that were identified as candidates to satisfy the criteria. It explores their salient properties and a practical vibration-based scheme to control them. It compares normal-incidence transmission loss and reflection coefficient estimates for the modules in their passive and active states. The transmission loss estimates are well suited for comparison with classical normal-incidence transmission loss formulations. The analysis reveals basic problems and performance limitations for some configurations, and provides recommendations that may be used to enhance their effectiveness. It also identifies mechanisms of higher-performance vibration-controlled ASPs to enhance the viability of ASPs as ASTC solutions.

The underlying active control objective for each module is to induce global vibration control of its various transmitting surfaces by directly minimizing normal vibration of its *principal* transmitting surface. The analysis shows why single-composite-leaf (SCL) ASP modules are typically unable to use this simple strategy effectively. It then introduces two double-composite-leaf (DCL) modules that use the strategy much more effectively to produce efficient global vibration control of transmitting surfaces and achieve high transmission loss over a broad frequency range. One module is especially noteworthy in that it is designed to enhance implementation of full ASP arrays while satisfying the seven design and performance criteria mentioned above.

The investigation explores physical configurations of individual ASP modules rather than complete ASP arrays. However, it does so while acknowledging basic array requirements and characteristics. Results are derived assuming the use of moving-coil actuators, but they are representative of results based on other transducer formats. One-dimensional acoustic fields, lumped mechanical elements, analogous electro-mechano-acoustic circuits,[28,29] and time-harmonic excitation and control are employed to efficiently model and evaluate the configurations. Relevant assumptions and simplifications are made to limit otherwise unwieldy circuit diagrams to planar (though still somewhat complicated) forms. These steps are carried out carefully to avoid detracting from key module characteristics.

The following sections present theoretical developments for the four module configurations and numerical examples to illustrate their characteristic behaviors. They also discuss additional ASP module limitations and capabilities from a general standpoint.

II. SELECTED ASP MODULE CONFIGURATIONS

A. Configuration 1: Single composite leaf (SCL) with a surrounding finite-impedance interstice

The first module configuration is shown in Fig. 1. It includes an actuator with its diaphragm and a surrounding interstice to which the actuator frame is connected. Little consideration has previously been given to the effects that actuator control has on supporting ASP structures such as
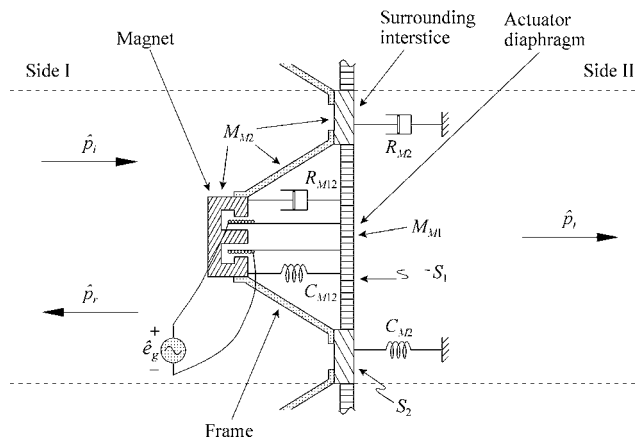
FIG. 1. Isolated view of ASP module configuration 1. An actuator diaphragm vibrates freely but snugly within the frictionless interstice opening. Both the actuator diaphragm and the interstice are assumed to vibrate as pistons in normal translational rigid-body modes. Plane waves are normally incident and reflected in the source space, and normally transmitted in the receiving space. The model isolates effects of acoustic forces and reactive motor forces on the interstice. (After Ref. 7.)

this interstice. A common tacit assumption is that supporting structures have infinite impedance. However, this notion is invalid for actuators supported by many partition elements. When exposed finite-impedance elements form transmitting interstitial components, their vibrations can pose significant ASTC problems.

The supporting interstice is assumed to form a single lumped-element system around the periphery of the actuator diaphragm with effective mechanical mass, compliance, and resistance. The diaphragm is assumed to vibrate freely but snugly within its frictionless opening such that no sound can transmit through an intervening crack. Although lumped mechanical resistance and compliance elements are shown on opposite sides of the diaphragm and interstice, they are assumed to apply uniformly over the pertinent cross sections. The interstice and diaphragm are assumed to vibrate only in normal translational rigid-body modes, so the entire transmitting surface area of the module is composed of piston-like elements. The actuator magnet and frame are considered to be acoustically unobtrusive lumped masses that firmly attach to the interstice and vibrate in unison with it.

As suggested by the dashed horizontal lines in the drawing, the semi-infinite source space (side I) and receiving space (side II) are assumed to behave as one-dimensional low-frequency fields, such as those constrained by rigid-walled plane-wave tubes. A constant incident pressure wave impinges normally upon the module from the source space while another wave reflects normally back into the source space. A transmitted wave propagates normally into the receiving space, which has the same characteristic impedance as the source space.

1. Equivalent circuit modeling and analytical results

An equivalent circuit representing the complete system under time-harmonic excitation and control is shown in Fig. 2 (see the List of Symbols for clarification of labels). The circuit includes four ideal area gyrators that couple the me-
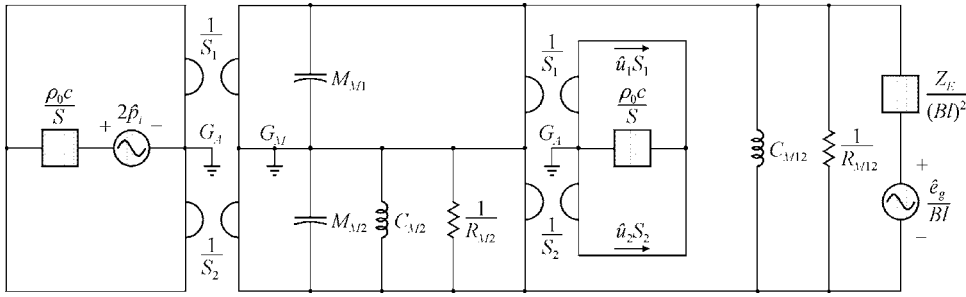
FIG. 2. Multiple-domain equivalent circuit representing ASP module configuration 1. (After Ref. 7.)

chanical mobility portion of the circuit to the acoustic impedance portions, the latter representing the source and receiving spaces. The control voltage input and combined electric impedance of the signal generator and actuator have been carried through an ideal transformer from the electric impedance domain to the mechanical mobility portion of the circuit.

Normal velocity amplitudes of the actuator diaphragm ($\hat{u}_1$) and the surrounding interstice ($\hat{u}_2$) may be derived directly from the circuit by (1) writing the defining equations for the ideal gyrators and ensuring they are satisfied with appropriate flow and potential senses; (2) writing the equations of motion in all physical domains using Kirchhoff's flow law (i.e., writing the nodal equations for all nodes except the ground reference nodes); and (3) solving the simultaneous algebraic equations for the velocities. Four nodal equations for the model reduce to two coupled equations of motion for the diaphragm and interstice, yielding the two complex amplitudes. They are functions of both incident pressure and electrical input voltage

$$\hat{u}_1 = \frac{2\hat{p}_i\left\{ Z_{M2}S_1 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]S \right\} + \hat{e}_g\frac{Bl}{Z_E}(2\rho_0cS_2 + Z_{M2})}{\left[2\frac{\rho_0c}{S}S_1^2 + Z_{M1} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right]\left[2\frac{\rho_0c}{S}S_2^2 + Z_{M2} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right] - \left[2\frac{\rho_0c}{S}S_1S_2 - Z_{M12} - \frac{(Bl)^2}{Z_E} \right]^2},$$ (1)

$$\hat{u}_2 = \frac{2\hat{p}_i\left\{ Z_{M1}S_2 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]S \right\} - \hat{e}_g\frac{Bl}{Z_E}(2\rho_0cS_1 + Z_{M1})}{\left[2\frac{\rho_0c}{S}S_1^2 + Z_{M1} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right]\left[2\frac{\rho_0c}{S}S_2^2 + Z_{M2} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right] - \left[2\frac{\rho_0c}{S}S_1S_2 - Z_{M12} - \frac{(Bl)^2}{Z_E} \right]^2}.$$ (2)

The total volume velocity produced by the vibrating surfaces is the same on both sides of the module:

$$\hat{U}_I = \hat{U}_{II} = \hat{u}_1S_1 + \hat{u}_2S_2 = \frac{2\hat{p}_i\left\{ Z_{M1}S_2^2 + Z_{M2}S_1^2 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]S^2 \right\} + \hat{e}_g\frac{Bl}{Z_E}(Z_{M2}S_1 - Z_{M1}S_2)}{\left[2\frac{\rho_0c}{S}S_1^2 + Z_{M1} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right]\left[2\frac{\rho_0c}{S}S_2^2 + Z_{M2} + Z_{M12} + \frac{(Bl)^2}{Z_E} \right] - \left[2\frac{\rho_0c}{S}S_1S_2 - Z_{M12} - \frac{(Bl)^2}{Z_E} \right]^2}.$$ (3)

This result is used to determine the transmitted pressure, sound power transmission coefficient, transmission loss, and source-side pressure-amplitude reflection coefficient under both passive and active conditions.

Under open-circuit passive conditions, $Z_E \rightarrow \infty$ and the inverse transmission coefficient becomes[30]

$$\frac{1}{\tau} = \left| \frac{\left(2\frac{\rho_0c}{S}S_1^2 + Z_{M1} + Z_{M12} \right)\left(2\frac{\rho_0c}{S}S_2^2 + Z_{M2} + Z_{M12} \right) - \left(2\frac{\rho_0c}{S}S_1S_2 - Z_{M12} \right)^2}{2\frac{\rho_0c}{S}(Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2)} \right|^2.$$ (4)

The reflection coefficient becomes

$$R = 1 - 2\frac{\rho_0 c}{S}\left[\frac{Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2}{\left(2\frac{\rho_0 c}{S}S_1^2 + Z_{M1} + Z_{M12}\right)\left(2\frac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}\right) - \left(2\frac{\rho_0 c}{S}S_1 S_2 - Z_{M12}\right)^2}\right]. \tag{5}$$

As stated in Sec. I, the active control strategy for the configuration is to minimize the normal velocity of its principal transmitting surface: the actuator diaphragm. Equation (1) demonstrates that it may be driven to zero (assuming sufficient actuator power handling and linear response) using the following input control voltage:

$$\hat{e}_g = -2\hat{p}_i \frac{Z_E}{Bl}\left\{\frac{Z_{M2}S_1 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E}\right]S}{2\rho_0 c S_2 + Z_{M2}}\right\}. \tag{6}$$

However, forces due to the impinging incident wave, coupling, and reaction of the actuator motor against the finite impedance interstice will cause it to vibrate with velocity amplitude

$$\hat{u}_2 = \frac{2\hat{p}_i(S_1 + S_2)}{2\rho_0 c S_2 + Z_{M2}}. \tag{7}$$

The force due to incident pressure on the actuator diaphragm is accordingly transferred by the actuator motor to the surrounding interstice. The residual volume velocity $\hat{U}_{\mathrm{II}} = \hat{u}_2 S_2$ leads to the controlled inverse transmission coefficient

$$\frac{1}{\tau} = \left|1 + \frac{Z_{M2}}{2\rho_0 c S_2}\right|^2. \tag{8}$$

Significantly, this expression is the same as that resulting from a *passive* single-leaf partition with mechanical impedance Z_{M2} and transmitting surface area S_2, instead of total surface area $S = S_1 + S_2$.

When the interstice is mass controlled, the reciprocal transmission coefficient becomes

$$\frac{1}{\tau} \approx 1 + \left(\frac{\omega \sigma_{M2}}{2\rho_0 c}\right)^2, \tag{9}$$

which corresponds to the normal-incidence mass law for a homogeneous single-leaf partition of surface density σ_{M2}. Notably, the interstice mass M_{M2} includes the mass of the actuator magnet and frame. Hence, this result shows that when the interstice is mass controlled, the active control scheme provides no added benefit to that which could be achieved by a *passive* module constructed solely of the interstitial material with a mass equivalent to that of the actuator magnet and frame added to it.

Equation (8) reveals several other noteworthy properties. First, it indicates that the transmission loss of the controlled module depends upon the interstice surface area and not the diaphragm surface area. Second, it shows that the transmission loss is independent of the actuator diaphragm mass. Hence, the configuration could employ any piston-like diaphragm with appropriate dimensions. Third, it demonstrates that transmission loss increases with increased interstice im-

pedance. For a lightweight ASP, the interstice does not necessarily have to be massive to have high impedance in the direction normal to the partition. It might be oriented so that it is deeper from the source side to the receiving side and proportionately narrower in the plane of the partition. This would significantly increase stiffness in the direction normal to the partition without increasing mass. It would also produce the simultaneous benefit of decreasing the transmitting interstice surface area S_2, which as seen in Eq. (8) would increase transmission loss even more.

Another important property of the controlled configuration is its pressure-amplitude reflection coefficient R as seen from the source space. If $|R| > 1$, the module adds energy to the semi-infinite one-dimensional field, at least relative to that encountered in the presence of a perfectly rigid boundary. On the other hand, if $|R| < 1$, it absorbs at least a portion of the incident energy.[31] Thus, to avoid exacerbating noise levels in the primary field, a desirable requirement might simply be that $|R| \leq 1$. The reflection coefficient determined from the residual volume velocity $\hat{U}_{\mathrm{I}}$ is

$$R = \frac{Z_{M2}}{2\rho_0 c S_2 + Z_{M2}}, \tag{10}$$

which satisfies this requirement. In the high-frequency limit (i.e., when the interstice is fully mass controlled), $R \to 1$.

Based on these analyses, it appears that if one could make the interstitial surface area vanishingly small and its normal stiffness extremely large, the sound transmission through the controlled module would become correspondingly small. However, Sec. II B discusses an important detail that has been neglected and that significantly limits the performance of realizable SCL configurations.

2. Numerical example

Table I provides several numerical values for a representative example of the configuration. The parameters are used with Eqs. (4) and (8) to generate normal-incidence transmis-

TABLE I. Parameter values used in the numerical example for ASP module configuration 1.

Parameter	Value	Parameter	Value
Bl	5 T·m	R_g	0.1 Ω
c	343 m/s	R_{M2}	3 kg/s
C_{M2}	200 μm/N	R_{M12}	1 kg/s
C_{M12}	500 μm/N	S	400 cm²
L_E	0.5 mH	S_1	75 cm²
M_{M1}	7 g	S_2	325 cm²
M_{M2}	325 g	ρ_0	1.21 kg/m³
R_E	6 Ω		

FIG. 3. Normal-incidence transmission loss for ASP module configuration 1 under open-circuit passive and active conditions (calculated using numerical values from Table I).

FIG. 5. Isolated view of ASP module configuration 2. An actuator diaphragm is connected to a surrounding resilient suspension, which is connected on its opposite side to a stiffened interstice. The surround is represented with lumped mechanical elements. The model isolates effects of acoustic forces and diaphragm forces on the resilient surround. (After Ref. 7.)

sion loss curves for passive and active conditions, as shown in Fig. 3. As the curves demonstrate, control of sound transmission is poor for both states at lower audible frequencies. However, performance gradually improves to modest levels at higher frequencies. Active control produces a 9-dB improvement over passive control at these higher frequencies, but it performs no better at very low frequencies and actually performs worse in the narrow band centered at 58 Hz.

The values in Table I are also used with Eqs. (5) and (10) to provide a similar comparison of source-side reflection coefficient moduli under passive and active conditions (see Fig. 4). In both Figs. 3 and 4, curve dips correspond to resonances of vibrating module elements. The dip centered at 87 Hz for the passive condition corresponds to the resonance of the resiliently mounted diaphragm. Because the diaphragm motion is minimized via active control, this dip vanishes in the active state. The dip centered at 20 Hz for both passive and active conditions corresponds to the dominant resonance of the interstice (with the mounted actuator).

B. Configuration 2: Single-composite leaf (SCL) with a surrounding suspension

The analysis of the first module configuration assumed the actuator diaphragm vibrated freely but snugly within the

FIG. 4. Source-side reflection coefficient modulus for ASP module configuration 1 under open-circuit passive and active conditions (calculated using numerical values from Table I).

surrounding interstice opening so that no sound could transmit through an intervening crack. In practice, this idealized behavior would be extremely difficult to achieve. An airtight seal between an actuator diaphragm and a surrounding interstice would be realized with a mechanical connection. This connection must be resilient to produce sufficient suspension compliance and reduce undesirable mechanical coupling between adjacent ASP elements (a presumed requirement for controller simplification). A compliant airtight suspension with finite transmitting surface area is therefore introduced between the diaphragm and interstice as shown in Fig. 5. This suspension is analogous to the surround of a typical moving-coil loudspeaker. When used in conjunction with an adequately spaced rear suspension (e.g., a spider), it forms a composite dual suspension that significantly restricts diaphragm motion to translational rigid-body motion. As a transmitting surface, a surrounding suspension possesses degrees of vibrational freedom that may be difficult to control. Since it responds somewhat independently to impinging sound waves, a knowledge of its behavior is crucial to understand the performance of the module when the normal velocity of the transmitting diaphragm is minimized through actuation.

1. Equivalent circuit modeling and analytical results

For this model, the interstice is assumed to have infinite stiffness in the direction normal to the partition plane, and its in-plane area is assumed to be negligible.[33] As a result, dominant effects of the suspension are well isolated and a useful representation of the module is given by the equivalent circuit in Fig. 6. The surrounding suspension is modeled as a lumped mass connected between the actuator diaphragm and the rigid interstice with compliance and resistance on both sides. Solving the circuit as described in Sec. II A 1, produces the following normal surface velocities for the diaphragm and surround:
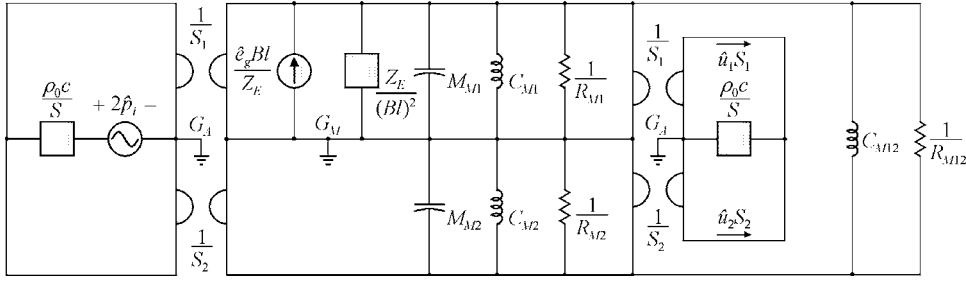
FIG. 6. Multiple-domain equivalent circuit representing ASP module configuration 2. (After Ref. 7.)

$$\hat{u}_1 = \frac{2\hat{p}_i(Z_{M2}S_1 + Z_{M12}S) + \hat{e}_g\dfrac{Bl}{Z_E}\left(2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}\right)}{\left[2\dfrac{\rho_0 c}{S}S_1^2 + Z_{M1} + \dfrac{(Bl)^2}{Z_E} + Z_{M12}\right]\left(2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}\right) - \left(2\dfrac{\rho_0 c}{S}S_1S_2 - Z_{M12}\right)^2}, \tag{11}$$

$$\hat{u}_2 = \frac{2\hat{p}_i\left\{\left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right]S_2 + Z_{M12}S\right\} - \hat{e}_g\dfrac{Bl}{Z_E}\left(2\dfrac{\rho_0 c}{S}S_1S_2 - Z_{M12}\right)}{\left[2\dfrac{\rho_0 c}{S}S_1^2 + Z_{M1} + \dfrac{(Bl)^2}{Z_E} + Z_{M12}\right]\left(2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}\right) - \left(2\dfrac{\rho_0 c}{S}S_1S_2 - Z_{M12}\right)^2}. \tag{12}$$

The total volume velocity is then

$$\hat{U}_\mathrm{I} = \hat{U}_\mathrm{II} = \frac{2\hat{p}_i\left\{\left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right]S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2\right\} + \hat{e}_g\dfrac{Bl}{Z_E}(Z_{M2}S_1 + Z_{M12}S)}{\left[2\dfrac{\rho_0 c}{S}S_1^2 + Z_{M1} + \dfrac{(Bl)^2}{Z_E} + Z_{M12}\right]\left(2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}\right) - \left(2\dfrac{\rho_0 c}{S}S_1S_2 - Z_{M12}\right)^2}. \tag{13}$$

While this configuration is notably different than configuration 1, its analytical representations for passive $1/\tau$ and R are identical to those given in Eqs. (4) and (5) (except that the definition of Z_{M1} differs as shown in the List of Symbols). The control voltage required to drive the normal diaphragm velocity $\hat{u}_1$ to zero is

$$\hat{e}_g = -2\hat{p}_i\frac{Z_E}{Bl}\left(\frac{Z_{M2}S_1 + Z_{M12}S}{2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}}\right). \tag{14}$$

When it is applied, the normal surround velocity becomes

$$\hat{u}_2 = \frac{2\hat{p}_i S_2}{2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}}. \tag{15}$$

The resulting volume velocity for the controlled condition then produces the inverse transmission coefficient

$$\frac{1}{\tau} = \left|1 + \frac{Z_{M2} + Z_{M12}}{2\dfrac{\rho_0 c}{S}S_2^2}\right|^2. \tag{16}$$

If the surround becomes mass controlled, $Z_{M2} + Z_{M12} \approx j\omega M_{M2} = j\omega\sigma_{M2}S_2$ and the inverse transmission coefficient becomes

$$\frac{1}{\tau} \approx 1 + \left(\frac{\omega\sigma_{M2}}{2\rho_0 c}\right)^2\left(\frac{S}{S_2}\right)^2. \tag{17}$$

The reflection coefficient for the controlled configuration is given by the relationship

$$R = \frac{Z_{M2} + Z_{M12}}{2\dfrac{\rho_0 c}{S}S_2^2 + Z_{M2} + Z_{M12}}. \tag{18}$$

This once again satisfies the requirement that $|R| \leqslant 1$ and in the high-frequency limit, $R \to 1$.

The results given above indicate that the ways to reduce sound transmission through the resilient surround are (1) to reduce its relative surface area and (2) to increase its mechanical impedance. However, increasing its mechanical mass, stiffness, or resistance has important disadvantages. Dramatically increasing its mass is not a good option if one is trying to minimize overall partition mass. The amount that its stiffness or resistance can be increased will likewise be restricted if the surround is to remain sufficiently resilient. Thus, although these quantities may be increased within prescribed limits, the most important approach to reducing

T. W. Leishman and J. Tichy: Active segmented partition modules

FIG. 7. Normal-incidence transmission loss for ASP module configuration 2 under open-circuit passive and active conditions (calculated using numerical values from Table II).

FIG. 8. Source-side reflection coefficient modulus for ASP module configuration 2 under open-circuit passive and active conditions (calculated using numerical values from Table II).

transmission through the surround is to minimize its exposed surface area. This too has practical limitations.

2. Numerical example

Figure 7 uses the analytical results and the parameter values listed in Table II to provide a numerical comparison of the module transmission loss under open-circuit passive and active conditions. Figure 8 provides a similar comparison of the source-side reflection coefficient moduli. In the passive curves of both figures, deep dips centered at 27 Hz are caused by the dominant resonance of the resiliently suspended diaphragm. The shallower and broader dips centered near 235 Hz are produced by the dominant surround resonance. Active control minimizes diaphragm motion, resulting in significant improvement to transmission loss at lower frequencies. However, since the surround remains free to vibrate, the dips near 235 Hz persist. In addition, from 125 to 225 Hz, active control produces transmission loss that is less than that produced by passive control. A comparison of Figs. 3 and 7 also reveals that this active module configuration performs better than configuration 1 at low frequencies (i.e., below 125 Hz) but worse at higher frequencies.

TABLE II. Parameter values used in the numerical example for ASP module configuration 2.

Parameter	Value	Parameter	Value
Bl	5 T·m	R_g	0.1 Ω
c	343 m/s	R_{M1}	2 kg/s
C_{M1}	400 μm/N	R_{M2}	2 kg/s
C_{M2}	200 μm/N	R_{M12}	2 kg/s
C_{M12}	200 μm/N	S	400 cm^2
L_E	0.5 mH	S_1	325 cm^2
M_{M1}	175 g	S_2	75 cm^2
M_{M2}	5 g	ρ_0	1.21 kg/m^3
R_E	6 Ω		

C. Configuration 3: Double-composite leaf (DCL) with acoustic transmitting surface actuation

The preceding sections have provided useful guidelines to increase transmission loss through interstices and actuator diaphragm assemblies. However, they have also shown that the vibration-controlled SCL modules are significantly limited in the ASTC performance they can achieve. Because of this limitation and other problems associated with their incorporation into ASP arrays, the authors chose to explore several other module configurations.[13]

One noteworthy configuration is depicted in Fig. 9. It is a DCL arrangement that incorporates many of the ideas discussed for the preceding configurations but also adds interesting new features. It includes an actuator with a piston-like diaphragm and surround. The actuator frame is rigidly mounted to a stiffened surrounding interstice. Because the interstice impedance normal to the partition is extremely large (assumed to be infinite) and the exposed transmitting surface area is very small, its vibration and contribution to sound transmission are considered negligible. The actuator
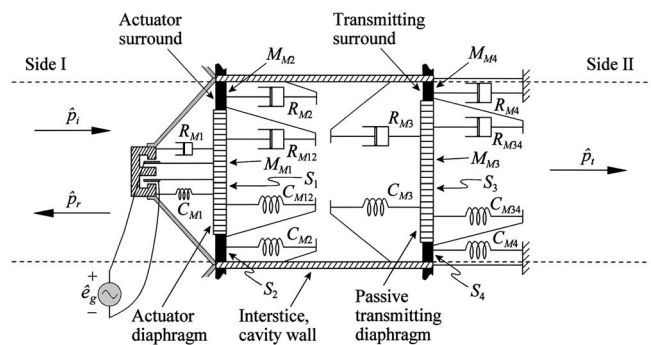
FIG. 9. Isolated view of ASP module configuration 3. An actuator diaphragm is connected to a surrounding resilient suspension, which is connected on its opposite side to a stiffened interstice. The surround is modeled with lumped mechanical elements. A passive transmitting diaphragm is similarly supported by a surrounding resilient suspension and a spaced rear suspension. The source and transmitting sides of the module are separated and acoustically coupled by a small isolated cavity. The model demonstrates useful methods mechano-acoustic actuation and segmentation.
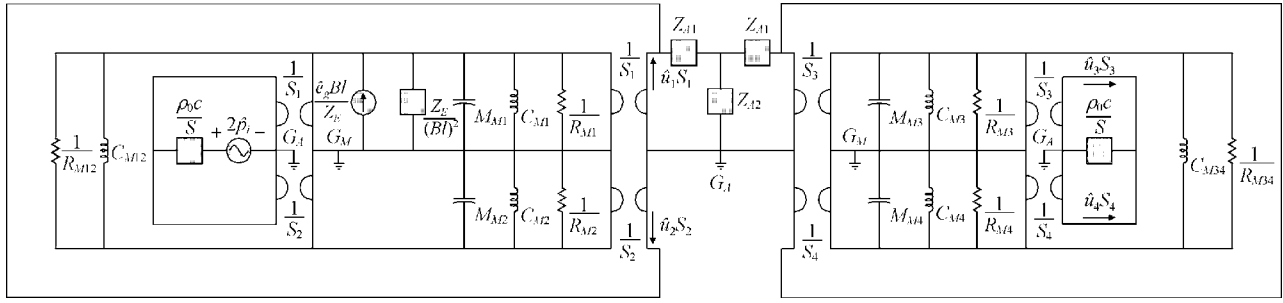
FIG. 10. Multiple-domain equivalent circuit representing ASP module configuration 3.

drives a small sealed cavity to *acoustically* actuate a passive piston-like diaphragm and surround. Both transmitting surfaces are significant. A primary focus of the model is to evaluate the effects of transmitting surround vibration when the module is passive and when its transmitting diaphragm vibration is minimized via active control. As suggested in Sec. II B, performance limitations imposed by transmitting surrounds are difficult to control passively and are therefore of considerable interest. The exposed surface area of the surround may be chosen to be small and its impedance may be chosen to be as large as possible within prescribed limits. The transmitting diaphragm is modeled with connection to both the resilient surround and a spaced rear suspension, the combination helping to constrain its rigid-body translational motion.

1. Equivalent circuit modeling and analytical results

A planar equivalent circuit representing the configuration is shown in Fig. 10. Excitation, radiation, and coupling in the acoustic impedance portions of the circuit are again represented for one-dimensional fields. The simplification neglects higher-order modes and certain near-field coupling effects in the cavity (both of which may merit further analysis), but it acceptably represents cross-sectional spatially averaged field quantities.

Nine nodal equations from the circuit reduce to four coupled equations of motion for the actuator diaphragm, actuator surround, transmitting diaphragm, and transmitting surround, yielding normal surface velocities that are again functions of both incident pressure and electrical input voltage. The velocities of the first two elements combine with their surface areas to produce the following source-side volume velocity:

$$\hat{U}_{\mathrm{I}} = \frac{2\hat{p}_i\left\{\left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right]S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2\right\} + \hat{e}_g\dfrac{Bl}{Z_E}(Z_{M2}S_1 + Z_{M12}S)}{\left\{\left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right]S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2\right\}Z_{A4} + \left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right]Z_{M2} + \left\{\left[Z_{M1} + \dfrac{(Bl)^2}{Z_E}\right] + Z_{M2}\right\}Z_{M12}}. \tag{19}$$

The normal surface velocities of the transmitting diaphragm and surround are

$$\hat{u}_3 = \hat{U}_{\mathrm{I}}\left[\frac{Z_{A2}(Z_{M4}S_3 + Z_{M34}S)}{(Z_{M3}S_4^2 + Z_{M4}S_3^2 + Z_{M34}S^2)Z_{A3} + Z_{M3}Z_{M4} + (Z_{M3} + Z_{M4})Z_{M34}}\right], \tag{20}$$

and

$$\hat{u}_4 = \hat{u}_3\left(\frac{Z_{M3}S_4 + Z_{M34}S}{Z_{M4}S_3 + Z_{M34}S}\right). \tag{21}$$

These velocities also combine with their surface areas to produce the transmitting volume velocity

$$\hat{U}_{\mathrm{II}} = \hat{U}_{\mathrm{I}}\left(\frac{Z_{A3} - Z_{A4}}{Z_{A2}}\right). \tag{22}$$

Under open-circuit passive conditions, the inverse transmission coefficient then becomes

$$\frac{1}{\tau} = \left|\frac{[(Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2)Z_{A4} + Z_{M1}Z_{M2} + (Z_{M1} + Z_{M2})Z_{M12}]Z_{A2}}{2\dfrac{\rho_0 c}{S}(Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2)(Z_{A3} - Z_{A4})}\right|^2, \tag{23}$$

and the source-side reflection coefficient becomes
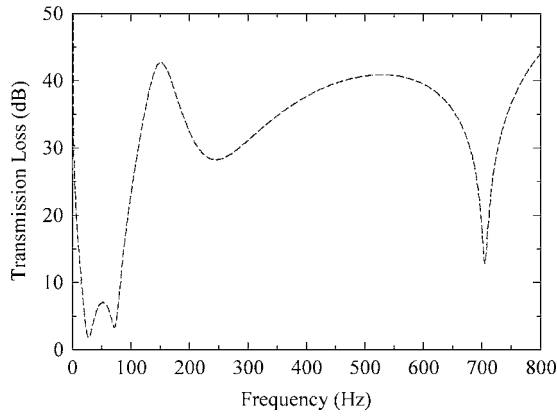
FIG. 11. Normal-incidence transmission loss for ASP module configuration 3 under an open-circuit passive condition (calculated using numerical values from Table III). The predicted infinite transmission loss produced by active control is not plotted.

TABLE III. Parameter values used in the numerical example for ASP module configuration 3.

Parameter	Value	Parameter	Value
Bl	5 T·m	R_E	6 Ω
c	343 m/s	R_g	0.1 Ω
C_{M1}	400 μm/N	R_{M1}	2 kg/s
C_{M2}	200 μm/N	R_{M2}	2 kg/s
C_{M3}	400 μm/N	R_{M3}	2 kg/s
C_{M4}	200 μm/N	R_{M4}	2 kg/s
C_{M12}	200 μm/N	R_{M12}	2 kg/s
C_{M34}	200 μm/N	R_{M34}	2 kg/s
L	25 cm	S	400 cm^2
L_E	0.5 mH	S_1	325 cm^2
M_{M1}	175 g	S_2	75 cm^2
M_{M2}	5 g	S_3	325 cm^2
M_{M3}	175 g	S_4	75 cm^2
M_{M4}	5 g	ρ_0	1.21 kg/m^3

$$R = 1 - 2\frac{\rho_0 c}{S}\left[\frac{Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2}{(Z_{M1}S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2)Z_{A4} + Z_{M1}Z_{M2} + (Z_{M1} + Z_{M2})Z_{M12}}\right]. \tag{24}$$

As seen from Eq. (20), the control strategy of minimizing normal vibration of the transmitting diaphragm is accomplished by driving the *source-side* volume velocity to zero such that $\hat{u}_1 S_1 = -\hat{u}_2 S_2$. Solving Eq. (19) for the required control voltage yields

$$\hat{e}_g = -2\hat{p}_i \frac{Z_E}{Bl}\left\{\frac{\left[Z_{M1} + \frac{(Bl)^2}{Z_E}\right]S_2^2 + Z_{M2}S_1^2 + Z_{M12}S^2}{Z_{M2}S_1 + Z_{M12}S}\right\}. \tag{25}$$

Inspection of Eqs. (21) and (22) further reveals that this voltage and the resulting zero source-side volume velocity simultaneously drive the transmitting surround velocity and the transmitting volume velocity to zero. Thus, with the interstice free from vibration (or otherwise producing negligible sound transmission), the transmission loss of the controlled configuration tends (within modeling constraints) to infinity (i.e., $1/\tau \to \infty$). The source-side reflection coefficient becomes $R = 1$.

2. Numerical example

Figure 11 provides an example of the open-circuit passive transmission loss of the module using the parameters listed in Table III. (The predicted infinite transmission loss is not shown for the active state.) A comparison of passive and active source-side reflection coefficient moduli is given in Fig. 12. The curves show multiple dips corresponding to resonances of the interacting module elements for the open-circuit passive state. Some of these effects are characteristic of common double-leaf partitions, while others result specifically from segmentation. The narrow dip centered at 705 Hz corresponds to the first cavity resonance. The broader dip, centered at about 235 Hz, corresponds to the source-side sur-

round resonance (compare Figs. 7 and 8). As with most passive double-leaf constructions, transmission loss is relatively high at high frequencies, but poor at very low frequencies. Nevertheless, active control produces a substantial increase in transmission loss at all frequencies as long as the module dimensions are small compared with the acoustic wavelength.

D. Configuration 4: Double-composite leaf (DCL) with resilient actuator support

An enhanced DCL configuration is shown in Fig. 13. While it is similar to configuration 3 in many respects, it incorporates a few distinguishing features that improve its integration and control in extended ASP arrays. The actuator frame in configuration 3 was rigidly mounted to the interstice, which was assumed to have infinite impedance. In practice, the interstice impedance would be large but finite.
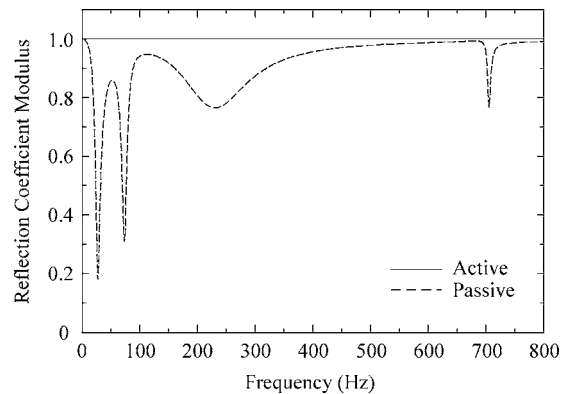
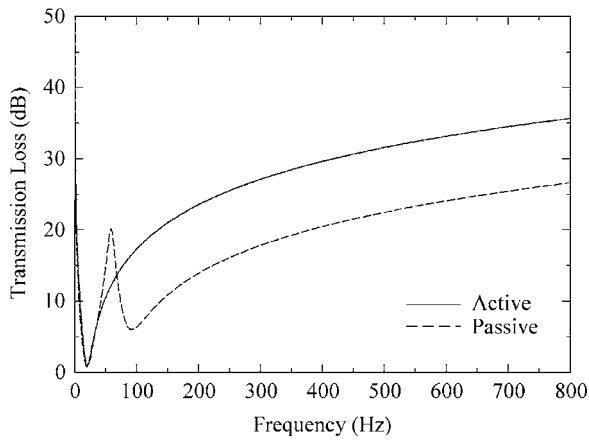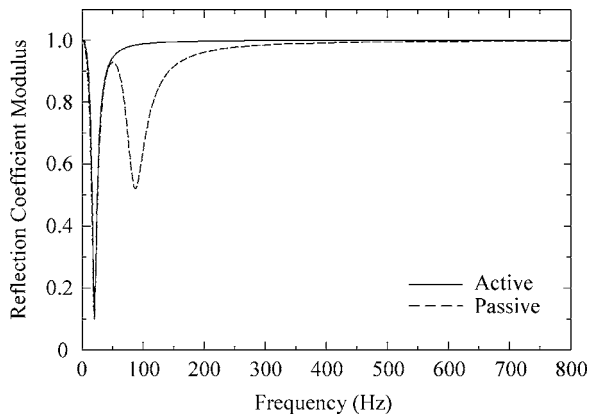FIG. 12. Source-side reflection coefficient modulus for ASP module configuration 3 under open-circuit passive and active conditions (calculated using numerical values from Table III).
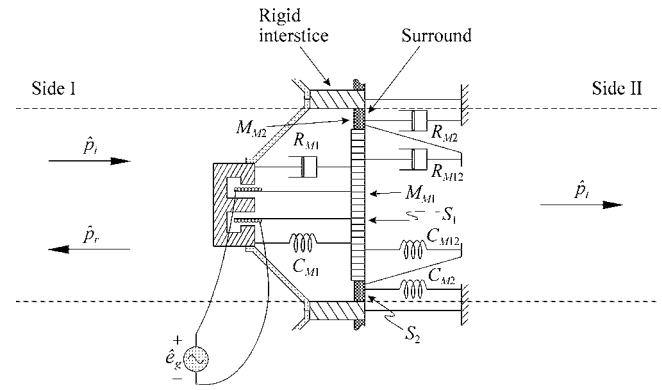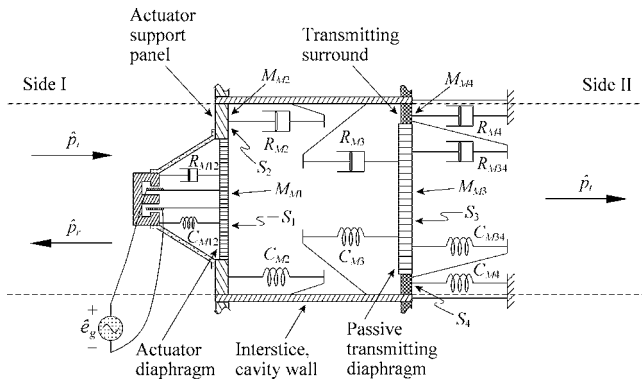
FIG. 13. Isolated view of ASP module configuration 4. An actuator diaphragm vibrates freely but snugly within a frictionless support panel opening. The support panel vibrates similarly within a surrounding interstice opening and is connected to the stiffened interstice by a resilient suspension. A passive transmitting diaphragm is supported by a surrounding resilient suspension and a spaced rear suspension, both modeled with lumped mechanical elements and connected on their opposite sides to the interstice. The source and transmitting sides of the module are separated and acoustically coupled by a small isolated cavity. The model demonstrates useful methods of mechano-acoustic actuation, segmentation, and isolation. (After Ref. 7.)

The reaction of the actuator frame against the interstice would then force it to vibrate to some degree. Configuration 4 accordingly adds a measure of isolation between the actuator and interstice through an inertial support panel. The actuator frame is rigidly mounted to the panel, but the panel is attached to the interstice via a resilient connection. The panel is characterized as a lumped-element piston that vibrates snugly within the frictionless interstice opening. The actuator diaphragm is likewise assumed to vibrate freely but snugly as a piston within the frictionless panel opening, with mechanical resistance and compliance coupling it to the actuator frame.

Sound transmission into the cavity through intervening cracks or surrounds is basically inconsequential. In practice, resilient surrounds must be incorporated with the actuator diaphragm and support panel. However, their areas are neglected here (as with configuration 1) to allow modeling with a planar equivalent circuit. This simplification is acceptable for the source-side elements because, as seen from the analysis of configuration 3, the control condition that minimizes transmitting surface vibrations is a zero source-side volume velocity. Vibration control of the principal transmitting surface and small-cavity acoustic coupling of the transmitting surfaces to the source-side surfaces ensure that the condition is essentially satisfied for any number of source-side surfaces. Hence, the actuator diaphragm and support panel are considered to be the only significant source-side surfaces without much loss in generality.

As with configuration 3, the single actuator controls vibrations of both the transmitting diaphragm and its surround through acoustic actuation. Although the concept of acoustic actuation of *distributed* double-leaf partition cavities is not new,[34–38] important distinctions of this configuration make its actuation and behavior unique in several ways. First, when incorporated into an array, several of these modules form an extended partition that is structurally segmented into isolated elements on both its source side and transmitting side. Both the actuator support panel suspension and the transmitting diaphragm suspension help reduce mechanical coupling to the interstice and adjacent modules. Second, the distinctive pair of diaphragms (actuator diaphragm and transmitting diaphragm) and the actuator support panel vibrate as discrete translation pistons. Third, the actuator diaphragm and support panel produce a zero volume velocity condition on the source side of the cavity. Fourth, the actuator and other elements are configured to produce local cross-sectional symmetries. Fifth, the overall partition cavity is segmented by the interstices into acoustically small cavities. This last characteristic has the important effect of *constraining* primary (noise) and secondary (control) pressure distributions within each cavity, and therefore upon the inner faces of the transmitting diaphragms and surrounds. The pressure distributions are thus compelled to match more closely, yielding improved active control performance.

In reality, acoustic pressure within a cavity is never completely uniform and becomes less uniform with increasing frequency. However, if the transmitting diaphragm is constrained to vibrate only as a translational rigid body, it will respond to the total *spatially averaged* pressure over its cavity-side face. Control of its normal velocity will produce an advantageous effect: the same actuation that minimizes its vibration via spatially averaged pressure will also suppress vibration of the transmitting surround over a broad frequency range. Accordingly, the approach of the configuration is not merely one of acoustic cavity actuation, but an important combination of mechano-acoustic isolation, actuation, and segmentation.

1. Equivalent circuit modeling and analytical results

An equivalent circuit representing the configuration is shown in Fig. 14. Its nine nodal equations again reduce to four coupled equations of motion for the actuator diaphragm, actuator support panel, transmitting diaphragm, and transmitting surround. The velocities and surface areas of the first two elements yield the source-side volume velocity

$$\hat{U}_I = \frac{2\hat{p}_i\left\{ Z_{M1}S_2^2 + Z_{M2}S_1^2 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]S^2 \right\} + \hat{e}_g\frac{Bl}{Z_E}(Z_{M2}S_1 - Z_{M1}S_2)}{\left\{ Z_{M1}S_2^2 + Z_{M2}S_1^2 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]S^2 \right\}Z_{A4} + Z_{M1}Z_{M2} + (Z_{M1} + Z_{M2})\left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right]}. \quad (26)$$
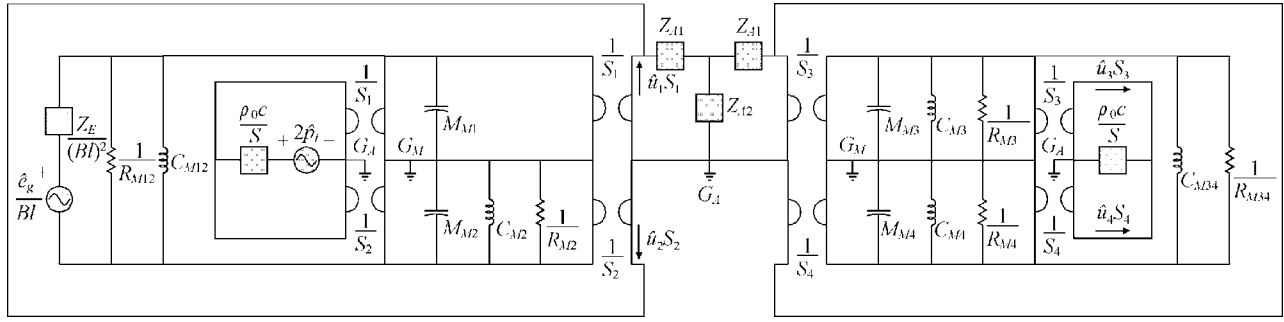
FIG. 14. Multiple-domain equivalent circuit representing ASP module configuration 4. (After Ref. 7.)

Using this result, $\hat{u}_3$, $\hat{u}_4$, and $\hat{U}_{\text{II}}$ are represented as shown in Eqs. (20)–(22). Although the source-side elements are quite different than those of configuration 3, the open-circuit passive representations for $1/\tau$ and R are also identical to those given in Eqs. (23) and (24) (except for the difference in Z_{M1}, as defined in the List of Symbols).

The normal transmitting diaphragm velocity may be driven to zero with the following actuator control voltage:[39]

$$
\hat{e}_g = -2\hat{p}_i \frac{Z_E}{Bl} \left\{ \frac{Z_{M1}S_2^2 + Z_{M2}S_1^2 + \left[Z_{M12} + \frac{(Bl)^2}{Z_E} \right] S^2}{Z_{M2}S_1 - Z_{M1}S_2} \right\}.
$$

(27)

As indicated earlier, this voltage simultaneously drives the transmitting surround velocity and transmitting volume velocity to zero through the source-side condition $\hat{u}_1 S_1 = -\hat{u}_2 S_2$.[40] The transmission loss thus tends (within modeling constraints) to infinity and the source-side reflection coefficient takes on the constant value of $R = 1$. While this reflection coefficient does not reduce acoustic energy in the semi-infinite source space, it does not add energy to the space above that which would occur with a completely rigid boundary.

2. Numerical example

Figure 15 provides an example of the open-circuit passive transmission loss of the module using the parameters
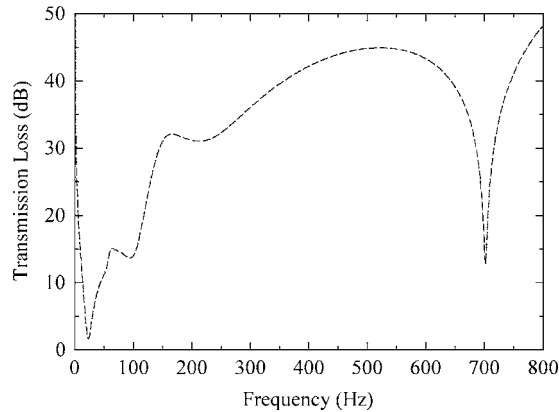
FIG. 15. Normal-incidence transmission loss for ASP module configuration 4 under an open-circuit passive condition (calculated using numerical values from Table IV). The predicted infinite transmission loss produced by active control is not plotted.

listed in Table IV. A comparison of passive and active source-side reflection coefficient moduli is given in Fig. 16. Because of the similarity in module constructions, the curves are similar in some respects to those given in Figs. 11 and 12. Differences follow from the changes to the source-side module elements. Once again, active control produces a considerable increase in transmission loss at all frequencies in which the module dimensions are small compared with the acoustic wavelength.

3. Design and performance criteria

This module configuration satisfies the seven design and performance criteria mentioned in Sec. I. First, it is designed to utilize simple localized sensors and actuators. This characteristic is important because consistent local control mechanisms allow the module to perform similarly despite changes in adjacent source and receiving spaces. It also allows the module to be self-contained with the possibility of decentralized control. Second, its design and control enable efficient global reduction of normal transmitting surface vibrations (although interstice vibration and transmission are controlled passively). While multiple actuation mechanisms are usually required to control multiple transmitting surface vibrations, this configuration permits efficient control of two vibrating surfaces using only a single actuator. Furthermore, because the control is induced by minimizing vibration of a single transmitting surface, the configuration simplifies error sensing.

TABLE IV. Parameter values used in the numerical example for ASP module configuration 4.

Parameter	Value	Parameter	Value
Bl	5 T·m	R_E	6 Ω
c	343 m/s	R_g	0.1 Ω
C_{M2}	200 μm/N	R_{M2}	3 kg/s
C_{M3}	400 μm/N	R_{M3}	2 kg/s
C_{M4}	200 μm/N	R_{M4}	2 kg/s
C_{M12}	500 μm/N	R_{M12}	1 kg/s
C_{M34}	200 μm/N	R_{M34}	2 kg/s
L	25 cm	S	400 cm^2
L_E	0.5 mH	S_1	75 cm^2
M_{M1}	7 g	S_2	325 cm^2
M_{M2}	325 g	S_3	325 cm^2
M_{M3}	175 g	S_4	75 cm^2
M_{M4}	5 g	ρ_0	1.21 kg/m^3

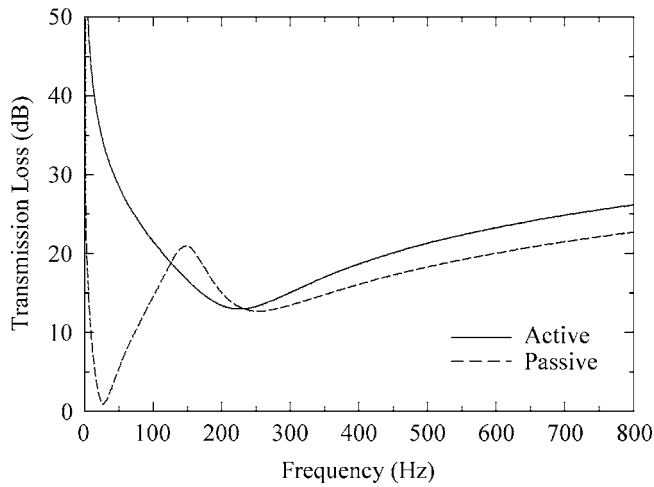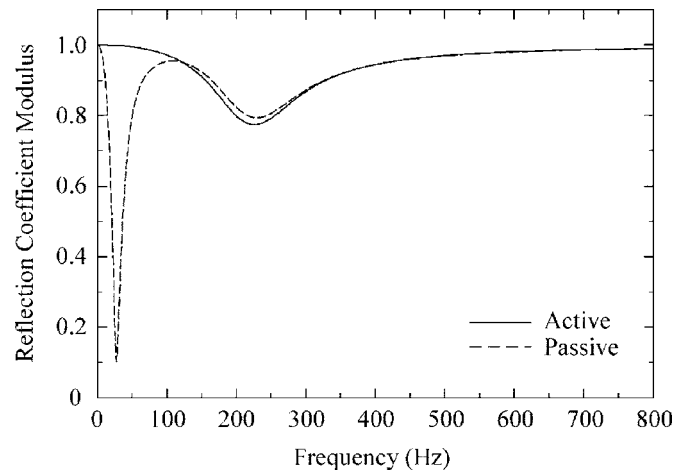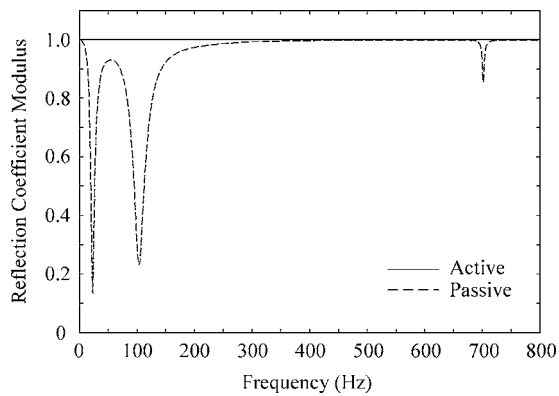FIG. 16. Source-side reflection coefficient modulus for ASP module configuration 4 under open-circuit passive and active conditions (calculated using numerical values from Table IV).

Third, the configuration may be constructed using appropriate lightweight materials. To ensure lumped-element behavior, key structural elements should be acoustically small and stiffness controlled over frequencies of interest. Fourth, the configuration adds no sound energy to a one-dimensional semi-infinite source space, at least compared to that encountered with a completely rigid boundary. Fifth, the configuration can accommodate simple passive enhancements and function as a hybrid active–passive device. One enhancement of particular interest is the addition of porous absorptive material to the module cavity to improve passive high-frequency transmission loss. The module could then function as an active device at low frequencies, then "cross over" into effective passive control at higher frequencies.

Sixth, the configuration uses small, isolated, lumped-element components with constrained fields and vibrational motions to reduce problems with controllability, observability, and spillover. Seventh, its segmentation and isolation minimize coupling between adjacent partition modules to simplify control requirements and increase the feasibility of decentralized control.

III. ADDITIONAL DISCUSSION OF CAPABILITIES AND LIMITATIONS

The idealized behavior of each module configuration is unavoidably affected by nonideal physical conditions. Some performance limitations result from nonideal vibrational responses. Controlled transmitting components may have residual vibrations or fail to vibrate as ideal rigid bodies. Bending or rotational rigid-body modes ("rocking modes") of transmitting surfaces complicate the behaviors of modules and may reduce their effectiveness. They also affect the assumed accuracy of uniform or spatially averaged pressures employed in equivalent circuits.

Residual vibrations of narrow transmitting surfaces include those of exposed interstices and transmitting diaphragm surrounds. When radiating into a semi-infinite three-dimensional receiving space, these surfaces are similar to baffled annular pistons or strips. Studies of such radiators[41,42] indicate that radiation from a thin interstice or surround becomes very inefficient when its in-plane width δ becomes very small compared to wavelength, i.e., $k\delta \ll 1$. Further-

more, it can be shown[41] that reducing δ while holding the characteristic cross-sectional dimension a of a module constant has greater impact on radiation reduction than reducing a by the same factor with δ held constant. Moreover, with a fixed-dimension array of many modules, minimizing δ remains the only practical method of surface area reduction.

Three of the module configurations in this investigation were assumed to incorporate narrow interstices with infinite impedance in the direction normal to the plane of the partition. If a narrow interstice has finite impedance and does vibrate to a degree, its radiation into a receiving space will be inherently inefficient. However, its vibration will also force limited vibration of the adjacent transmitting diaphragm and surround. Despite this problem, the normal vibration of the principal transmitting surface is still controlled with appropriate actuation.

When translational motion of a transmitting rigid body is effectively reduced, low-amplitude residual rocking may persist. Nevertheless, if a baffled transmitting face of a rigid body rocks symmetrically with respect to its center point (i.e., with odd symmetry), it produces zero net volume velocity. Modules should accordingly be designed to ensure that any residual transmitting surface vibrations fall into this category. When transmitting faces of an ASP are small compared to wavelength, such consistent reduction in volume velocity provides beneficial reduction in sound transmission.[9,10,43] Notably, the method described here is more than just volume velocity control involving large, spatially dependent surface velocities. Instead, the approach produces sound transmission control in two steps. First, it significantly suppresses normal vibration of efficiently transmitting modes. Second, it leaves only low-amplitude residual vibration with a distribution that transmits inefficiently into the receiving space.

The ASP design and performance criteria discussed in Sec. II D 3 require that normal vibrations of *all* transmitting surfaces be minimized to the greatest extent possible. Because module configurations 3 and 4 were found to accomplish the task well, they are expected to simultaneously reduce transmission into both receiving space near fields and far fields. However, near fields produced by narrow residually vibrating surfaces and rocking surfaces possess reactive acoustic energy that cannot be wholly neglected. Such residual vibrations must be carefully considered when listeners or acoustic receivers are positioned close to partitions.

Simultaneous control of the source-side reflection coefficient or the source-space sound field may be desirable while an ASP is engaged in controlling sound transmission. Additional actuation mechanisms may be incorporated into modules to help provide this control or help control residual interstice vibration.[13] While the requirements to accomplish these additional tasks make the problem more complex, the concepts could find application in certain sound transmission problems.

It is anticipated that carefully designed ASP modules will function effectively in complete ASP arrays using simple adaptive controllers. Decentralized controllers may require reduction of coupling between adjacent partition modules to weaken cross-coupling control paths (i.e., trans-

fer functions between actuators and error sensors of neighboring modules).[44] Of the module configurations presented in this investigation, configuration 4 appears to be most capable of producing effective isolation.

IV. CONCLUSIONS

This work has provided a theoretical and numerical investigation of several vibration-controlled modules for use in active segmented partitions (ASPs). These partitions consist of physically segmented arrays of acoustically and structurally small modules that may be regulated using adaptive controllers. The control objective for each module was to induce global control of its various normal transmitting surface vibrations by directly minimizing normal vibration of its *principal* transmitting surface. In some cases, passive means were used to control interstitial surface vibrations. Analyses of the modules were realized using electro-mechano-acoustic circuits. While certain simplifying assumptions were required to make the circuits manageable, they clearly demonstrated important module behaviors. Transmission loss estimates from the analyses are well suited for comparison with classical normal-incidence transmission loss formulations.

The investigation has revealed the inability of two characteristic single-composite-leaf (SCL) configurations to simultaneously control vibration of their transmitting diaphragms, resilient suspensions, and partition interstices under the active control scheme. A numerical example showed that the first configuration produced poor active sound transmission control (ASTC) at low frequencies and moderate ASTC at high frequencies. Another example showed that the second configuration improved low-frequency performance but at the expense of high-frequency performance. In both cases, the introduction of active control was shown to *decrease* transmission loss below passive levels at some frequencies.

Despite these limitations, the investigation provided useful recommendations to enhance ASTC capabilities of these and other vibration-controlled configurations. They suggest that lightweight ASPs should incorporate interstices with very high impedance in the direction normal to the partition. Interstitial widths should also be very narrow in the plane of the partition. Resilient transmitting suspensions that connect stiff diaphragms to interstices should likewise be thin in the plane of the partition and have reasonably large impedances (within prescribed limits to maintain desired resiliency).

Two double-composite-leaf (DCL) configurations were introduced that utilize distinctive combinations of acoustic actuation and mechano-acoustic segmentation. In their passive states, both show transmission loss characteristics typical of double-leaf partitions but add other characteristics unique to their segmented natures. In their active states, they efficiently and globally control transmitting surface vibrations to produce high transmission loss over a broad frequency range. Moreover, because of their localized control mechanisms, their performances remain relatively unaffected by typical time-varying changes to surrounding environments. The second of the two configurations incorporates special segmentation and isolation to reduce coupling between adjacent partition modules and enhance the effectiveness of extended arrays using simple decentralized controllers.

This work has shown that appropriate vibration control of properly configured ASP modules can yield effective local control of sound transmission. When incorporated into full partition arrays, these modules have the potential to enhance the practicality and performance of ASTC systems through their straightforward design, implementation, actuation, error sensing, and application to many sound transmission problems.

[1] When appropriate (e.g., for analogous circuit modeling), spatially averaged acoustic pressures and particle velocities may be assumed over relevant cross-sectional areas.

[2] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000).

[3] J. E. Cole III and M. C. Junger, "Active noise control for machinery enclosures," NSF Final Report U-1944-379F (1991).

[4] D. R. Thomas, P. A. Nelson, and S. J. Elliott, "An experimental investigation into the active control of sound transmission through stiff light composite panels," Noise Control Eng. J. **41**, 273–279 (1993).

[5] D. R. Thomas, P. A. Nelson, R. J. Pinnington, and S. J. Elliott, "An analytical investigation of the active control of the transmission of sound through plates," J. Sound Vib. **181**, 515–539 (1995).

[6] J. E. Cole, K. F. Martini, and A. W. Stokes, "Active noise control for machinery enclosures," NSF Final Report U-2413-393 (1996).

[7] T. W. Leishman and J. Tichy, "A fundamental investigation of the active control of sound transmission through segmented partition elements," Proceedings of Noise-Con 97, University Park, PA, 1997, Vol. **2**., pp. 137–148.

[8] S. L. Sharp, G. H. Koopmann, and W. Chen, "Transmission loss characteristics of an active trim panel," Proceedings of Noise-Con 97, University Park, PA, 1997, Vol. **2**., pp. 149–160.

[9] M. E. Johnson and S. J. Elliott, "Active control of sound radiation from vibrating surfaces using arrays of discrete actuators," J. Sound Vib. **207**, 743–759 (1997).

[10] R. L. St. Pierre, Jr., G. H. Koopmann, and W. Chen, "Volume velocity control of sound transmission through composite panels," J. Sound Vib. **210**, 441–460 (1998).

[11] T. W. Leishman and J. Tichy, "An experimental evaluation of individual partition segment configurations for the active control of sound transmission," J. Acoust. Soc. Am. **104**, Pt. 2, 1776(A) (1998).

[12] T. W. Leishman and J. Tichy, "An experimental investigation of a novel active segmented partition for sound transmission control," J. Acoust. Soc. Am. **105**, Pt. 2, 1156(A) (1999).

[13] T. W. Leishman, "Active control of sound transmission through partitions composed of discretely controlled modules," Ph.D. thesis, The Pennsylvania State University, University Park, PA, 2000.

[14] S. M. Hirsch, J. Q. Sun, and M. R. Jolly, "An analytical study of interior noise control using segmented panels," J. Sound Vib. **231**, 1007–1021 (2000).

[15] S. M. Hirsch, N. E. Meyer, M. A. Westervelt, P. King, F. J. Li, M. V. Petrova, and J. Q. Sun, "Experimental study of smart segmented trim panels for aircraft interior noise," J. Sound Vib. **231**, 1023–1027 (2000).

[16] T. W. Leishman, "Vibration-controlled modulues for use in active segmented partitions," J. Acoust. Soc. Am. **114**, Pt. 2, 2385(A) (2003).

[17] C. R. Fuller, V. L. Metcalf, R. J. Silcox, and D. E. Brown, "Experiments on structural control of sound transmitted through an elastic plate," Proceedings of the 1989 American Control Conference, Pittsburgh, PA, Vol. **3**, pp. 2079–2084.

[18] C. R. Fuller, "Active control of sound transmission/radiation from elastic

plates by vibration inputs: I. Analysis," J. Sound Vib. **136**, 1–15 (1990).

[19] V. L. Metcalf, C. R. Fuller, R. J. Silcox, and D. E. Brown, "Active control of sound transmission/radiation from elastic plates by vibration inputs. II. Experiments," J. Sound Vib. **153**, 387–402 (1992).

[20] D. R. Thomas, P. A. Nelson, and S. J. Elliott, "Active control of the transmission of sound through a thin cylindrical shell. I. The minimization of vibrational energy," J. Sound Vib. **167**, 91–111 (1993).

[21] D. R. Thomas, P. A. Nelson, and S. J. Elliott, "Active control of the transmission of sound through a thin cylindrical shell. II. The minimization of acoustic potential energy," J. Sound Vib. **167**, 113–128 (1993).

[22] S. D. Snyder and N. Tanaka, "On feedforward active control of sound and vibration using vibration error sensors," J. Acoust. Soc. Am. **94**, 2181–2193 (1993).

[23] C. R. Fuller, C. H. Hansen, and S. D. Snyder, "Active control of sound radiation from a vibrating rectangular panel by sound sources and vibration inputs: An experimental comparison," J. Sound Vib. **145**, 195–215 (1991).

[24] S. D. Snyder and C. H. Hansen, "The design of systems to control actively periodic sound transmission into enclosed spaces. II. Mechanisms and trends," J. Sound Vib. **170**, 451–472 (1994).

[25] S. D. Snyder and C. H. Hansen, "Mechanisms of active noise control by vibration sources," J. Sound Vib. **147**, 519–525 (1991).

[26] J. Pan, C. H. Hansen, and D. A. Bies, "Active control of noise transmission through a panel into a cavity. I. Analytical study," J. Acoust. Soc. Am. **87**, 2098–2108 (1990).

[27] J. Pan and C. H. Hansen, "Active control of noise transmission through a panel into a cavity. II. Experimental study," J. Acoust. Soc. Am. **90**, 1488–1492 (1991).

[28] L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1986).

[29] H. F. Olson, *Acoustical Engineering* (Professional Audio Journals, Philadelphia, 1991).

[30] The transmission loss (TL) follows directly from this result as shown in the List of Symbols.

[31] These statements are not necessarily true for enclosed one-dimensional source spaces (Refs. 13 and 32).

[32] T. W. Leishman and J. Tichy, "On the significance of reflection coefficients produced by active surfaces bounding one-dimensional sound fields," J. Acoust. Soc. Am. **113**, 1475–1482 (2003).

[33] Transmission through the interstice is assumed to be completely controlled by its passive characteristics. Active control of interstice vibration is also possible (Ref. 13).

[34] F. W. Grosveld and K. P. Shepherd, "Active sound attenuation across a double wall structure," J. Aircr. **31**, 223–227 (1994).

[35] W. Yamada, T. Sueto, Y. Okudaira, and H. Hamada, "Active control of sound in the cavity of a double wall to increase the sound transmission loss," Proceedings of Inter-Noise 94, Yokohama, Japan, 1994, pp. 1311–1314.

[36] P. Sas, C. Bao, F. Augusztinovicz, and W. Desmet, "Active control of sound transmission through a double panel partition," J. Sound Vib. **180**, 609–625 (1995).

[37] C. Bao and J. Pan, "Experimental study of different approaches for active control of sound transmission through double walls," J. Acoust. Soc. Am. **102**, 1664–1670 (1997).

[38] J. Pan and C. Bao, "Analytical study of different approaches for active control of sound transmission through double walls," J. Acoust. Soc. Am. **103**, 1916–1922 (1998).

[39] This equation corrects Eq. (13) in Ref. 7. Equations (14) and (15) of that paper also have minor errors that may be corrected by solving for $\hat{u}_1$ and $\hat{u}_2$ under general conditions (as outlined in the present paper), then substituting the control voltage.

[40] The concept of minimizing volume velocity into the cavity leads to another possibility for localized error sensing. If the cavity is sufficiently small, it may be feasible to minimize acoustic pressure at a point within the cavity to produce the desired reduction of volume velocity and subsequent control of transmitting surface vibrations. However, the scheme is slightly precarious; it is possible to minimize acoustic pressure at a point in the cavity while the transmitting diaphragm is still vibrating. If small interstice vibrations or extraneous sources in the receiving space cause the transmitting diaphragm to vibrate, the acoustic pressure minimization would not necessarily eliminate sound transmission through the diaphragm and surround. In addition, if a pressure error sensor happened to be located at an uncontrolled pressure node in the cavity, the error signal might not be sufficiently observable and control complications could arise.

[41] W. Thompson, Jr., "The computation of self- and mutual-radiation impedances for annular and elliptical pistons using Bouwkamp's integral," J. Sound Vib. **17**, 221–233 (1971).

[42] S. P. Lipshitz, T. C. Scott, and B. Salvy, "On the acoustic impedance of baffled strip radiators," J. Audio Eng. Soc. **43**, 573–580 (1995).

[43] M. E. Johnson and S. J. Elliott, "Active control of sound radiation using volume velocity cancellation," J. Acoust. Soc. Am. **98**, 2174–2186 (1995).

[44] S. J. Elliott and C. C. Boucher, "Interaction between multiple feedforward active control systems," IEEE Trans. Speech Audio Process. **2**, 521–530 (1994).

An experimental investigation of two module configurations for use in active segmented partitions

Timothy W. Leishman
Acoustics Research Group, Department of Physics and Astronomy, Brigham Young University,
Eyring Science Center, Provo, Utah 84602

Jiri Tichy
Graduate Program in Acoustics, The Pennsylvania State University, Applied Science Building,
University Park, Pennsylvania 16802

Individual modules intended for active segmented partitions should be carefully analyzed before they are advanced as viable tools for active sound transmission control. In this paper we present experimental evaluations of two vibration-controlled modules: a single-composite-leaf (SCL) configuration and a double-composite-leaf (DCL) configuration. Experimental apparatuses and procedures are introduced to assess their normal-incidence transmission losses over a bandwidth from 40 to 1970 Hz. The average transmission loss of the passive SCL module is found to be 21 dB. If its transmitting diaphragm acceleration is minimized through active control, transmission loss increases somewhat at lower frequencies but decreases at higher frequencies, producing no increase in the average value. The average transmission loss of the passive DCL module is found to be 43 dB. After its transmitting diaphragm acceleration is actively minimized, the transmission loss increases substantially at all frequencies (especially at low frequencies), to produce an average value of 77 dB. Thus, while both configurations have the same underlying control objective, the DCL module yields a 56 dB average improvement over the SCL module through simple configurational changes. An alternative control scheme is investigated that minimizes acoustic pressure in the DCL module cavity, but it is found to be less effective. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992747]

I. INTRODUCTION

An active segmented partition (ASP) consists of an array of interconnected modules or elements with active and passive components that are configured and controlled to reduce sound transmission. Proper design, control, and evaluation of individual modules are crucial if ASPs are to provide effective control of sound transmission over useful bandwidths. In the past, research of ASPs has been obscured by more extensive research in other areas of active sound transmission control (ASTC).[1] It has nevertheless been pursued by a few researchers, who have seen its potential to enhance sound isolation capabilities of existing segmented partition structures, reduce partition masses, simplify active control requirements, and enhance ASTC in other ways.

Acousticians have long recognized the need for more efficient control of sound transmission through lightweight partitions, particularly at low frequencies. Some have anticipated that the active control of partitions would help fulfill this need because of its typical low-frequency benefits. However, difficulties encountered in the control of distributed structural and fluid media have imposed serious limitations on both the performance and practicality of many ASTC systems. Several criteria have been proposed to address these limitations, suggesting that properly implemented ASPs could provide needed ASTC improvements.[2,3]

Most researchers advocating ASPs in the past have done so to propose a means of controlling sound transmission through fuselage structures into aircraft cabins.[4–10] This concept is appealing because interior fuselage trim panels already exist as segmented partition elements. Others have suggested ASPs to control sound transmission through lightweight machinery enclosure walls[11,12] (a concept reminiscent of earlier passive techniques[13]). The authors of this paper have explored many different ASP configurations for general applications and ASTC improvements.[2,3,14–17]

While past research efforts have provided useful contributions to ASPs and their advancement, they have often lacked in-depth analyses and comparisons of module configurations, control methods, and associated ASTC performances. Recent theoretical work has filled some of these gaps through the exploration of several distinct module configurations.[2,3,17] However, a presentation of related experimental results is needed to validate theoretical predictions, to further compare the configurations, and to establish those capable of producing very high transmission loss using simple control strategies.

Sound isolation capabilities of experimental ASP implementations have often remained ambiguous to readers of the literature. Apparent shortcomings related to their performance and consistency have limited their acceptance or utility for many applications. However, uncertainty surrounding their capabilities has arisen in part because irregular measurement techniques are often used to quantify sound isolation of active partitions. While it is true that their measure-

ments are somewhat unique and must account for special conditions produced by active control,[18] they should reasonably conform to recognized norms, providing a meaningful evaluation of actual transmission loss capabilities. Experimental methods have been developed as part of this work to address these needs.

The research reported in this paper was undertaken to increase general understanding of ASPs and to provide greater insight into individual module configurations, their performance capabilities, and their limitations. It was also intended to substantiate theoretical predictions for two specific ASP module configurations. The first incorporated a single electromechanically actuated diaphragm with a surrounding airtight suspension (compare configuration 2 in Ref. 17). Theoretical work predicted that this single-composite-leaf (SCL) configuration would show significant transmission loss deficiencies. The second involved an electroacoustic actuator driving an isolated cavity and thereby actuating a passive transmitting diaphragm with a surrounding airtight suspension (compare configuration 3 in Ref. 17). Theoretical work predicted that this double-composite-leaf (DCL) configuration would perform much better than the first, achieving very high transmission loss over a broad bandwidth while satisfying several other design and performance criteria.[2,3,17]

Several experimental tools were developed to evaluate the modules and their control schemes. The underlying control objective for each was to minimize sound transmission through global reduction of transmitting surface vibrations, using only simple actuation and error-sensing techniques.

In the following sections we describe the experimental methods used in the research effort. We then present results and analyses of measured normal-incidence transmission losses for the two modules in their passive and active states. The results are compared to theoretical predictions and transmission losses of six passive benchmark partitions for further experimental validation and insight.

II. EXPERIMENTAL METHODS

A. Transmission loss measurement system

Normal-incidence acoustic properties of ASP modules may be readily determined using plane-wave tubes and the two-microphone transfer function technique developed by Chung and Blaser.[20,21] The common uses of these tools for the measurement of absorptive materials and acoustic filters naturally extend to the measurement of an electromechanoacoustic device separating a source tube and receiving tube.[3,14,15,22] The approach may be used to decompose the upstream and downstream fields adjacent to an ASP module into normally incident, reflected, and transmitted components, producing reliable transmission loss measurements. The possible residual reflection from an anechoic receiving tube termination suggests a need to decompose the downstream field into traveling-wave components using downstream microphones.

Figure 1 depicts the plane-wave tube system developed for the measurement of ASP modules and passive benchmark partitions in this work. It is shown with an arbitrary partition
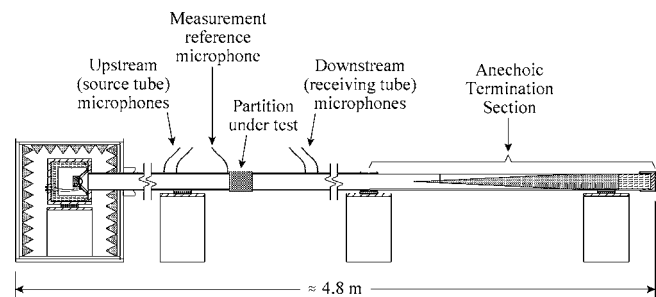
FIG. 1. Experimental system for measuring normal-incidence transmission loss of an individual partition module (ASP module or passive benchmark structure). The arrangement is shown with an arbitrary partition positioned between the source tube and receiving tube.

positioned between the source tube and receiving tube. Time-averaged sound power incident upon a partition $\langle W_i \rangle_t$ and transmitted through the partition $\langle W_t \rangle_t$ are detected using the upstream and downstream microphone pairs. The transmission loss then follows from the expression

$$\text{TL} = 10 \log \left(\frac{\langle W_i \rangle_t}{\langle W_t \rangle_t} \right). \tag{1}$$

Both tubes were constructed of heavy plastic with 10.1 cm inside diameters.[23] The source tube had an overall length of approximately 1 m. The receiving tube consisted of a measurement section and an anechoic termination section. The termination was designed following empirical concepts[24] and validated[25] using the two-microphone transfer function technique.[26] It was a pyramidal wedge cut from a block of 25.6 kg/m^3 open-cell polyurethane foam rubber, tapering over a 91 cm span from its apex to its 10.1 cm diameter circular base. Its tip was supported by a thin wire set a small distance from the apex to prevent it from sagging. The circular base was a 25 cm thick continuous extension of the wedge, backed by an 18 cm thick air cavity filled with loose glass-wool insulation and capped with a heavy cast-iron plug. In overall length, the anechoic termination was approximately 134 cm. Its measured cutoff frequency (above which its normal-incidence pressure-amplitude reflection coefficient modulus was consistently $|R| \leq 0.10$) was approximately 135 Hz. However, its absorption coefficient α exceeded 0.5 down to 40 Hz. The usable length of the receiving tube measurement section, from a given partition face to the anechoic wedge tip, was approximately 184 cm.

The source tube excitation loudspeaker was driven by an amplified swept sinusoid, stepped at 1 Hz increments below the cutoff frequency of the first tube cross mode (1990 Hz). Many partitions introduced boundary conditions that resulted in significant axial source tube resonances. However, acoustic pressure was measured by a measurement reference microphone at the source-side face of each partition to ensure that its level was maintained at a nearly constant value over frequency. This was accomplished through automatic level control of the excitation signal. The process ensured sufficiently linear behavior of the source tube sound field and its transducers, and maintained an adequate transmission loss signal-to-noise ratio at all frequencies of interest. Nevertheless, because of the automatic control, the excitation loud-

FIG. 3. Single-composite-leaf (SCL) ASP module, shown with an accelerometer mounted to the transmitting diaphragm (aluminum honeycomb sandwich disk). Compare configuration 2 of Ref. 17.
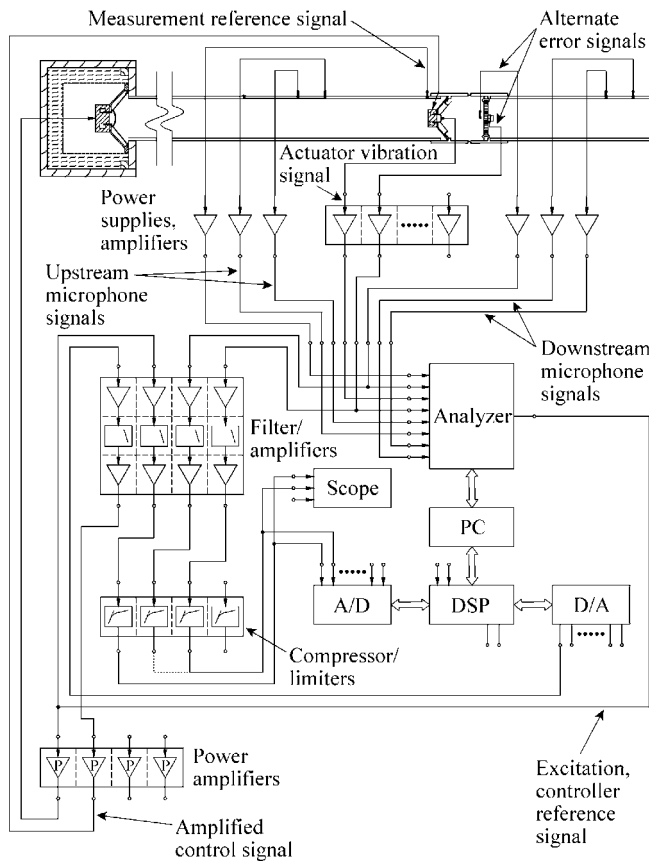
FIG. 2. Block diagram showing electronic equipment and interconnections of the measurement system. The interconnections shown correspond to the evaluation of the double-composite-leaf (DCL) ASP module discussed in Sec. III B.

speaker vibrated more vigorously at some frequencies (i.e., at source tube antiresonance frequencies) than at other frequencies (i.e., at source tube resonance frequencies). Structure-borne flanking transmission consequently became more conspicuous near antiresonance frequencies.

A block diagram of the measurement system, its transducers, and electronics is shown in Fig. 2. Rough placements of microphones and accelerometers are suggested in the figure. More specific placements are discussed below and shown in subsequent figures. Holes were drilled into the tops of the source and receiving tubes to accept microphones at key positions and multiple spacings. This enabled broadband decomposition of the fields within established error constraints.[26,27] The microphones were regularly calibrated to each other and to a standard reference microphone using switching[20,26] and uniform-pressure calibration techniques.

As seen from the figure, three microphones were actually used to probe the source tube. The microphone positioned closest to the partition (approximately 1 cm upstream) was the measurement reference microphone mentioned earlier. The other two microphones were a relatively calibrated pair used to determine the one-dimensional acoustic properties in the source tube. The rightmost microphone of the pair was fixed 30 cm (approximately 3 tube diameters) from the end of the tube. This distance provided a margin beyond the common single-diameter rule,[26] allowing a greater decay of strongly reflected higher-order modes and ensuring the plane
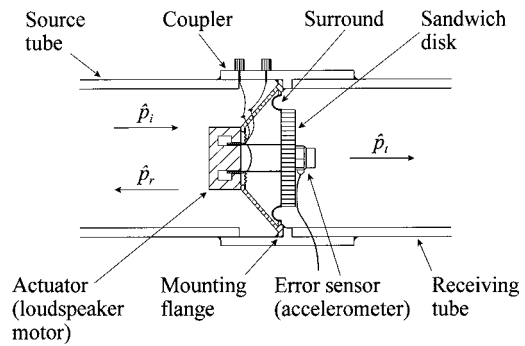
wave mode became more predominant at the microphone pair. The leftmost microphone of the pair was positioned at one of two upstream positions to provide either a 7.0 or 42.0 cm microphone spacing for valid measurements at higher or lower frequencies, respectively. A second relatively calibrated microphone pair was used in a mirror-image configuration on the opposite side of the partition to probe the receiving tube. A usable measurement bandwidth of approximately 40 to 1970 Hz was made possible through the combination of measurements made with the two spacings. A few passive properties of ASP modules were evaluated at frequencies below 40 Hz. Error conditions did not fall into the ideal range for these frequencies,[27] but they were considered reasonable for the purposes of this study. Other details of the measurement system, control system, and electronics are discussed in Ref. 3.

B. ASP modules

1. Single-composite-leaf (SCL) module

The first experimental ASP module was adapted from a moving-coil loudspeaker, as shown in Fig. 3. The loudspeaker diaphragm was replaced by an aluminum honeycomb sandwich disk with a 7 cm diameter and an 11 g total mass. Because its first bending mode frequency (for a free-edge condition) was determined to be well above the cutoff frequency of the first plane-wave tube cross mode, the disk was assumed to behave as a rigid body over the frequency range of interest. A small 1.3 g acrylic tube was glued between the existing voice-coil former and disk to enable its direct electromechanical actuation. The inner edge of the existing half-roll doped-cloth surround was glued to the back face of the disk around its periphery. The loudspeaker frame incorporated large openings, making it somewhat unobtrusive to incident waves.

The loudspeaker mounting flange was trimmed to fit snugly within a heavy PVC plastic tube coupler and firmly adhered to the inner coupler surfaces to produce an airtight seal between the source tube and receiving tube. The coupler was likewise joined to the source tube and receiving tube with airtight seals. An accelerometer was mounted to the center of the transmitting face of the disk as an error sensor. When the module was excited by normally incident waves, a control signal could be applied to the actuator terminals to
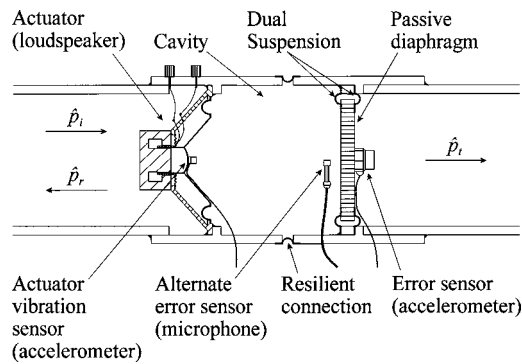
FIG. 4. Double-composite-leaf (DCL) ASP module, shown with accelerometers mounted to the hardened actuator diaphragm dust dome and to the passive transmitting diaphragm. A cavity microphone is also shown behind the center of the transmitting diaphragm. The module coupler housings are attached via a resilient airtight connection to reduce structure-borne flanking transmission from the source tube to the receiving tube. Compare configuration 3 of Ref. 17.

minimize the normal acceleration of the disk. In its uncontrolled state, the actuator terminals were simply open circuited. Electro-mechano-acoustic parameters for the module (defined in Ref. 17) were determined using common electric impedance testing techniques,[28] acoustical testing techniques,[22] direct measurements,[3] and published specifications.

2. Double-composite-leaf (DCL) module

The second experimental ASP module is depicted in Fig. 4. It was constructed using the same type of loudspeaker as in the first module, but with the original diaphragm assembly still intact. The diaphragm dust dome was coated with several layers of epoxy to provide a sufficiently rigid surface for accelerometer mounting. A miniature accelerometer was mounted to the dome to observe its vibrational behavior. In this configuration, the loudspeaker functioned as an electroacoustic actuator driving a small acoustic cavity (11.3 cm diameter and 9.1 cm nominal length) between its own diaphragm and a passive transmitting diaphragm. The latter consisted of a 19 g stiff aluminum honeycomb sandwich disk supported by a dual surrounding suspension. The suspension consisted of two half-roll butyl 50 rubber surrounds that were firmly glued to the peripheries of the front and back disk faces. An accelerometer was mounted to the center of the transmitting face as an error sensor. A small microphone was also positioned in the cavity to observe the enclosed field or to act as an alternate error sensor. Electro-mechano-acoustic parameters for the module (defined in Ref. 17) were again determined using a variety of methods.

As with the first module, the actuator and transmitting diaphragm assemblies were carefully fitted and adhered within tube coupler housings with airtight seals. The couplers were also joined to the source tube and receiving tube with airtight seals. However, because of complications arising from flanking transmission during very high transmission loss measurements, the two couplers were typically spaced by about 1 cm and joined with a resilient airtight connection. The connection reduced structure-borne flanking trans-

mission into the receiving tube, most of which resulted from mechanical source tube vibrations induced by the excitation loudspeaker or control actuator.

Once the module was excited by normally incident waves, a control signal was applied to the actuator terminals to minimize either normal acceleration of the transmitting diaphragm or acoustic pressure at a specified point within the cavity. In its uncontrolled state, the actuator terminals were open-circuited.

C. Passive benchmark partitions
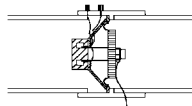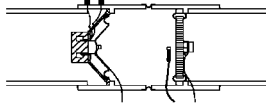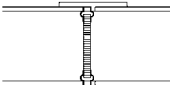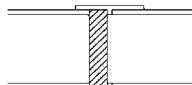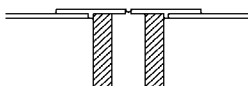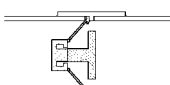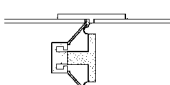
As a means of validating results provided by the measurement system, the active and passive transmission losses of the experimental ASP modules were compared with transmission losses of several passive benchmark partitions. Table I summarizes the designs and experimental purposes of the various active and passive structures.

Benchmark partition 1 was an open tube (no structure), expected to provide very little if any measurable transmission loss. Benchmark partition 2 was the isolated passive transmitting diaphragm assembly from the DCL ASP module. The transmission loss of this diaphragm was expected to behave like that of a single-leaf partition with effective mass, compliance, and resistance over an appreciable bandwidth.

Benchmark partition 3 incorporated a single 2.5 cm thick, 11.3 cm diameter solid steel plug spanning the entire cross section of a PVC plastic or steel tube coupler, sealed around its periphery to the inner surfaces of the coupler. The coupler was firmly connected and sealed to both the source and receiving tube. While the structure had no readily predictable transmission loss behavior, it was expected to provide a high transmission loss benchmark with significant structural flanking paths. Benchmark partition 4 (also shown in Fig. 5) incorporated a pair of sealed steel plugs—one in a PVC coupler and one in an adjacent steel coupler. As with the DCL ASP module, one coupler was mounted and sealed to the source tube and the other was mounted and sealed to the receiving tube. They were typically separated from each other by a 1 cm spacing and joined with a resilient airtight connection. This resulted in a plug spacing of approximately 4 cm. The arrangement provided a very high transmission loss benchmark. The resilient connection could also be removed so that the couplers could be firmly connected or separated by a large distance (e.g., 50 cm). Additional physical separation further reduced direct and flanking transmission, providing an extremely high transmission loss, near or exceeding the highest measurable transmission loss of the experimental system.

Benchmark partition 5 was an arrangement similar in basic form to the SCL ASP module (see Fig. 3). However, the transmitting diaphragm, surround, voice-coil assembly, and spider were removed from the actuator and replaced with a heavy steel plug (see Fig. 6). The diameter of the plug face was the same as the diameter of the transmitting aluminum honeycomb sandwich disk. The diameter of the plug shaft was approximately the same as the diameter of the voice-coil former and extension. The base of the shaft was firmly affixed to the actuator magnet structure. Since the resilient sur-

TABLE I. Listing of measured ASP modules and passive benchmark partitions.

Name	Description	Purpose	Diagram
SCL ASP module	Single-composite-leaf ASP module	TL evaluation of configuration 2 from Ref. 17.	
DCL ASP module	Double-composite-leaf ASP module, resilient coupler connection	TL evaluation of configuration 3 from Ref. 17.	
Benchmark partition 1	Open tube	Validation of minimal TL	
Benchmark partition 2	Single-leaf passive transmitting diaphragm	Validation of predictable single-leaf partition TL	
Benchmark partition 3	Single steel plug	High TL with appreciable structure-borne flanking	
Benchmark partition 4	Dual steel plugs, resilient coupler connection	Very high TL with minimal structure-borne flanking	
Benchmark partition 5	Modified SCL module, steel plug, no surround	TL evaluation of magnet and diaphragm constriction	
Benchmark partition 6	Modified SCL module, steel plug, surround	TL evaluation of isolated surround	

round was removed, a direct but constricted sound transmission path was introduced around the perimeter of the plug.

Benchmark partition 6 was nearly identical to benchmark partition 5, except that the resilient surround was once again mounted. As shown in Fig. 6, its inner edge was glued to the steel plug and its outer edge was glued to the actuator frame. This mounting again established the airtight seal between the source and receiving tubes and provided an approximate passive means of evaluating sound transmission through the surround alone.

D. Measurement capabilities and limitations

Measurements and analyses of the first four passive benchmark partitions were conducted to verify that the mea-

surement system functioned as expected and to identify some of its capabilities and limitations. The open tube (benchmark partition 1) showed little if any transmission loss (never exceeding 0.4 dB) over the measurement bandwidth. The isolated passive transmitting diaphragm assembly (benchmark partition 2) closely followed its theoretically predicted behavior as a single-leaf partition. The single and dual steel plug partitions (benchmark partitions 3 and 4) demonstrated the high transmission loss measurement capabilities of the system, but also exposed certain limitations. Maximum reliable transmission loss values were limited by flanking transmission, electronic cross-talk, ambient acoustic noise, electronic noise, etc. Benchmark partition 3 yielded a
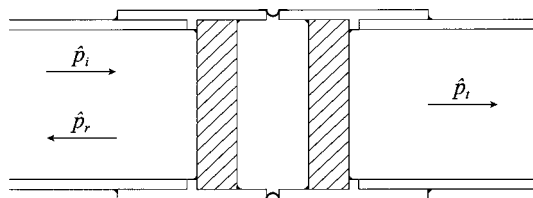
FIG. 5. Passive benchmark partition 4: a pair of spaced 2.5 cm thick solid steel plugs spanning the entire cross sections of the tube couplers to which they are sealed. The couplers are shown as being attached to each other by a resilient airtight connection to reduce structure-borne flanking transmission from the source tube to the receiving tube.
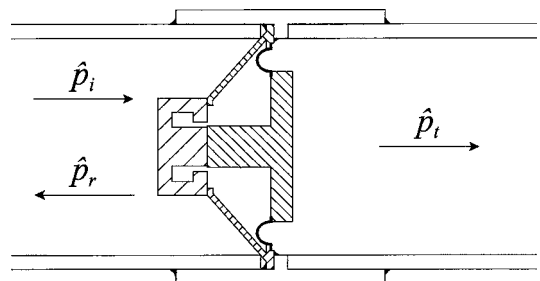
FIG. 6. Passive benchmark partition 6: a modified SCL module with a steel plug replacing the transmitting diaphragm assembly. The surround is intact.
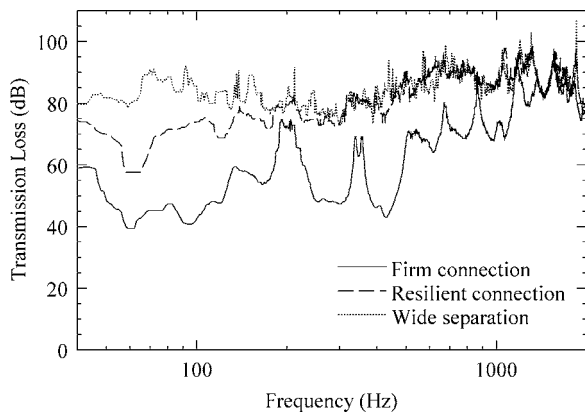
FIG. 7. Measured transmission losses for passive benchmark partition 4 (see Fig. 5). Measurement results are shown for three distinct coupler interconnection conditions.

transmission loss ranging from approximately 50 dB at the lowest measurement frequencies to approximately 70 dB at the highest frequencies, with an average value of 60 dB.[29] However, the transmission loss curve had several significant undulations over frequency that corresponded to decreases and increases in structural flanking transmission near the axial source tube resonance and antiresonance frequencies, respectively (see Sec. II A).

Many of the same flanking transmission effects were evident (especially below 700 Hz) for the dual steel plug partition (benchmark partition 4) when the couplers were rigidly connected. When the couplers were spaced with the resilient connection between them, these effects largely disappeared below 1350 Hz and the measurable transmission loss dramatically increased. The value ranged from approximately 70 dB at the lowest frequencies to approximately 90 dB above 500 Hz, with an average transmission loss of 86 dB. If the resilient connection was removed and the couplers were widely separated with no mechanical connection at all, a significant increase in the measured transmission loss occurred at only the lowest frequencies, below about 160 Hz. In this case, the transmission loss ranged from approximately 80 dB at the lowest frequencies to approximately 90 dB at the highest frequencies, producing only a slightly higher average value. Transmission loss plots for the three coupler conditions are shown in Fig. 7.

The benchmark partitions also increased the understanding of the experimental ASP modules. As discussed in Sec. III A 2, measurements of benchmark partitions 5 and 6 were particularly useful for increasing understanding of the SCL module. Other validations of the measured module properties are discussed in Ref. 3.

The transmission loss curves in Fig. 7 demonstrate a degree of random fluctuation at decibel values above about 80 dB. In such a high decibel range, measurement limitations could be caused by electronic cross-talk between measured signals or an insufficient electroacoustic signal-to-noise ratio. A series of measurements specifically demonstrated that electronic cross-talk could impact the associated downstream signals—at least near 200 Hz, 350 Hz, and in the band between 500 and 1000 Hz.[3] Electroacoustic noise in the downstream measurements included ambient noise and self-noise

of the microphones and their accompanying electronics. Because of the signal processing methods used in the measurements, the upstream incident intensity level was intended to exceed the transmission loss by an amount greater than the associated background intensity level at all frequencies of interest. For partitions with extremely high transmission loss, this could pose a significant problem. The incident intensity level had to be limited to prevent clipping of source-space microphone signals and other system nonlinearities. A level of approximately 124 dB (re 10^{-12} W/m^2) was chosen, producing a maximum transmission loss signal-to-noise ratio from roughly 80 dB at the lowest frequencies to 100 dB at the highest frequencies. The average value over the measurement bandwidth was approximately 95 dB.[3]

A comparison of the measured transmission loss of benchmark partition 4 (with resilient connection) to the maximum measurable values (limited by either electronic crosstalk or transmission loss signal-to-noise ratio) revealed that the partition pushed the dynamic limits of the measurement system—especially between 500 and 800 Hz. However, at most other frequencies, the system was capable of measuring its average 86 dB transmission loss with sufficient margin. Other measurement limitations and sources of measurement error are considered elsewhere.[3,22,26,27]

E. Measurement challenges posed by adaptive control

Additional experimental challenges needed to be surmounted to perform accurate transmission loss measurements of the ASP modules under adaptive control. Simultaneous (and sometimes conflicting) dynamic range requirements dictated by the measurement system and the adaptive controller had to be consistently satisfied. The adaptive controller also had to remain stable and maintain an adequate rate of convergence as it tracked the swept-sinusoidal stimulus.

1. Simultaneous dynamic range requirements

As seen in Fig. 2, several components of the measurement system shared signals with the adaptive controller. The challenging simultaneous dynamic range requirements of the components became further complicated by the fact that the measurement system utilized multiple signal channels and a large bandwidth. To maximize usable dynamic ranges, it was necessary to optimize signal path gain structures. The analyzer automatically adjusted every A/D input range with internal amplification circuits; they were checked, adjusted, and recorded with each swept-sine step. However, the adaptive controller had no automatic gain control for its A/D inputs. The associated signals had to be independently conditioned and optimized to ensure that the controller dynamic range capabilities could be met.

As indicated above, the ambient acoustic and electronic noise present in the receiving tube measurements required the source tube to be sufficiently but not excessively driven to bring transmitted acoustic energy above the noise floor. This was a primary reason why the stimulus signal level was adjusted over frequency to maintain a nearly constant sound pressure level at the source-side face of the partitions. The

process necessitated frequency-dependent compensation for the excitation loudspeaker sensitivity and strong source tube resonances. However, the same signal driving the excitation loudspeaker power amplifier also drove the adaptive controller reference input. Thus, because of the extreme amplitude adjustments needed to satisfy the constant sound pressure level requirement, the varying signal level essentially pushed the adaptive controller input to its dynamic limits. Had the signal at the input been maintained at a consistent optimal level, the dynamic range of transmission loss measurements would have suffered. Thus, the ideal requirements of the measurement system and adaptive controller were essentially at odds with one another.

The solution chosen to remedy this problem involved both the use of the measurement reference microphone discussed earlier and the multichannel dbx 1046 compressor/limiter, which conditioned the controller reference input signal (see Fig. 2). The latter was set with an infinite compression ratio and a threshold that maintained a nearly constant controller reference input level over the entire measurement bandwidth. Another channel of the compressor/limiter was used to explore dynamic range control of the ASP module error signals.[30]

Because of the automatic stimulus signal level adjustments, the excitation loudspeaker required protection from being overdriven at some frequencies. This was accomplished by specifying a fixed maximum stimulus signal level in the swept-sine control. The output signal of the reference microphone was also maintained at a maximum level that ensured that the sound pressure at all other source-space microphones would remain high, but without overdriving them or their accompanying electronic circuitry.

2. Tracking of the swept-sine stimulus

Another experimental challenge involved the ability of the adaptive controller to track rapid changes in the stepped-sinusoidal excitation of the measurement system. With measurements spanning nearly 2000 Hz with 1 Hz step intervals, overall measurement durations could be quite long. It was therefore desirable to adjust the sweep rate to be as fast as possible and the settling time after each step to be as short as possible. However, the adaptive controller had to be able to track the sweep, adapting rapidly and stably within the settling times.

As the stimulus frequency approached axial source tube resonance frequencies, the adaptive controller typically took much longer to adapt after each 1 Hz frequency increment. This was likely due to dramatic changes in system amplitudes and phases near these frequencies. On occasion, it was necessary to increase settling times or pause measurements and increase adaptive controller convergence parameters to shorten convergence times with each step.

III. RESULTS

The experimental ASP modules were tested under the various conditions summarized in Table II. Three measure-

TABLE II. Measurement conditions for the experimental ASP modules.

ASP Module	Mode	Error Signal	Error Signal Compression
SCL	Passive	N/A	N/A
	Active	Transmitting diaphragm acceleration	No
		Transmitting diaphragm acceleration	Yes
DCL	Passive	N/A	N/A
	Active	Transmitting diaphragm acceleration	Yes
		Cavity pressure, behind diaphragm center	Yes
		Cavity pressure, behind diaphragm periphery	Yes

ments were conducted for the SCL module and four for the DCL module. The following sections provide additional details of these measurements and their results.

A. Single-composite-leaf (SCL) module

The SCL module depicted in Fig. 3 was evaluated under two basic conditions. First, its behavior was measured in its passive mode, with the actuator terminals open circuited. Subsequently, its normal transmitting diaphragm acceleration was minimized through adaptive control. The acceleration control was first achieved without error signal compression, then with error signal compression. The following paragraphs report measurements and observations for the module under the various conditions. They also compare evaluations of passive benchmark partitions 5 and 6 introduced in Sec. II C.

1. Passive mode

Figure 8 shows the measured normal-incidence transmission loss for the passive module plotted against the transmission loss for the active module (discussed below). Passive and active theoretical predictions are also plotted using analytical results[17] and module parameters. The minimum
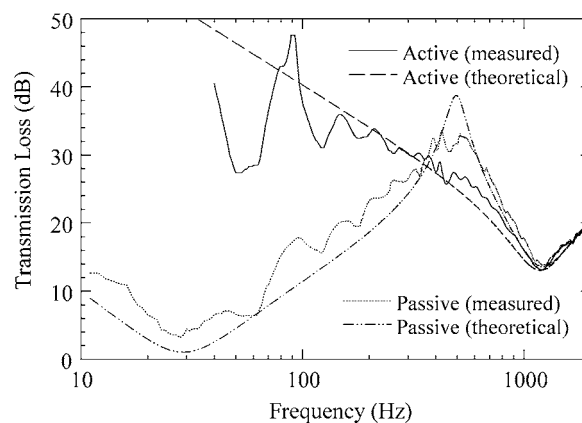
FIG. 8. Measured transmission losses for the SCL ASP module in its passive mode (actuator terminals open circuited) and in its active mode (transmitting diaphragm acceleration minimized). The measured curves are plotted against theoretical predictions for the two modes.

value of the passive transmission loss curve corresponds to the diaphragm assembly resonance frequency at approximately 28 Hz. Below this frequency, the assembly becomes stiffness controlled. Between about 50 and 500 Hz, it appears to behave as a mass-controlled single-leaf partition; its value consistently rises at approximately 6 dB per octave.[31] One might expect the device to remain mass controlled at higher frequencies, but this behavior clearly breaks down above 500 Hz. The transmission loss drops from a maximum of 32 dB at 500 Hz to 14 dB at about 1220 Hz. It then rises again gradually, but to a value no greater than 20 dB. The dip centered at 1220 Hz corresponds to the primary resonance of the damped surround. When averaged over the entire measurement bandwidth, the transmission loss of the passive module is approximately 21 dB.

2. Transmitting diaphragm acceleration control

The normal acceleration of the module diaphragm was effectively minimized through direct electromechanical actuation. As indicated earlier, this actuation could be adaptively controlled either with or without compression of the acceleration error signal. Despite improvements in error signal reduction provided by compression,[32] transmission losses produced by the module were essentially identical in either case.

The actively controlled transmission loss shown in Fig. 8 generally agrees with the corresponding theoretical prediction, except at the lowest frequencies. The measured curve decreases with undulation from moderate values (e.g., 35 dB) at low frequencies to a minimum value of 13 dB at 1220 Hz. The curve then rises by approximately 1 dB per 100 Hz increase to a value of about 20 dB at 1970 Hz. The average transmission loss over the entire measurement bandwidth is again 21 dB. Thus, on average, the actively controlled module offers no benefit over the passively controlled module. However, below 350 Hz, it does offer useful improvement. Between 40 and 350 Hz, the average transmission loss is 33 dB, compared to the 24 dB average of the uncontrolled module. At the lowest frequencies, active control produced a 20 to 30 dB improvement over the passive transmission loss results.

As predicted theoretically, the uncontrolled module provided better transmission loss than the controlled module between about 350 and 1300 Hz. Above 1300 Hz, the controlled and uncontrolled states produced essentially identical results. Theoretical analysis suggested that these high-frequency active performance limitations are caused by uncontrolled surround vibrations.[17] Passive benchmark partitions 5 and 6 were evaluated to validate this assertion. Benchmark partition 5 provided means of evaluating the constriction effects of the module magnet structure, frame, and rigid transmitting diaphragm core. Benchmark partition 6 was identical to benchmark partition 5, except that it added the flexible airtight surround (see Fig. 6). Measured transmission losses of the two benchmark partitions are compared in Fig. 9.

Without the surround, the transmission loss of benchmark partition 5 was very small. Over the measurement bandwidth, its average value was only 1 dB. It produced a
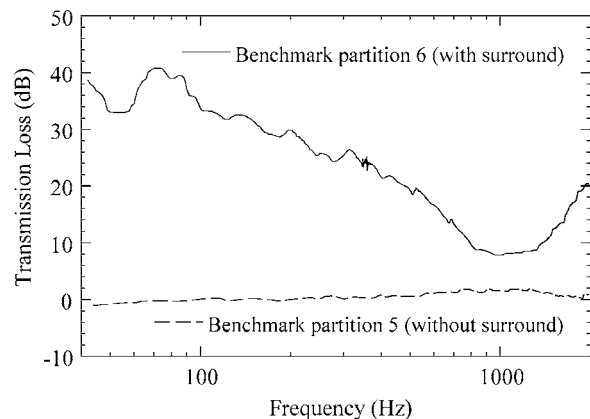
FIG. 9. Measured transmission losses for passive benchmark partitions 5 and 6 (see Fig. 6). Both are modified SCL modules with a steel plug replacing the transmitting diaphragm assembly. Benchmark partition 5 is configured without a surround while benchmark partition 6 is configured with a surround.

transmission loss plateau of just under 2 dB between 750 and 1275 Hz. At higher and lower frequencies, the transmission loss tapered to values closer to those of the open tube, with an average transmission loss of less than 0.2 dB. The transmission loss of benchmark partition 6 was quite similar to that of the acceleration-controlled ASP module shown in Fig. 8. These measurements accordingly confirm that the module surround was an important partition element with a resonant transmission loss minimum near 1200 Hz.

B. Double-composite-leaf (DCL) module

The DCL module depicted in Fig. 4 was evaluated under three basic conditions. It was first measured in its passive mode, with its actuator terminals open circuited. It was next measured in its first active mode, with its normal transmitting diaphragm acceleration minimized through acoustic actuation. It was subsequently measured in its second active mode, with its enclosed cavity pressure minimized at a point through acoustic actuation. For the second active mode, the error signal was first produced by a microphone located on the chamber axis, centered about 1 cm behind the transmitting diaphragm. The microphone was then moved to a position about 1 cm from the edge of the chamber, approximately 1 cm behind the transmitting diaphragm surround. Because of the success of error signal compression with the SCL module, the compressor/limiter was again used to process the module error signals (see Fig. 2). In the following sections we provide evaluations of the module under the various conditions.

1. Passive mode

The measured normal-incidence transmission loss for the passive DCL module is shown in Fig. 10. A theoretical curve based on analytical results[17] and module parameters is also shown for comparison. The dip in the curve to its minimum value near 35 Hz is well predicted by the theoretical curve and is caused by the impedances of the actuator and transmitting diaphragm assemblies. Transmission loss is characteristically poor in the surrounding spectral region. Al-
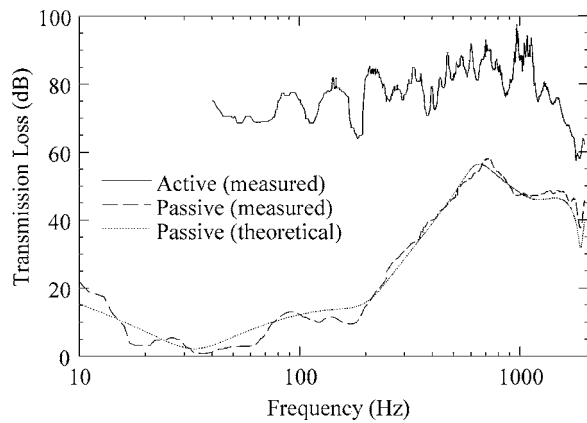
FIG. 10. Measured transmission losses for the DCL ASP module in its passive mode (actuator terminals open circuited) and in its first active mode (transmitting diaphragm acceleration minimized). The measured passive curve is plotted against a theoretical prediction.

FIG. 11. Measured acceleration error signal attenuation produced by a single-channel filtered-x controller with the DCL ASP module. The result is plotted against an attenuation curve for the cavity pressure immediately behind the center of the transmitting diaphragm under acceleration control.

though the average transmission loss over the entire measurement bandwidth is 43 dB, it is only 10 dB between 40 and 200 Hz. These low-frequency limitations are typical of passive partitions.

The dip near 200 Hz in the measured curve corresponds approximately to the mass-air-mass resonance frequency of the passive double-leaf partition.[31] Immediately above this dip, the transmission loss increases at slightly more than 20 dB per octave, as anticipated for frequencies just above the resonance. As the frequency continues to increase, one might expect the transmission loss to continue rising at a rate of about 18 dB per octave. However, the trend stops at approximately 720 Hz, at a maximum value of 58 dB. The reduction of the transmission loss above this frequency (following primarily from the presence of the actuator diaphragm surround) suggests a limitation to the idealized double-leaf partition behavior that might be ascribed to the passive module. From 720 to 950 Hz, transmission loss gradually recedes to approximately 47 dB, where it temporarily levels off. Near 1890 Hz, a pronounced dip to a value of 38 dB corresponds to the first axial resonance frequency of the enclosed cavity.

2. Transmitting diaphragm acceleration control

Once the normal acceleration of the transmitting diaphragm was minimized via active control, the transmission loss of the module dramatically increased over the entire measurement bandwidth (see Fig. 10). It increased by an average of 34 dB and by more than 70 dB at some lower frequencies. The transmission loss was not infinite, as predicted by limited theoretical modeling,[17] but it was substantial. Its general value gradually increased from about 70 dB near 40 Hz to over 90 dB near 725 Hz. Above 725 Hz, it decreased until it dropped to roughly 60 dB at the highest measurement frequencies. The average over the entire measurement bandwidth was approximately 77 dB. From 40 to 200 Hz, the average was 73 dB—a 63 dB improvement over the corresponding average under passive conditions. Between 40 and 1125 Hz, the average transmission loss was 82 dB.
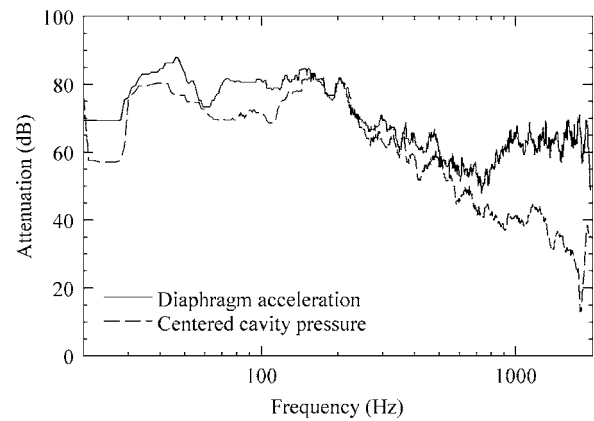
As shown in Fig. 11, the active attenuation of the normal acceleration error signal ranged from a low of about 50 dB at some frequencies to more than 80 dB at several frequencies below 200 Hz. (Such high values for the 12-bit controller I/O system were made possible in part by error signal compression.) The attenuation notably decreased between the lowest frequencies and 750 Hz. Above 750 Hz, it again increased until it leveled off at roughly 64 dB. The average reduction over the entire measurement bandwidth was 64 dB.

The attenuation of the cavity pressure directly behind the center of the transmitting diaphragm is also shown in Fig. 11 as a result of the acceleration control scheme. It clearly follows the acceleration attenuation up to about 750 Hz, but continues its decline at higher frequencies, dropping well below the acceleration attenuation curve. It eventually dips to a conspicuous minimum of 13 dB centered near 1800 Hz. The center frequency of this dip coincides closely with the cutoff frequency of the first cavity cross mode. At frequencies above the dip, the cavity pressure attenuation again rises to approximately 35 dB. On average, cavity pressure attenuation over the measurement bandwidth is 45 dB, a full 19 dB less than the corresponding average acceleration attenuation.

Another noteworthy characteristic of the controlled module involves the normal acceleration at the center of the *actuator* diaphragm. Under controlled conditions, its mean-square value was found to decrease from its passive mean-square value below 800 Hz. However, above 800 Hz, it actually *increased* in order to minimize the normal transmitting diaphragm acceleration. The greatest increase in its mean-square acceleration occurred at approximately 1850 Hz. Theoretical work has indicated that the transmitting diaphragm acceleration control is accomplished by minimizing the source-side volume velocity into the cavity.[2,3,17] Although the rigid body motion of the actuator diaphragm could not be guaranteed over the entire measurement bandwidth, the fact that the relative central diaphragm acceleration increased with increasing frequency suggests that the diaphragm had to counteract at least one other transmission path into the cavity. A principal candidate for such a path was the resilient diaphragm surround.
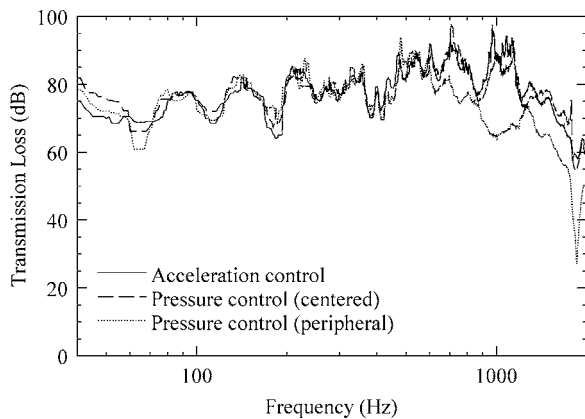
FIG. 12. Measured transmission losses for the DCL ASP module in three active modes: (1) transmitting diaphragm acceleration control, (2) cavity pressure control with the error microphone centered behind the transmitting diaphragm, and (3) cavity pressure control with the error microphone directly behind the transmitting diaphragm surround.

FIG. 13. A comparison of measured transmission losses for the SCL and DCL ASP modules in their passive and acceleration-controlled active modes.

3. Cavity acoustic pressure control

a. Central cavity position Figure 12 shows that when acoustic pressure was controlled at the central position behind the transmitting diaphragm, the module transmission loss characteristics were very similar to those attained under transmitting diaphragm acceleration control. The average transmission loss over the measurement bandwidth was again 77 dB.

The error signal reduction for this scheme was also similar to that produced by the acceleration control scheme over most of the measurement bandwidth. The average attenuation was again 64 dB. This stands in contrast to the 45 dB average pressure reduction realized under acceleration control at the same cavity position. Over a small bandwidth centered at approximately 180 Hz (near the passive mass-air-mass resonance frequency of the module) there was a significant difference between the error signal attenuation of the two control schemes. Acoustic pressure reduction under cavity pressure control was approximately 12 dB less than acceleration reduction under transmitting diaphragm acceleration control. Acoustic pressure was even reduced more through acceleration control at these frequencies than through cavity pressure control.

As one might anticipate, normal transmitting diaphragm acceleration was substantially reduced under the cavity pressure control scheme because of an attendant decrease in the source-side volume velocity into the cavity. This reduction was nearly identical to the cavity pressure reduction under transmitting diaphragm acceleration control over most of the measurement bandwidth (see Fig. 11). Its average attenuation of 43 dB was about 2 dB lower than the cavity pressure attenuation under acceleration control. The most pronounced difference between the two attenuation curves was again concentrated in the small spectral region centered near the passive mass-air-mass resonance frequency. Relative to cavity pressure reduction under acceleration control, acceleration reduction under cavity pressure control suffered a 13 dB dip in this region. Thus, although the two control schemes demonstrated similar behaviors, the acceleration control scheme was slightly more effective.
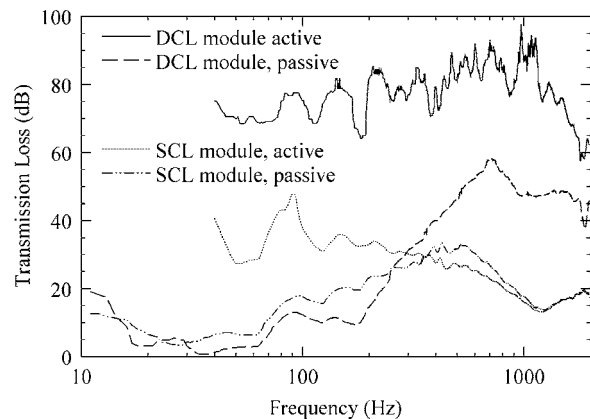
b. Peripheral cavity position Figure 12 also shows that the controlled transmission loss of the module suffered more significantly after the error microphone was moved from behind the center of the transmitting diaphragm to a peripheral position behind its surround. The transmission loss followed that generated under acceleration control and that generated under centered cavity pressure control up to about 600 Hz. However, above 600 Hz, its relative transmission loss significantly decreased, dropping from roughly 85 dB at 600 Hz to about 50 dB at 1970 Hz. The average value over the entire measurement bandwidth was 68 dB—a full 9 dB lower than that achieved under the other two schemes. The approach also produced a more pronounced dip to 27 dB near 1850 Hz. Interestingly, the transmission loss was even 12 dB lower at that frequency than that produced by the module in its passive state.

Error signal attenuation for this latter scheme once again resembled that of the acceleration control scheme, except near 180 Hz. At that frequency, its relative attenuation dropped by approximately 15 dB. Average attenuation over the measurement bandwidth was 63 dB. The attenuation of normal transmitting diaphragm acceleration was similar to that achieved by the centered cavity pressure control scheme up to about 400 Hz. However, above 400 Hz, it dropped until it became roughly 20 dB less than that achieved by the former scheme over the remaining measurement bandwidth. The attenuation dip centered at 1800 Hz (see Fig. 11) also shifted up slightly to a center frequency of 1850 Hz. The average diaphragm acceleration reduction over the entire measurement bandwidth was 29 dB—a full 14 dB less than that achieved when pressure was minimized at the central cavity position.

IV. DISCUSSION

Figure 13 compares the transmission loss curves for the two ASP modules in their passive and active acceleration-controlled states. While both modules utilized the same basic control objective, the DCL module yielded an average transmission loss improvement of 56 dB over the SCL module.

This significant improvement followed from a simple configurational difference that produced very high transmission loss over a broad frequency range.

Experimental results indicated that both normal transmitting diaphragm acceleration control and centered cavity pressure control of the DCL module yielded similar transmission losses. Both schemes performed best at frequencies well below the cutoff frequency of the first cavity cross mode and the first cavity axial resonance frequency. Assuming the transmitting diaphragm behaved as a translational piston, one would expect diaphragm acceleration control to be less sensitive to error sensor location than cavity pressure control. This assertion is partly supported by the degraded experimental results for cavity pressure control when the error microphone was displaced to the periphery of the cavity. Transmitting diaphragm acceleration control also provided slightly superior error signal reduction than either cavity pressure control scheme.

Many of the results described for the cavity pressure control scheme give evidence of acoustic wave effects increasing in the cavity with increasing frequency. Since these effects make the scheme less consistent at higher frequencies, error microphone placement must be optimized to reduce their impact. Another detrimental aspect of the scheme appears at lower frequencies near the mass-air-mass resonance frequency of the module. As indicated in Ref. 17, cavity pressure control is slightly precarious because it is possible to minimize acoustic pressure at a point in the cavity while the transmitting diaphragm is still vibrating. If an error microphone is positioned near an uncontrolled cavity pressure node, control complications may arise. Because of these limitations of the pressure control scheme, acceleration control might be considered the more effective strategy for the module.

The transmission loss of the SCL module was relatively low at most frequencies. As a result, limitations posed by flanking transmission, electronic cross-talk, and noise were not generally significant during its measurements. With the module rigidly attached to both the source tube and receiving tube, Figs. 7 and 8 suggest that flanking transmission may only have caused a slight effect on the measured transmission loss in a narrow frequency band near 90 Hz.

Because of the very high transmission loss produced by the DCL module, one might question whether the measurement system could accurately evaluate its full transmission loss capabilities. In Sec. II D we indicated that measurement constraints due to flanking transmission, electronic cross-talk, and noise posed significant limitations. To further consider the impact of these limitations, the transmission loss of the acceleration-controlled DCL module is plotted again in Fig. 14, against the transmission loss signal-to-noise ratio[3] and the transmission loss for benchmark partition 4 with its resilient coupler connection. An inspection of the curves reveals that the transmission loss of the DCL module was generally within the transmission loss signal-to-noise ratio, although it pushed these limits at some frequencies. The DCL module also pushed the limits suggested by the benchmark partition up to about 1150 Hz. However, above 1150 Hz, its transmission loss receded, creating a significant margin.
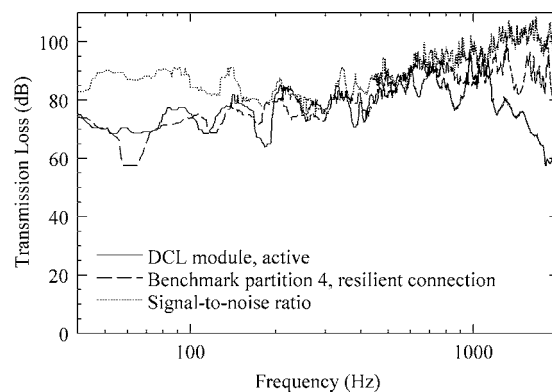
FIG. 14. A comparison of the measured transmission loss of the DCL ASP module under transmitting diaphragm acceleration control to (1) the measured transmission loss of benchmark partition 4 with a resilient coupler connection (see Fig. 5) and (2) the transmission loss signal-to-noise ratio.

Results from Sec. II D suggest that the benchmark partition transmission loss may have been slightly corrupted by flanking transmission below about 160 Hz (see Fig. 7). An inspection of Fig. 14 suggests that the low-frequency portions of the measured ASP module transmission loss may also have been corrupted. However, an additional comparison of Fig. 7 suggests that effects of flanking transmission could have been much worse had the two module sections not been mechanically isolated through their resilient connection.

Transmission loss for the benchmark partition could also have been slightly corrupted by electronic cross-talk between upstream and downstream signals—most predominantly in the spectral region between 500 and 1000 Hz. Once again, this possibility could have implications for the DCL module. Since identical adaptive controller convergences could not be guaranteed for consecutive measurements, upstream and downstream acoustic pressures were measured simultaneously. Otherwise, the effects of electronic cross-talk might have been further reduced.

As shown in Figs. 8 and 10, the measured and theoretically predicted transmission loss curves for the ASP modules have been found to agree within reason. This provides useful validation of the models presented in Ref. 17. Measured and theoretically predicted source-side reflection coefficients of the modules have also been found to agree.[3,17]

V. SUMMARY AND CONCLUSIONS

Experimental methods have been developed to assess the normal-incidence transmission losses of individual active segmented partition (ASP) modules over a bandwidth from 40 to 1970 Hz. The methods fulfill simultaneous dynamic range requirements of a measurement system and an adaptive controller while addressing adaptive controller stability and convergence. Steps were taken to evaluate and reduce flanking transmission, electronic cross-talk, ambient acoustic noise, and electronic noise—all of which limited the maximum reliably measured transmission loss of the system. The module measurements were validated by comparing their

measured transmission losses to theoretical predictions and measured transmission losses of passive benchmark partitions with predictable or extreme behaviors.

Two experimental ASP modules were created to substantiate theoretical predictions of their behaviors. The first was a single-composite-leaf (SCL) configuration with an electromechanically actuated diaphragm and a resilient surround. The second was a double-composite-leaf (DCL) configuration with an electroacoustically actuated cavity bounded by a passive transmitting diaphragm and a resilient surround. The underlying active control objective for each was to induce global reduction of its normal transmitting surface vibrations through simple actuation and error sensing.

A filtered-x adaptive controller successfully controlled both modules over the entire measurement bandwidth as they were excited by a swept-sinusoidal stimulus. Electronic signal compression was found to increase error signal attenuations, sometimes to exceptional values for the 12-bit controller I/O system. Nevertheless, actual transmission loss performances of the modules depended more significantly upon their physical configurations and control schemes.

Under active transmitting diaphragm acceleration control, the low-frequency transmission loss of the SCL module increased somewhat over its passive transmission loss, but the high-frequency transmission loss decreased, leaving the average value unchanged. The decline in high-frequency performance was apparently caused by uncontrolled surround vibrations. Under similar control, the transmission loss of the DCL module increased substantially at all frequencies. Increases at low frequencies were particularly noteworthy. Flanking transmission, electronic cross-talk, and noise may have had some influence on the measurement of its very high transmission loss values.

The adaptive controller was also used in an alternative control scheme to minimize acoustic pressure at discrete points within the DCL module cavity. Transmitting diaphragm acceleration control was found to be slightly more effective than cavity pressure control when the error microphone was centered directly behind the transmitting diaphragm. It was found to be much more effective than cavity pressure control when the microphone was moved to the periphery of the cavity. Inefficiencies of the cavity pressure control scheme stem from acoustic wave effects in the cavity at higher frequencies and other problems related to error signal observability. Transmitting diaphragm acceleration control and cavity pressure control both performed best at frequencies well below the cutoff frequency of the first cavity cross mode and the frequency of the first axial cavity resonance.

Under active control, the mean-square acceleration at the center of the DCL module actuator diaphragm actually increased from its passive value at higher frequencies. This increase was apparently required to counteract one or more transmission paths into the module cavity (other than the central part of the diaphragm itself) and to reduce the source-side volume velocity. The actuator surround was a likely candidate for an alternative path into the cavity.

While both modules used the same underlying control objective, the DCL module yielded an average transmission loss that was 56 dB higher than that of the SCL module. This substantial improvement was accomplished through a simple configurational change that required no more complicated actuation or error sensing. While both modules improved low-frequency transmission loss through active control, the DCL module consistently provided much higher transmission loss over the entire measurement bandwidth. It therefore proved to be the better candidate for ASP applications.

Additional work is required to assess normal-incidence transmission loss characteristics of other ASP module configurations and multiple modules mounted in ASP arrays. Oblique and random-incidence transmission loss measurements should also be conducted for these modules and arrays.

ACKNOWLEDGMENTS

[1] T. W. Leishman, "Research in the field of active sound transmission control," J. Acoust. Soc. Am. **114**, 2390(A) (2003).

[2] T. W. Leishman and J. Tichy, "A fundamental investigation of the active control of sound transmission through segmented partition elements," Proceedings of Noise-Con 97, University Park, Pennsylvania, USA, 1997, Vol. 2, pp. 137–148.

[3] T. W. Leishman, "Active control of sound transmission through partitions composed of discretely controlled modules," Ph.D. thesis, The Pennsylvania State University, University Park, Pennsylvania, 2000.

[4] D. R. Thomas, P. A. Nelson, and S. J. Elliott, "An experimental investigation into the active control of sound transmission through stifflight composite panels," Noise Control Eng. J. **41**, 273–279 (1993).

[5] D. R. Thomas, P. A. Nelson, R. J. Pinnington, and S. J. Elliott, "An analytical investigation of the active control of the transmission of sound through plates," J. Sound Vib. **181**, 515–539 (1995).

[6] S. L. Sharp, G. H. Koopmann, and W. Chen, "Transmission loss characteristics of an active trim panel," Proceedings of Noise-Con 97, University Park, Pennsylvania, 1997, Vol. 2, pp. 149–160.

[7] R. L. St. Pierre, Jr., G. H. Koopmann, and W. Chen, "Volume velocity control of sound transmission through composite panels," J. Sound Vib. **210**, 441–460 (1998).

[8] S. M. Hirsch, J. Q. Sun, and M. R. Jolly, "An analytical study of interior noise control using segmented panels," J. Sound Vib. **231**, 1007–1021 (2000).

[9] S. M. Hirsch, N. E. Meyer, M. A. Westervelt, P. King, F. J. Li, M. V. Petrova, and J. Q. Sun, "Experimental study of smart segmented trim panels for aircraft interior noise," J. Sound Vib. **231**, 1023–1027 (2000).

[10] G. P. Mathur, C. L. Chin, M. A. Simpson, and J. T. Lee, "Structural acoustic prediction and interior noise control technology," NASA/CR-2001-211247, 2001.

[11] J. E. Cole, III and M. C. Junger, "Active noise control for machinery enclosures," NSF Final Report U-1944-379F, 1991.

[12] J. E. Cole, K. F. Martini, and A. W. Stokes, "Active noise control for machinery enclosures," NSF Final Report U-2413-393, 1996.

[13] R. S. Jackson, "The performance of acoustic hoods at low frequencies," Acustica **12**, 139–152 (1962).

[14] T. W. Leishman and J. Tichy, "An experimental evaluation of individual partition segment configurations for the active control of sound transmission," J. Acoust. Soc. Am. **104**, 1776(A) (1998).

[15] T. W. Leishman and J. Tichy, "An experimental investigation of a novel active segmented partition for sound transmission control," J. Acoust. Soc. Am. **105**, 1156(A) (1999).

[16] T. W. Leishman, "Vibration-controlled modules for use in active segmented partitions," J. Acoust. Soc. Am. **114**, 2385(A) (2003).

[17]T. W. Leishman and J. Tichy, "A theoretical and numerical analysis of vibration-controlled modules for use in active segmented partitions," J. Acoust. Soc. Am. **118**, 1424-1438 (2005).

[18]One easily overlooked complication in the measurement of active partitions involves the impact of actuation upon both the receiving space and source space sound fields. When comparing passive and active control conditions for a partition, a change to either field has impact on the measured transmission loss. Another complication involves the interaction between active partitions, as secondary sources of sound, and primary sources of sound in the source space (Ref. 19). Other complications are discussed in this paper.

[19]T. W. Leishman and J. Tichy, "On the significance of reflection coefficients produced by active surfaces bounding one-dimensional sound fields," J. Acoust. Soc. Am. **113**, 1475–1482 (2003).

[20]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).

[21]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. II. Experiment," J. Acoust. Soc. Am. **68**, 914–921 (1980).

[22]B. E. Anderson and T. W. Leishman, "An acoustical measurement method for the derivation of loudspeaker parameters," AES 115th Convention, New York, NY, USA, 2003, preprint 5865, pp. 1–13.

[23]Although ASP modules are typically rectangular in cross section, circular tube and module geometries are investigated in this work with no loss of generality for the evaluation of individual modules.

[24]L. L. Beranek and H. P. Sleeper, Jr., "The design and construction of anechoic sound chambers," J. Acoust. Soc. Am. **18**, 140–150 (1946).

[25]"Standard test method for impedance and absorption of acoustical materials by the impedance tube method," American Society for Testing and Materials, ASTM Designation: C384–88 (1988).

[26]"Standard test method for impedance and absorption of acoustical materials using a tube, two microphones, and a digital frequency analysis system," American Society for Testing and Materials, ASTM Designation: E1050–90 (1990).

[27]H. Bodén and M. Åbom, "Influence of errors on the two-microphone method for measuring acoustic properties in ducts," J. Acoust. Soc. Am. **79**, 541–549 (1986).

[28]J. D'Appolito, *Testing Loudspeakers* (Audio Amateur Press, Peterborough, NH, 1998).

[29]Arithmetic decibel averages are used in this paper. Because they are based on fixed 1 Hz frequency increments, these averages are most easily visualized with transmission loss plots that use linear frequency scales.

[30]When used for error signal processing, the compressor/limiter was adjusted to send the adaptive controller consistently high error signal levels throughout the adaptation processes. With sinusoidal excitation and a compression ratio of approximately 8:1, it allowed adaptation to proceed without significant effect on convergence or stability. It also enabled the adaptive controller to significantly improve its error signal reductions (Ref. 3).

[31]F. J. Fahy, *Sound and Structural Vibration: Radiation, Transmission, and Response* (Academic Press, New York, 1985).

[32]Without the use of the compressor/limiter, the average acceleration reduction over the measurement bandwidth was approximately 63 dB. With its use, the value increased to approximately 66 dB—a 3 dB improvement. However, between 20 and 100 Hz, the compressor/limiter produced an average improvement of 10 dB, while between 200 and 600 Hz it produced an average improvement of 9 dB. Over the remainder of the bandwidth, the compressed and uncompressed error signals yielded similar reductions, with a few exceptions. Error signal compression might be considered for the enhancement of adaptive controllers with inadequate dynamic ranges, at least when used with sinusoidal stimuli.

Linear independence method for system identification/secondary path modeling for active control

Benjamin J. Kim[a)] and David C. Swanson

Applied Research Laboratory and Graduate Program in Acoustics, The Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804

A method for noninvasive system identification/secondary path modeling has been developed for single- and multi-channel filtered-*x* least-mean-square (LMS)-based active noise control (ANC). The problem of on-line secondary path modeling is recognized as one of linear dependence associated with an underdetermined system, a one-equation/two-unknown problem in which the highly correlated primary source and secondary source contributions to the error signal are not readily distinguishable. The method resolves this uniqueness issue by introducing a second equation with similar unknowns. The critical linear independence of the two equations, hence the proposed designation, is achieved with a single perturbation of the control filter output, thereby rendering the system solvable. This secondary path modeling strategy was implemented using an innovative real-time DSP control architecture and tested on a "transducerless" system devised to investigate behavior of ANC algorithms. Results of narrowband, broadband, and multi-channel tests reveal response estimates that are accurate in both magnitude and phase; bias due to primary noise and other secondary sources is notably absent in the obtained secondary path models. The rapidity with which the system is identified can also contribute to the stability and performance of filtered-*x* LMS-based controllers. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992727]

I. INTRODUCTION

The filtered-*x* least-mean-square (FXLMS) algorithm is easily the most widely used and studied adaptive algorithm in feedforward active noise control (ANC) due in no small measure to its relative simplicity, proven robustness, and ease of implementation.[1–6] An extension of the venerable LMS algorithm conceived by Widrow and Hoff in 1959,[7,8] the FXLMS algorithm was developed (by Morgan in 1980 and independently by Widrow *et al.* and Burgess in 1981)[9–11] to accommodate the electro-acoustic feedforward or secondary path inherent in ANC systems (see Fig. 1). Accommodation would entail filtering the reference or *x* signal with a response estimate or model of the composite secondary path prior to performing the cross correlation in the LMS recursion (hence the designation "filtered-*x*"). This prefiltering would compensate for the delay in the feedforward path by effectively time aligning the reference and error signals, allowing for a usable gradient estimate and thus rendering the LMS method practicable for ANC.

Obtaining the necessary secondary path response estimates requires that the FXLMS adaptive filter be partnered with a secondary path modeler, preferably one that is not only accurate in its measurement of the "plant" response but adaptive in order to track any changes that the path might undergo during system operation. Given that the stability and performance of the (otherwise robust) FXLMS algorithm depend almost entirely on the quality of the plant models, it is a rather uncompromising arrangement. The secondary path modeler must *lead* the adaptive control filter with response

estimates that sufficiently reflect the path at the time of the filter updates, else the controller can diverge. For a time-varying environment, this ideal modeler—one that is both accurate and adaptive—is preferred. To this end, secondary path modeling continues to be a focus of research in active control.

The level of accuracy that secondary path modelers should attain for stable FXLMS operation has been analyzed extensively.[9,12–14] With a tonal reference the FXLMS algorithm is stable provided the phase of the response estimate is within ±90° of that of the true plant response at the disturbance frequency. Numerical results suggest that errors of ±40° hardly affect the convergence of the algorithm;[9] beyond ±40°, however, convergence slows with adaptation actually "stalling" as phase errors approach ±90°. The convergence coefficient, or step size, can be reduced to help with stability but doing so can further slow adaptation and limit convergence. Errors in the magnitude of the response estimate, considered more forgiving than its phase counterpart, can also cause instability but can be rectified by an appropriate adjustment of the step size (in proportion to the inverse of the error).[5] This rescaling of step sizes based on error, however, is deemed unfeasible as it requires knowledge of the true plant response which if known would certainly be used in lieu of the estimated response. Nevertheless, the findings establish that secondary path models need only be sufficiently accurate—not necessarily exact—for a FXLMS controller to perform satisfactorily. It is therefore desirable that secondary path modelers demonstrate consistency in meeting such tolerances.

The adaptive requirement for the ideal modeler assumes an on-line secondary path modeling implementation. The

---
[a)]Electronic mail: bkim@bbn.com and benjkim@hotmail.com
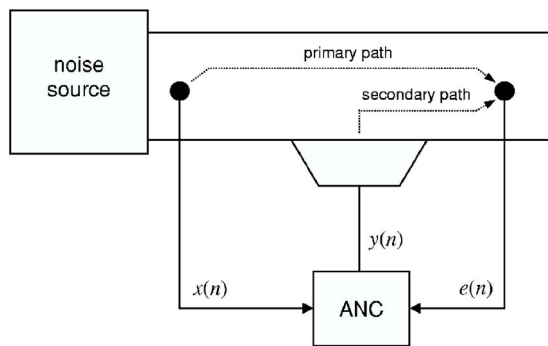
FIG. 1. Secondary path (and primary path) in feedforward ANC system. Secondary path extends from the control filter output to the error sensor input (includes the digital-to-analog and analog-to-digital converters).

secondary path response can be estimated off-line prior to ANC operation and is usually done in the absence of primary noise. In most practical cases, however, the primary noise cannot be switched off. Moreover, the secondary path can be time varying which may render the initial model obsolete and unusable by the FXLMS controller. It follows that there are two basic approaches to on-line secondary path modeling: invasive and noninvasive. The former, often referred to as active system identification, involves the injection of interrogatory signals such as uncorrelated random noise through the control path, a technique known as additive random noise.[15] Besides the counterproductive nature of injecting auxiliary noise for the purposes of noise control, this technique suffers from misalignment in the secondary path response estimate due to residual noise[1] and typically long convergence times if convergence is achieved at all.[4] Variations on this approach have been suggested that attempt to alleviate these problems.[16–18]

The preferred on-line modeling approach is thus noninvasive wherein the control signals are used to perform secondary path modeling (see Fig. 2), an approach often referred to as passive system identification. One such method employs an extended least squares technique whereby the
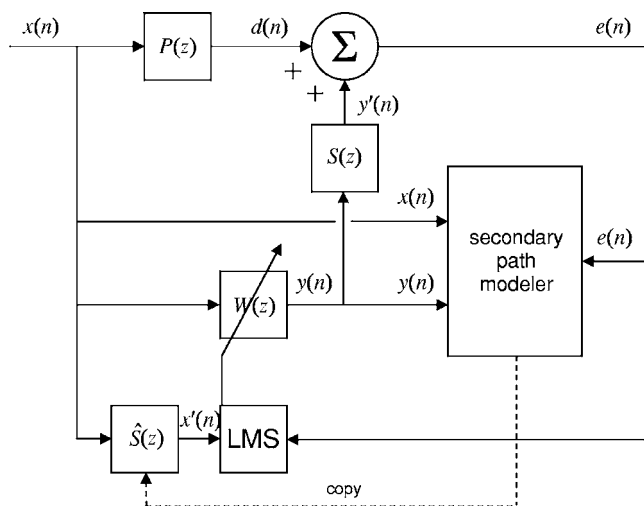
FIG. 2. Noninvasive secondary path modeler within FXLMS framework. $W(z)$, $P(z)$, $S(z)$, and $\hat{S}(z)$ represent the z transforms of the control filter, primary path, secondary path, and secondary path response estimate, respectively.

primary path and secondary path are simultaneously modeled in an "overall" fashion (the primary path modeling would aid in removing from the secondary path response estimate the bias due to primary noise).[19–21] The minimization of the difference between its estimated and measured error signals, however, yields solutions for the primary and secondary path models that are not necessarily unique.[4–6,19] This was observed by Kim in a "two-trial test", the results of which are given in Fig. 3.[22] In this real-time implementation, the primary and secondary path response estimates were recorded upon minimization of both the modeling error and error signal (a 1-kHz sinusoid). The results are indicative of an underdetermined system and can account for the observed instability of FXLMS-based control systems utilizing the overall modeling technique.[22–24] Kuo and Jiang suggested a procedural modification of this overall modeler wherein the primary and secondary path responses are estimated serially rather than in parallel in a "stage switching" process.[25] Numerous on-line noninvasive secondary path modeling methods have been proposed.[4,26–32] The difficulties associated with on-line secondary path modeling have led some to explore alternatives to the FXLMS algorithm, i.e., adaptive algorithms that do not require an explicit model of the plant.[33–39]

In this paper, a new method for system identification/secondary path modeling is presented. The difficulty in identifying the secondary source contribution within the error signal is first sketched in a physical and algebraic context. This description serves then as the point of departure for the development of the frequency-domain-based linear independence method for secondary path modeling. The methodology is formulated for single-channel system identification for both narrowband (single- and multiple-frequency)[22,40] and broadband signals, then generalized for multi-channel system identification. The proposed method is implemented atop a real-time frequency-domain FXLMS controller within the framework of a novel DSP control architecture, and tested on linear "transducerless" ANC systems devised for straightforward analysis of ANC algorithm behavior. Tests were conducted for both periodic and stochastic signals. To quantitatively assess the performance of the new method, particularly in how it manages a changing plant, all secondary path models are compared with the true plant models.

II. THE LINEAR INDEPENDENCE METHOD

A. Underlying problem in on-line secondary path modeling

Modeling of the secondary path is essentially a transfer function measurement of the path between the output of the digital controller that feeds the control source and the input of the error sensor. What renders such a routine measurement nontrivial in active control is the infiltration of primary noise through the secondary path during the identification process as this introduces bias in the secondary path response measurement. In most ANC applications, it is unlikely that the primary noise source can be switched off for system identification purposes (such as electrical transformers and air-
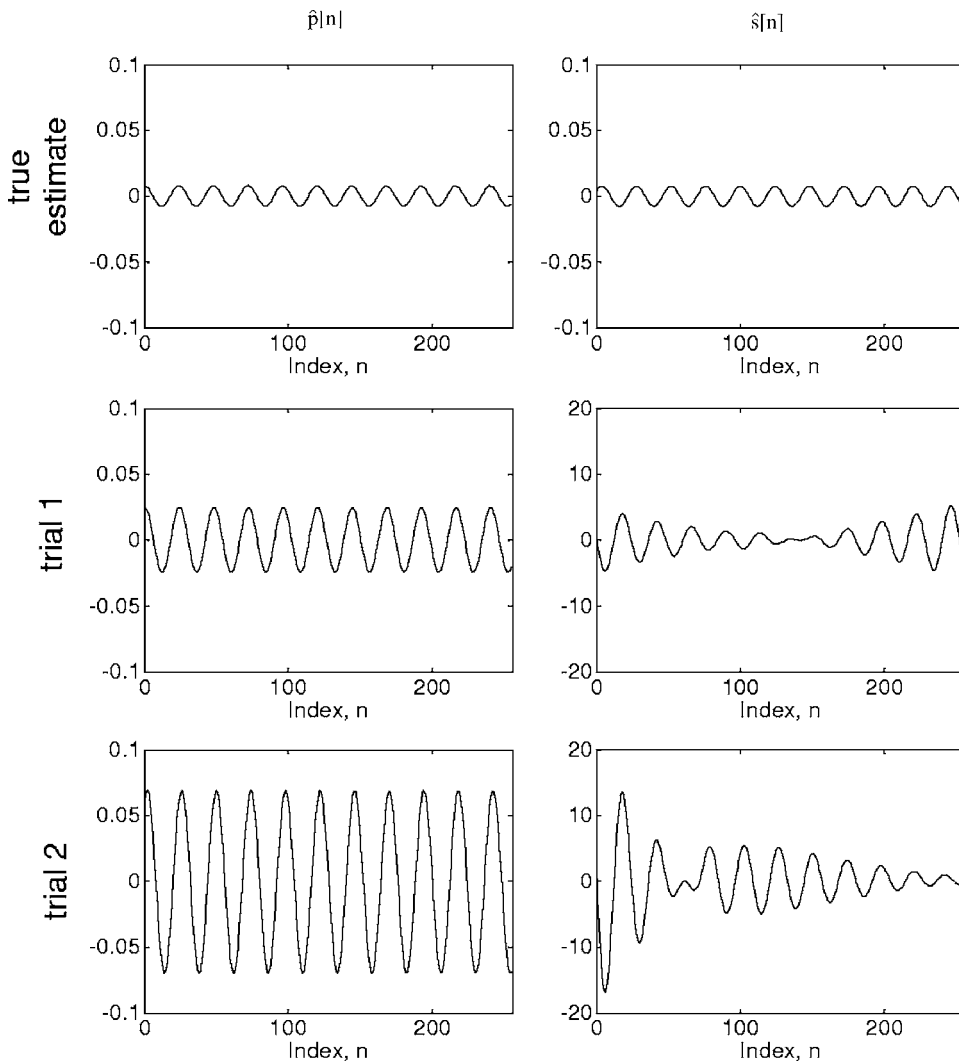
FIG. 3. Two-trial results for overall modeling algorithm. Note relative phases of response estimates and scale of y axes for $\hat{s}[n]$.

plane engines). Many times the secondary path model obtained off-line does not reflect the path during system operation (as in fan-induced duct noise).

The predicament is clear when considering the superposition of two sinusoids of like frequency each with some specified amplitude and phase, a superposition which results in a third sinusoid of equal frequency with corresponding amplitude and phase. If the object was to find the first two sinusoids with knowledge of only the third, even the casual observer would recognize that it is not possible. In fact, an infinite set of sinusoids exists (at the given frequency) that can yield the third sinusoid. It is this correlation that makes on-line secondary path modeling the challenge it has been: it disallows the ability to correctly distinguish the secondary noise contribution in the error signal which when known aids in the accurate determination of the secondary path response. Determining what two numbers make up a specified sum such as in the equation $P+S=1$ summarizes the inherent problem algebraically.

In any single-channel ANC system where source-to-reference feedback is negligible, the time-domain equation relating the physical contributions of the primary noise source and the secondary noise source to the error signal is

$$[x(n) * p(n)] + [y(n) * s(n)] = e(n), \tag{1}$$

where $*$ denotes linear convolution; $x(n)$ and $y(n)$ are the reference and filter output signals, respectively; $p(n)$ and $s(n)$ represent the primary and secondary path impulse responses, respectively; and $e(n)$ is the error signal. The quantity $[x(n)*p(n)]$ corresponds to the primary source contribution to the error signal $e(n)$ [reference signal $x(n)$ traverses primary path $p(n)$] while the quantity $[y(n)*s(n)]$ corresponds to the secondary source contribution [control filter output $y(n)$ traverses secondary path $s(n)$]. The filter output is related to the reference signal as in the equation

$$y(n) = \sum_{l=0}^{L-1} w_l(n)x(n-l), \tag{2}$$

where $w_l(n)$ is the lth coefficient of a length L finite impulse response (FIR) filter. Applying the Fourier transform to Eq. (1) gives

$$[X(k)P(k)] + [Y(k)S(k)] = E(k), \tag{3}$$

where k denotes the kth frequency bin. $X(k)$, $Y(k)$, and $E(k)$ are transforms of the readily observable quantities $x(n)$, $y(n)$, and $e(n)$, respectively, that are obtained via blocks of

time-aligned data, or "snapshots". The unknowns are $P(k)$ and $S(k)$, the primary path and secondary path transfer functions, respectively. For the purposes of comparison, Eq. (3) can be rewritten as

$$\{XP + YS = E\}_k. \tag{4}$$

The similarity between the aforementioned "$P+S=1$" example and Eq. (4) is now evident. For given data set $\{X,Y,E\}$ there exists an infinite set of solutions for P and S.

B. Formulation

The insufficiency of one equation in determining a unique solution for the requisite secondary path response estimate $\hat{S}(k)$ was made salient in the last section. The formulation of the linear independence method begins with the notion that it may be possible to locate the correct set $\{\hat{P}(k),\hat{S}(k)\}$ within the infinite pool of estimates that satisfy Eq. (3) by introducing another equation with a similar set of unknowns. Taking two consecutive snapshots of data allows for a set of two equations with two unknowns,

$$X_A(k)P(k) + Y_A(k)S(k) = E_A(k), \tag{5a}$$

$$X_B(k)P(k) + Y_B(k)S(k) = E_B(k), \tag{5b}$$

where $\{X_A(k),Y_A(k),E_A(k)\}$ corresponds to the first snapshot of data and $\{X_B(k),Y_B(k),E_B(k)\}$ corresponds to the second snapshot of data. $Y_A(k)$ and $Y_B(k)$ represent the control filter outputs according to the equations

$$Y_A(k) = W_A(k)X_A(k), \tag{6a}$$

$$Y_B(k) = W_B(k)X_B(k), \tag{6b}$$

where $W(k)$ is the FFT of the filter's impulse response. Linear independence of the two equations [Eqs. (5a) and (5b)] is achieved by ensuring the inequality

$$W_A(k) \neq W_B(k). \tag{7}$$

Primary and secondary paths are assumed to remain stationary in between and during snapshots. The solution is thus

$$\hat{P}(k) = \frac{E_A(k)Y_B(k) - Y_A(k)E_B(k)}{X_A(k)Y_B(k) - Y_A(k)X_B(k)}, \tag{8a}$$

$$\hat{S}(k) = \frac{X_A(k)E_B(k) - E_A(k)X_B(k)}{X_A(k)Y_B(k) - Y_A(k)X_B(k)}. \tag{8b}$$

This formulation can be conveniently written in matrix form beginning with Eqs. (5a) and (5b):

$$\begin{bmatrix} E_A \\ E_B \end{bmatrix}_k = \begin{bmatrix} X_A & Y_A \\ X_B & Y_B \end{bmatrix}_k \begin{bmatrix} P \\ S \end{bmatrix}_k. \tag{9}$$

Matrix algebra gives

$$\begin{bmatrix} \hat{P} \\ \hat{S} \end{bmatrix}_k = \begin{bmatrix} X_A & Y_A \\ X_B & Y_B \end{bmatrix}_k^{-1} \begin{bmatrix} E_A \\ E_B \end{bmatrix}_k, \tag{10}$$

and inversion of the square matrix results in the equation

$$\begin{bmatrix} \hat{P} \\ \hat{S} \end{bmatrix}_k = \left\{ \frac{1}{X_A Y_B - Y_A X_B} \begin{bmatrix} Y_B & -Y_A \\ -X_B & X_A \end{bmatrix} \right\}_k \begin{bmatrix} E_A \\ E_B \end{bmatrix}_k. \tag{11}$$

This is equivalent to Eqs. (8a) and (8b).

1. Narrowband noise

Narrowband signals may render the 2×2 matrix in Eq. (10) noninvertible for the unexcited FFT bins where, again, k denotes the FFT bin number. These unexcited bins contain zeros, thus the inversion of these matrices would result in null sets of solutions. Although it may be feasible to add low-level broadband noise to excite these bins, spectral leakage prohibits the direct inversion of even the matrices that are associated with frequency bins corresponding to the narrowband signal.[22]

Since the FXLMS algorithm requires only the secondary path response estimate $\hat{S}(k)$, Eq. (8b) will be considered from this point forward though the method outlined in the following can be applied to estimate the primary path $\hat{P}(k)$ (valid for single-channel only). Noting that FFTs are made up of complex numbers, a routine complex conjugate multiply done on this expression yields

$$\hat{S}(k) = \frac{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*}{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*}, \tag{12}$$

where * denotes complex conjugate. The denominator is now real valued. For narrowband signals the denominator is still zero for the unexcited frequency bins. The numerator is a spectrum that happens to preserve all phase information related to the secondary path response. More simply, the numerator is *correct in phase*. For sinusoidal (single-frequency) signals the numerator need only be scaled in magnitude by a single factor to yield an approximate, if not exact, plant estimate. Two methods of numerator scaling are suggested. One method uses the total power of the spectrum in the denominator to scale the numerator for which case utilizing Parseval's theorem would be appropriate. The other method uses the peak value of the spectrum in the denominator. For sinusoidal signals the equations are thus

$$\hat{S}(k) = \frac{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*}{\{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{power}} \tag{13a}$$

and

$$\hat{S}(k) = \frac{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*}{\{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{peak}}. \tag{13b}$$

An extension of this procedure yields the general solution for multiple-frequency narrowband signals. The spectrum is divided into subbands, each subband spanning bins associated with each sinusoid. Each subband in the numerator is then scaled by the peak value or power within the corresponding subband in the denominator. Thus Eqs. (13a) and (13b) can be generalized to accommodate multiple-frequency signals like so:

$$\hat{S}_{B1}(k) = \frac{\{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{B1}}{\{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{power\ or\ peak,B1}}, \tag{14a}$$

$$\hat{S}_{B2}(k) = \frac{\{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{B2}}{\{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{power\ or\ peak,B2}}, \tag{14b}$$

$$\vdots$$

$$\hat{S}_{BN}(k) = \frac{\{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{BN}}{\{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*\}_{power\ or\ peak,BN}}, \tag{14c}$$

where B stands for "band" and N denotes the Nth subband containing the Nth frequency signal. The sinusoidal case is a subset of this solution. This is similar to the scaling of bin-dependent step sizes described by Swanson for controlling multi-frequency narrowband noise using the frequency-domain FXLMS algorithm.[41] Where to divide the spectrum can be determined using peak recognition, a spectral processing and analysis tool that is used extensively in such areas as condition-based maintenance, speech recognition, and mass spectroscopy.

Care must be taken when performing the spectral multiplies in Eq. (14) to guard against circular convolution/correlation, or wraparound, errors. The overlap-save method is preferred for its simplicity whereby time series data are appropriately buffered and pre-/post-zero padded prior to transforming and multiplying. This method is described in detail by Shynk,[42] Reichard, and Swanson.[41,43] For a frequency-domain FXLMS implementation, a control filter of length L necessitates a secondary path response estimate of length $2L$ to carry out the cross spectrum in the LMS update. To yield a $2L$-length secondary path response estimate, data snapshots of lengths $4L$ and $8L$ are needed to properly execute Eq. (14) in this regard. This is summarized in Table I. Longer snapshots can be taken as long as the final time-domain secondary path model is truncated to meet the $2L$ length requirement. A comprehensive step-by-step procedure for implementing the linear independence method equation is given in the Appendix.

2. Broadband noise

Due to the stochastic nature of broadband noise, time averaging is required. The transfer function estimates $\hat{P}(k)$ (for single-channel systems) and $\hat{S}(k)$ are thus calculated according to the equations

$$\hat{P}(k) = \frac{1}{q}\sum_{i=1}^{q} \frac{E_{A,i}(k)Y_{B,i}(k) - Y_{A,i}(k)E_{B,i}(k)}{X_{A,i}(k)Y_{B,i}(k) - Y_{A,i}(k)X_{B,i}(k)}, \tag{15a}$$

$$\hat{S}(k) = \frac{1}{q}\sum_{i=1}^{q} \frac{X_{A,i}(k)E_{B,i}(k) - E_{A,i}(k)X_{B,i}(k)}{X_{A,i}(k)Y_{B,i}(k) - Y_{A,i}(k)X_{B,i}(k)}, \tag{15b}$$

where q is the number of averages. It is assumed here that the secondary path contains a delay no longer than that in the primary path in order to meet the causality constraint for broadband feedforward control systems. Although not necessary for random noise signals, the overlap-save method can be applied nonetheless to ensure that the secondary path response estimate is free of any circular convolution errors.[41] Note that the expressions in Eq. (15) do not incorporate the scaling methodology suggested for narrowband signals.

As stated in the previous section, the lengths of the data snapshots can be varied as long as the secondary path response estimate length requirement is met. Particularly for broadband noise, longer data snapshots provide better resolution and generally yield greater accuracy in the truncated impulse response.

TABLE I. Summary of lengths of relevant quantities relative to control filter length for frequency-domain FXLMS controller and proposed secondary path modeler.

Quantity	Description	Length
w	Control filter	L
s	Secondary path response estimate	$2L$
x_A, y_A, e_A	First data snapshot	$8L$ or $4L$
x_B, y_B, e_B	Second data snapshot	$4L$ or $8L$

3. Multi-channel system identification

Secondary path modeling for multi-channel active control systems presents an even greater challenge due to the additional canceling sources contributing to the control region. Imposing linear independence via the control filter makes a control source physically "audible" to an error sensor regardless of the ambient noise in the control field. This makes the proposed method ideal for multi-channel use where the control field of each error can contain contributions from multiple sources. Sequential and orderly "perturbation" of the control sources in the multi-channel system (in a round robin fashion) allows each of the error sensors to observe and identify the contribution from each secondary source. The equation for multi-channel linear independence secondary path modeling is

$$\hat{S}_{ij}(k) = \frac{X_A(k)E_{B,i}(k) - E_{A,i}(k)X_B(k)}{X_A(k)Y_{B,j}(k) - Y_{A,j}(k)X_B(k)}, \tag{16}$$

where i indicates the ith error sensor and j the jth secondary source. The jth source is perturbed to obtain the path models associated with that particular source. For example, in a two-input/two-output configuration, source 1 is perturbed to obtain path models $\hat{S}_{11}$ and $\hat{S}_{21}$, and source 2 is perturbed to obtain $\hat{S}_{12}$ and $\hat{S}_{22}$. The expression is evaluated in the manner described in the last two sections depending on the characteristics of the noise signal. Again, solving for $\hat{P}(k)$ is not possible in a coupled multi-channel configuration (primary sources cannot be perturbed, therefore bias due to other sources in the system cannot be removed).

C. Notes on implementation

1. Imposing linear independence and on-line modeling

Imposing linear independence forces two lines that are otherwise parallel to intersect. For the two lines defined by Eqs. (5a) and (5b), adjusting the filter coefficients so that $w_B(n) \neq w_A(n)$ changes the slope of the second line relative to the first allowing them to cross. The common point is $(P(k), S(k))$ and it is of course the coordinate $S(k)$ that secondary path modeling targets. Clearly there are countless ways in which this inequality condition can be met. There are, however, important considerations. The linear independence method works by perturbing a secondary source output so that an error sensor physically "hears" the secondary source within the highly correlated noise. Too small a perturbation and the error sensor may not detect the change in the system. This corresponds to an exceedingly shallow intersection of the two lines and, algebraically, a near-singular matrix in Eq. (9). For real systems, the finite precision of the DSP and limited dynamic range of system components in effect broaden and give the otherwise infinitely thin lines width. Though very slight, at such shallow angles this can cause the crossing point to become indistinct. The worst case is that the lines remain parallel despite an intended perturbation.

Thus a minimum perturbation is necessary to achieve unambiguity. This minimum perturbation is measurable and varies with the system.

Even with minimum perturbation considered, there are still countless ways in which one can perturb the control output. A relatively simple method entails adjusting the magnitude of the current control filter coefficients to generate the second set of filter coefficients according to the equation

$$w_B(n) = (1 + \varepsilon)w_A(n), \tag{17}$$

where ε is a real number (positive or negative) except zero. ε should be sufficient in magnitude to make the two-line intersection unambiguous. If it is too large, however, excessive changes in the control source output can result which may be undesirable. The linear independence method can be applied in this manner whenever secondary path models need updating. It can also be performed repeatedly before each and every update of the control filter. When the error signal is minimized, however, the constant perturbations can contribute residual noise. This can be alleviated by setting a threshold below which system identification is not performed.

It is possible to utilize the intrinsic FXLMS updating of the control filter for system identification. Data snapshots would be those taken before and after a filter update. Reference and error signal data from the second snapshot would be used with the newly generated secondary path model to calculate the gradient estimate (this second snapshot would then serve as the first snapshot for the next modeling). The control filter is updated and a new second snapshot taken. This process would be repeated. When the error signal is minimized, however, the changes in the control filter may not be sufficient to yield well-conditioned data sets as described earlier. This again can be alleviated by setting a threshold for the error signal magnitude below which system identification is not performed. A more robust option would be to monitor the *changes* in the control filter (if a change does not result in a perturbation that exceeds the minimum perturbation, then a secondary path model would not be calculated).

This exploitation of the FXLMS adaptation works for single-channel narrowband control and yields a true on-line modeler since "the control system is kept in operation during this repeated measurement" of the secondary path response.[6] For multi-channel implementations, it is arguable whether or not the linear independence method, or any method that is perturbation based, constitutes an on-line modeler since the controller must be suspended momentarily while the individual control sources are perturbed. For broadband ANC as well, the necessary averaging requires that the FXLMS updates be suspended until enough averages have been performed to yield adequate models. Depending on the characteristics of the noise to be canceled, the number of channels, and implementation, the linear independence method may or may not technically constitute on-line modeling. However, the method does effectively remove measurement bias without having to switch off the primary noise and secondary noise sources which can be viewed as requisite for on-line modeling.

2. Update rate and minimum time constraint

Time delays present in the feedforward path should be considered in the implementation of adaptive active noise controllers. The electro-acoustic secondary paths can have delays on the order of milliseconds (it can take that amount of time for any change in the control source output to appear at the error sensor input). If a controller updates itself within this delay period, it can be thought of as running blindly for those iterations. Data used for one update should reflect the change imposed from the previous update, else it is a wasted effort and stability can be compromised.[44] Also of concern is the transients introduced into the system by the altering of the control filter. The FXLMS algorithm and the proposed secondary path modeler are based on linear time invariance, therefore these transients should be allowed to decay out before subsequent measurements are taken.[1,44] This minimum waiting period depends on the system. For infinitely damped or nonreverberant systems this waiting period equals roughly the propagation delay between the control source and the error sensor. For lightly damped or reverberant systems this waiting period equals the propagation delay plus the time it takes for reverberations to decay out. The waiting period is defined as $\tau_{\text{prop+trans}}$ where "prop" denotes propagation and "trans" denotes transients. A block FXLMS implementation (a frequency-domain version is described later) lends itself naturally for filter update rate control and linear independence secondary path modeling.

The linear independence method assumes that the control system is unchanged in between and during the two snapshots required for the modeling. For narrowband noise, as stated in Sec. II B 1, a total of $12L$ data points is required to calculate the response estimate ($8L$ points for one set and $4L$ points for the other set). This amounts to $(12L)f_s^{-1}$ s where f_s is the sampling rate. Allowing the system to reach steady state after perturbation adds additional time. Thus, for the proposed method the minimum time requirement for secondary path modeling is

$$\tau_{\text{min}} = \frac{12L}{f_s} + \tau_{\text{prop+trans}}. \tag{18}$$

The situation is clearly different for broadband signals since averaging is required. The minimum time can vary depending on the desired resolution and accuracy. Increasing the number of averages can yield a more accurate secondary path model but will increase the minimum time for system identification. Larger data blocks can enhance resolution of the model (model can be truncated to appropriate length for x-filtering) but that too will increase the minimum time. Since system identification is best performed as quickly as possible, the specification of these parameters will involve a balancing act of sorts and will depend on the system and application. One implementation is described later in Sec. IV A 4.

III. EXPERIMENTAL SETUP

A. Control architecture and hardware

Normally, DSP system boards are programmed to handle all signal processing tasks, i.e., filtering, calculations and up-
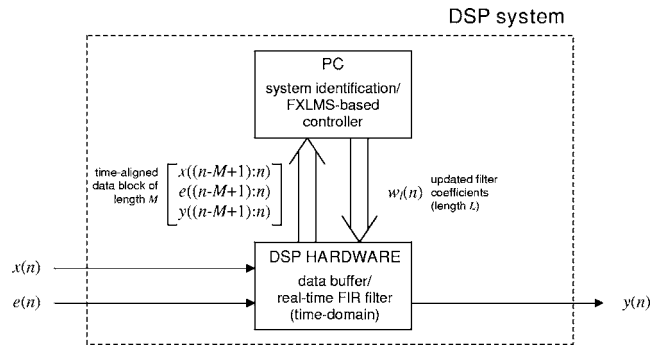
FIG. 4. DSP control architecture for block processing implementation, single channel.

dates to the control filters, system identification if any, etc. For evaluation, troubleshooting, and debugging purposes, this implementation can be cumbersome due to the difficulty in monitoring the processing of the signals. It can be considerably more cumbersome when hardware is programmed in assembly. An innovative DSP control architecture was designed and implemented that is advantageous with respect to ease of use, signals analysis, and operator control.

Figure 4 depicts the control architecture. Signal processing tasks are assigned to both the DSP system board and the PC in which it operates. The DSP board doubles as a real-time filter and data buffer: it performs time-domain filtering on the reference signal using the filter weights "handed over" by the PC, and stores in a circular buffer the incoming signal data $x(n)$ and $e(n)$ and the control output data $y(n)$. The PC meanwhile downloads data from this buffer (data block size specified by the user), uses the data to calculate the control filter coefficients, and uploads the new filter weights to the DSP board. This not only lightens the computational load on the DSP permitting it to operate at higher sample rates and/or run longer filters, this sharing of information with the PC allows the user easy access to data for efficient signals analysis and troubleshooting using relatively syntax-free tools such as MATLAB. It is even possible to do real-time signal processing (minus the control filtering) with a tool such as MATLAB depending on the software and hardware configuration.

The key advantage to this control architecture with specific regards to FXLMS control and linear independence system identification is the relative ease with which one can adjust the control filter coefficients and the rate at which the filter is updated. While this is fundamental to the proper implementation of the proposed linear independence method, it is conducive to the stability of the FXLMS algorithm particularly when applied to systems that have large secondary path delays[44] and/or are lightly damped where transients introduced into the control field can linger.

The DSP hardware used for the following tests comprised of Spectrum DSP32C system boards based on the WE® DSP chip. Each has two 16-bit analog I/O channels onboard with a ±3-V range on both the input and output. Each board, programmed in assembly, was optimized to run a 128-tap filter (128 coefficients) at 24 kHz. The multiplexed circular buffer held 131 072 samples of each of the two in-
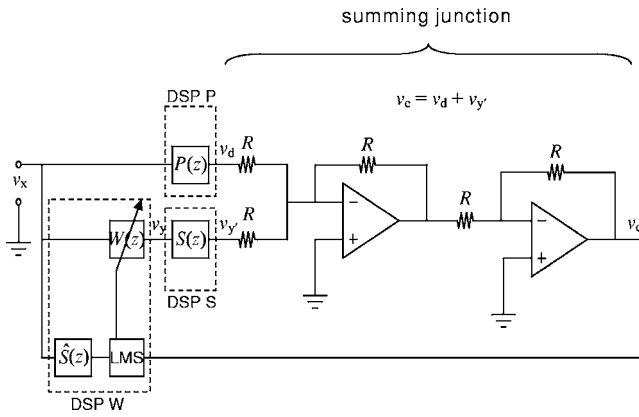
FIG. 5. Transducerless ANC test circuit, single channel. Measured impulse responses can be uploaded to filters **P** and **S** to simulate real environments.

FIG. 6. Transducerless ANC test circuit, two channel. Channels are coupled.

puts and control output, approximately 5.46 s worth of data (at 24 kHz). The cutoff frequency (-3 dB) was set at 9 kHz for the low-pass reconstruction and antialiasing filters. The PC ran on a 200-MHz Intel Pentium processor; Microsoft Visual C/C++ v6.0 was used for the necessary programming and operation of the DSP hardware. The reference signals were taken from a Hewlett Packard 35670A Dynamic Signal Analyzer.

B. The transducerless ANC test system

Op-amp-based transducerless ANC test systems were developed for controlled real-time testing of ANC algorithms. A schematic of the single-channel test system is given in Fig. 5. The summing junction acts as the point of interference between the desired and control signals, albeit in the electrical rather than the usual acoustical domain. There are several key advantages to such a test system: (1) it allows control over most system variables which aids in better understanding algorithm behavior; (2) it allows the testing of active noise controllers without having to go on-site or build large physical systems; (3) it ensures reproducibility over time; and (4) it is inexpensive. A notable feature of this test system is the interchangeable primary and secondary paths made possible by the insertion of digital filters at the appropriate points in the circuit. DSP **P** and DSP **S** help shape the primary and secondary paths, respectively. Like the control filter in DSP **W**, **P** and **S** are FIR filters that operate according to the equations

$$d(n) = \sum_{l=0}^{L-1} p_l(n)x(n-l), \tag{19a}$$

$$y'(n) = \sum_{l=0}^{L-1} s_l(n)y(n-l), \tag{19b}$$

the coefficients of which, namely $p_l(n)$ and $s_l(n)$, are determinable by the user. Each filter has a length of 128 and is run at 24 kHz. One can conceivably insert measured impulse responses of real systems into these filters to simulate those systems.
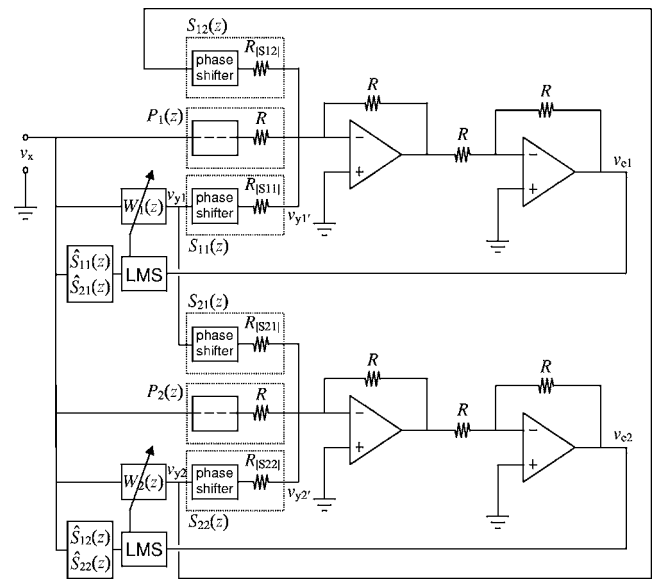
A schematic for the two-channel transducerless ANC test system is given in Fig. 6. Here, in addition to a second summing junction for the second channel, circuit paths were added to couple the two channels as indicated by S_{12} and S_{21} in the figure. Due to the limited number of DSP system boards, analog phase shifters were inserted to allow phase variation of sinusoidal signals in the secondary paths.[45] Magnitudes are determined by the resistor values at the inputs of the summing junctions.

C. LMS control filter implementations

A frequency-domain version of the normalized leaky FXLMS algorithm was implemented for the single-channel controller, expressed as

$$W_{m+1}(k) = \nu W_m(k) - \alpha_m(k)R_m^*(k)E_m(k), \tag{20}$$

where

$$R_m(k) = X_m(k)\hat{S}(k) \tag{21}$$

is the filtered-x signal spectrum, ν is the leakage factor, and $\alpha_m(k)$ is the convergence coefficient. The following frequency-domain normalized leaky multiple-error FXLMS algorithm[6] was implemented for the two-channel test:

$$W_{1,m+1}(k) = \nu W_{1,m}(k) - \alpha_{1,m}(k)\{R_{m,11}^*(k)E_{1,m}(k)$$
$$+ R_{m,21}^*(k)E_{2,m}(k)\}, \tag{22a}$$

$$W_{2,m+1}(k) = \nu W_{2,m}(k) - \alpha_{2,m}(k)\{R_{m,12}^*(k)E_{1,m}(k)$$
$$+ R_{m,22}^*(k)E_{2,m}(k)\}, \tag{22b}$$

where

$$R_{m,ij}(k) = X_m(k)\hat{S}_{ij}(k) \tag{23}$$

and i and j indicate the ith sensor and jth secondary source. For these tests ν was set to 0.999. Assuming the secondary path model is phase accurate the prefiltering operation in Eqs. (21) and (23) time aligns the reference and error

signals to allow for a valid cross spectrum in the gradient estimate. Again, care must be taken to avoid circular convolution/correlation errors. The FFT quantities needed for proper evaluation of these expressions using the overlap-save method are summarized below:

$$W_m(k) = \mathrm{FFT}[w_{mL}(0) \cdots w_{mL}(L-1) \, 0 \cdots 0], \tag{24}$$

$$R_m(k) = \mathrm{FFT}[r(mL-L) \cdots r(mL-1)$$
$$r(mL) \cdots r(mL+L-1)], \tag{25}$$

$$E_m(k)$$
$$= \mathrm{FFT}[0 \cdots 0 \, e(m2L+2L-L) \cdots e(m2L+2L-1)], \tag{26}$$

$$\hat{S}_m(k) = \mathrm{FFT}[s_{m2L}(0) \cdots s_{m2L}(2L-1) \, 0 \cdots 0], \tag{27}$$

$$X_m(k) = \mathrm{FFT}[x(m2L-2L) \cdots x(m2L-1)$$
$$x(m2L) \cdots x(m2L+2L-1)], \tag{28}$$

where L is the length of the control filter.

The convergence coefficient $\alpha_m(k)$ for the narrowband tests was made frequency *in*dependent so that

$$\alpha_m = \frac{\tilde{\alpha}}{[\Pi_{rr,m}(k)]_{\max}}, \tag{29}$$

where $\tilde{\alpha}$ is a normalized step size, $\Pi_{rr,m}(k)$ is the power spectral density of the filtered-x signal, and $[\Pi_{rr,m}(k)]_{\max}$ is the maximum absolute value in the power spectral density vector. This was done to sidestep the problems of spectral division mentioned in Sec. II B 1, namely those of spectral leakage and division by zero for unexcited frequency bins.[41] Each of the two FXLMS update equations in Eq. (22) uses a convergence coefficient that is scaled by the larger maximum of the two power spectral densities in its respective equation. The convergence coefficient for the broadband tests was made frequency dependent so that

$$\alpha_m(k) = \frac{\tilde{\alpha}}{\Pi_{rr,m}(k)}, \tag{30}$$

where, again, $\tilde{\alpha}$ is the normalized step size and $\Pi_{rr,m}(k)$ is the filtered-x signal power spectral density.[6] For both narrowband and broadband tests, the normalized step size $\tilde{\alpha}$ was set to 0.9.

IV. RESULTS AND DISCUSSION

Presented in this section are five tests that demonstrate the implementation and efficacy of the linear independence method. Efficacy of a secondary path modeler is best demonstrated by comparing its response estimates with the true models. The true models were obtained using the following adaptive modeler:

$$\hat{S}_{m+1}(k) = \hat{S}_m(k) + \alpha_m Y_m^*(k) E_m^y(k). \tag{31}$$

The corresponding block diagram is given in Fig. 7. Equation (31) is a basic LMS recursion that estimates the secondary path in an off-line fashion. The primary noise is made
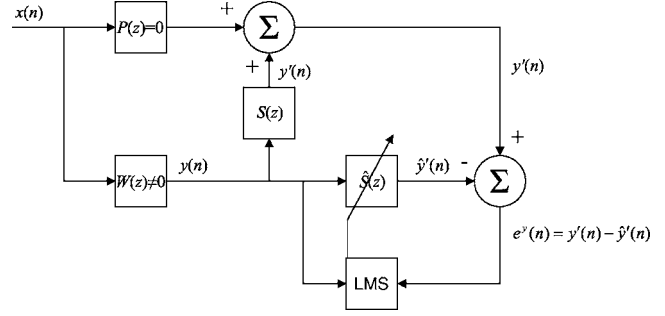
FIG. 7. Block diagram of LMS filter used in determining true secondary path model.

quiet by a simple disconnect in the transducerless ANC test circuit to allow for an unbiased secondary path model.

A. Single channel

The transducerless ANC test circuit pictured in Fig. 5 was used for the following single-channel tests. For both the narrowband and broadband tests, a pure delay of 100 samples was inserted into the primary path via the filter in DSP **P**; $P(z)$, the z transform of the primary path, was approximately z^{-100}. This was done primarily to allow for causality for the broadband test (causality is of little concern for narrowband signals). For the narrowband tests four impulse responses were generated using random numbers each with length 128 and range -0.2 to 0.2; these would be inserted into DSP **S** to shape the secondary path. The impulse responses and their respective magnitude and phase responses are shown in Fig. 8 and are labeled S1, S2, S3, and S4. While these do not represent responses of real systems, their value lies in the variation in phase and magnitude as the impulse responses are "swapped out". The secondary path modeler would have to cope with the changes in the path regardless of authenticity. For the broadband test pure delays were manufactured for the secondary paths so that $S(z) \approx \{z^{-10}, z^{-25}, z^{-50}, z^{-75}\}$.

1. Single frequency, changing secondary path

A sinusoidal signal of frequency 1 kHz and amplitude 1 V_p was used for the reference signal with the desired signal derived from this reference. DSP **S** is initialized with secondary path S1. Noise cancellation performance is given in Fig. 9. At ANC system startup, the control filter is zeroed $[W_A(z)=0]$ resulting in no secondary source output. A snapshot of data is taken (iteration 0); the control filter is then adjusted so that $W_B(z)=0.1$; a second snapshot is taken that reflects the change in the control field (iteration 1); the control filter is set back to zero; the secondary path model is calculated. The model is then used by the FXLMS filter for 14 iterations. The secondary path is then interchanged with S2 after iteration 15. A snapshot of data is taken (iteration 16), then the control filter coefficients present at iteration 16 are multiplied by 1.1 ($\varepsilon=0.1$) prior to acquisition of the second data set so that $w_B(n)=1.1w_A(n)$. This multiplication by 1.1 is done for the subsequent secondary path modeling for paths S3 and S4. This particular value of ε was chosen for no
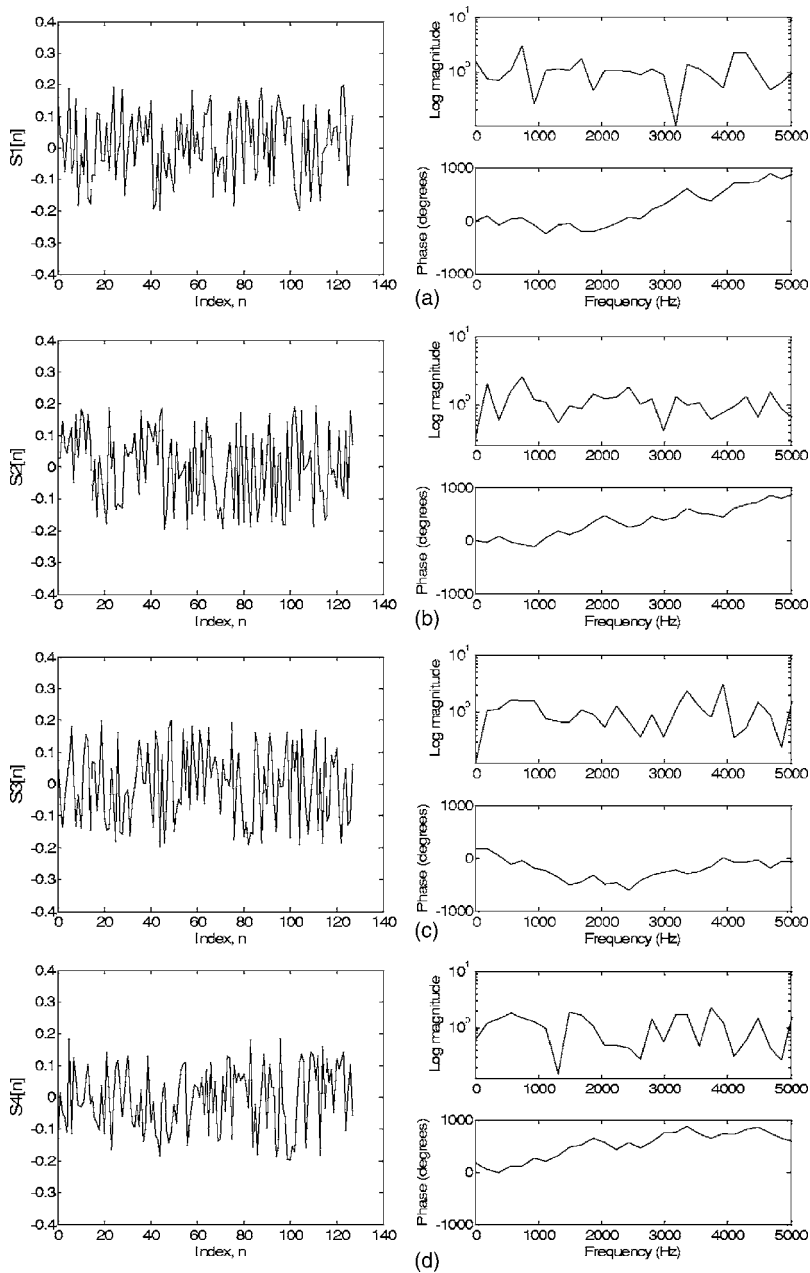
FIG. 8. Random number generated secondary path impulse responses of length 128 and their respective magnitude and phase responses: (a) S1, (b) S2, (c) S3, and (d) S4.

reason other than to ensure that the perturbation would be visible in the noise cancellation performance plots.

Figure 9 compares the noise cancellation of the controller using three different secondary path models: (1) the true model, (2) the model obtained by scaling the numerator by denominator power, and (3) the model obtained by scaling the numerator by denominator peak value. The respectable performance of the controller—attenuation of ~50 dB after only ten iterations—can be attributed to the high accuracy of the secondary path models, revealed in Fig. 10, as well as the controlled update rate. In this near-infinitely damped system with negligible propagation delay, ten iterations (two for "sys id" and eight for FXLMS adaptation) corresponds to 0.15 s or 150 ms ($12L$ samples for sys id and $2L \times 8$ samples for FXLMS adaptation where $L=128$). Phases are exact while magnitudes are very closely approximated to the true magnitude. The models obtained using the peak scaling method are acceptable because they are consistent—consistently be-
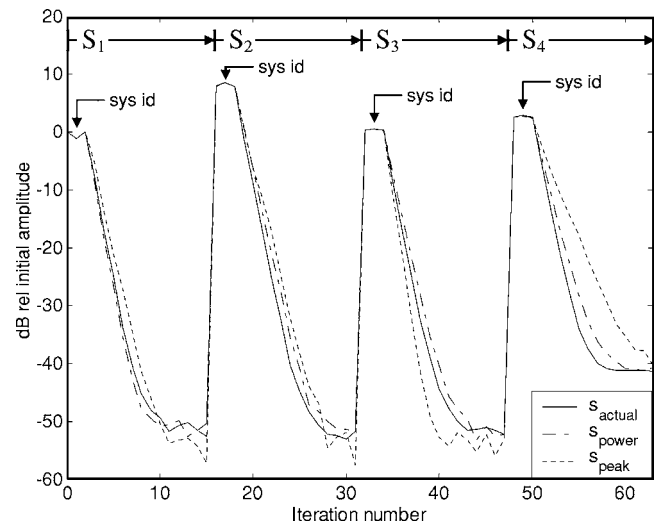
FIG. 9. Cancellation performance of 1-kHz sinusoid in changing secondary path.
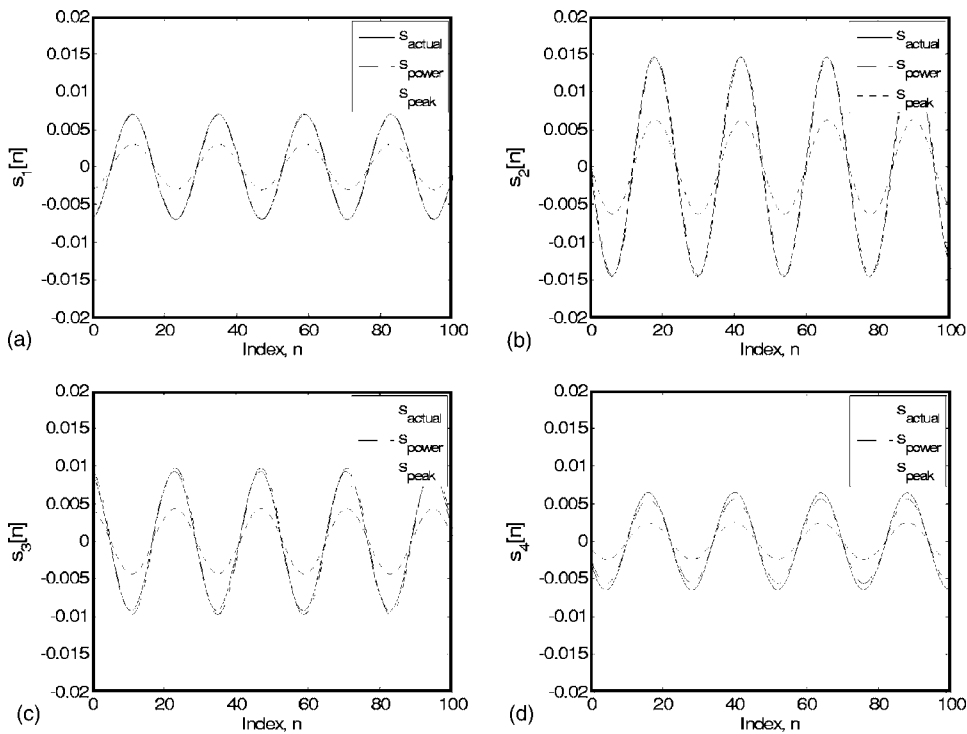
FIG. 10. Secondary path model comparison for 1-kHz sinusoid in changing path: (a) $\hat{s}_1$, (b) $\hat{s}_2$, (c) $\hat{s}_3$, and (d) $\hat{s}_4$. Models reveal phase delays of (a) 11, (b) 18, (c) 21, and (d) 16 samples. Amplitude of 0.0078 corresponds to unity gain.

low the actual magnitude—and therefore require no rescaling of step size to maintain FXLMS controller stability that an otherwise overestimated magnitude can compromise.

Phase delay information can be readily extracted from the plots of these single-frequency secondary path models. The center point or peak of the leftmost crest serves as the reference point. The distance from this point to the y axis is the phase delay in samples. In Fig. 10(a), the first crest is centered at approximately index 11 indicating an 11-sample phase delay in the path *at 1 kHz*. This phase delay changes according to the secondary path as observed in the other models in the figure. Note that phase delay is ambiguous within an additive constant of $2\pi N$ rad where N is any integer. Extracting absolute magnitude response requires knowledge of the FFT size. Here, an amplitude of 0.0078 corresponds to unity gain.

Before continuing, it is worthwhile to review why the secondary path models in Fig. 10 do not resemble the random number generated impulse responses in Fig. 9, e.g., why $s_1[n] \neq S1[n]$. Because the control signal is being used to estimate the secondary path, the models can reflect only those frequencies present in that signal. As Fig. 10 illustrates only the response at 1 kHz is estimated, not a broad frequency range. The same holds true for the control filters: filter weights will reflect only those frequencies present in the system. Thus the system will need to be reidentified when there is a change in frequency content (the control filter must correspondingly adapt to the new frequencies). This is the motivation for the next test.

2. Stationary secondary path, shifting frequency

Noninvasive secondary path modelers use the control signals to perform system identification, therefore any change in the frequency content will require the modeler to

reidentify the system to reflect the current frequencies. For this test, the frequency of the reference signal (still 1 V_p amplitude) was shifted from 1 to 4 kHz in 1-kHz steps. The secondary path was made stationary (path S1 was used throughout the test). The ANC performance is shown in Fig. 11. The procedure for imposing linear independence and adapting the control filter (14 iterations) is identical to that used for the first test. The secondary path response estimates are revealed in Fig. 12. As in the prior test, the phases correspond exactly with those of the actual models. The magnitudes are similar to the results of the last section.

3. Multiple frequency, changing secondary path

Two-frequency signals were used to validate the multiple frequency generalization of the linear independence
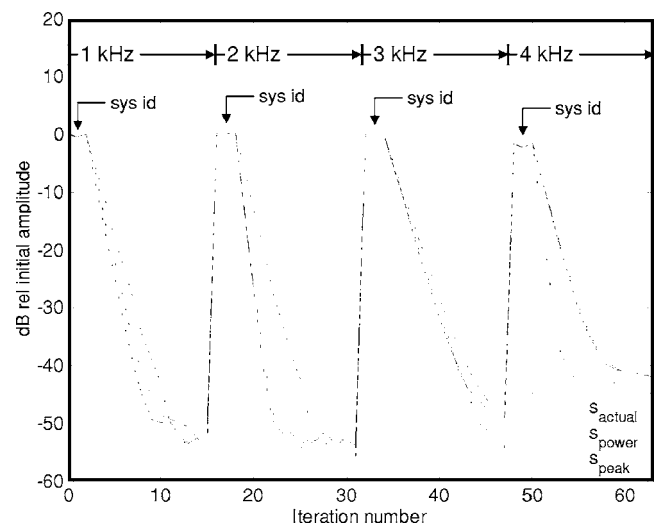
FIG. 11. Cancellation performance of changing frequency in stationary secondary path.
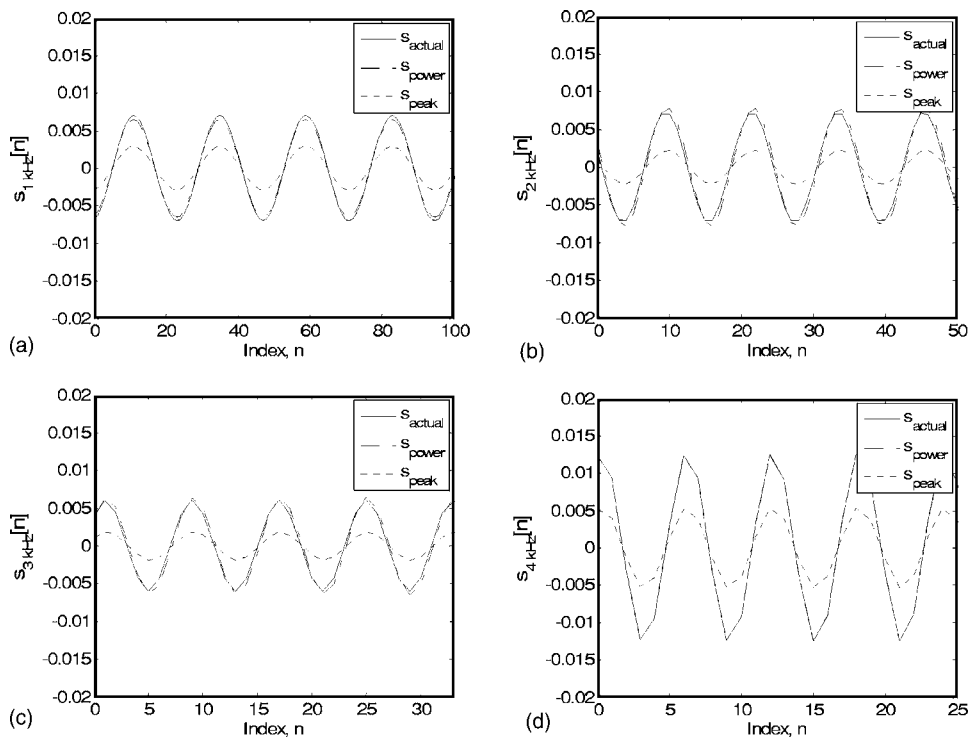
FIG. 12. Secondary path model comparison for changing frequency in stationary path: (a) $\hat{s}_{1\text{ kHz}}$, (b) $\hat{s}_{2\text{ kHz}}$, (c) $\hat{s}_{3\text{ kHz}}$, and (d) $\hat{s}_{4\text{ kHz}}$. Models reveal phase delays of (a) 11, (b) 9, (c) 2, and (d) 0 samples. Amplitude of 0.0078 corresponds to unity gain.

method. Peak recognition capability would assist in dividing the frequency spectrum into the appropriate subbands particularly in systems where the frequencies are time varying. Validation of the proposed method, however, does not require such capability. Subbands can be set *a priori* if the frequencies in the control system are unchanging. This is what is done here.

The two frequencies used for this test were 1 and 2 kHz, each with amplitude 0.5 V_p (lowered from 1 V_p to avoid clipping of the A/D and D/A converters). The spectrum was divided into two subbands at the midpoint between the two peaks with each subband accommodating its respective sinusoid. The procedure for imposing linear independence is

identical to that used in the previous tests. For the FXLMS adaptation, 30 iterations are allowed after the modeling of each secondary path.

Figure 13 shows the cancellation performance for the two-frequency error signal and Fig. 14 reveals the secondary path models. Again, the models scaled using the peak subband values have magnitudes that are consistently less than that of the true magnitudes. The models scaled by power generally have magnitudes that vary relative to the true magnitudes. This is an artifact of the spectrum subbanding where frequency bins that contribute to the power of one sinusoid can lie in an adjacent subband. The contributions of these outlying bins are generally small, hence the correspondingly small variations in the magnitude. Again, the phase is accurate.

4. Random noise, changing secondary path

Random noise, band limited at 6.4 kHz, was used for the broadband test. The controller was programmed to execute the linear independence method using 1024- and 2048-sample data blocks (overlap-save method was utilized). The resulting secondary path impulse response is 512 points long but would be truncated to 256 points (longer time data increases the resolution of the FFTs, resulting in more accurate secondary path impulse responses). Fifty averages were performed. Rather than impose linear independence 50 times, however, which would require the control filter to be adjusted twice that number of times (w_A to w_B and back to w_A), only one adjustment was made in between two snapshots, each snapshot of sufficient length to provide for 50 data windows of lengths 1024 and 2048 with 50% overlap. The first and second snapshots are thus 26 112 and 52 224 samples in

FIG. 13. Cancellation performance of two-frequency signal (1 and 2 kHz) in changing secondary path.
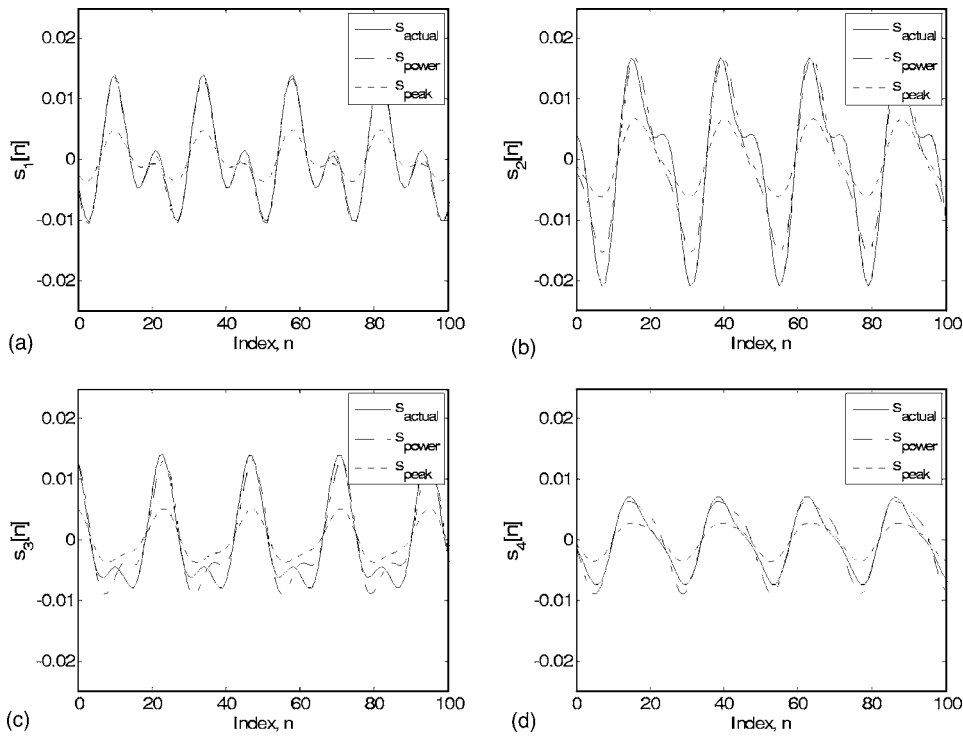
FIG. 14. Secondary path model comparison for two-frequency signal in changing path: (a) $\hat{s}_1$, (b) $\hat{s}_2$, (c) $\hat{s}_3$, and (d) $\hat{s}_4$.

length or 1.088 and 2.176 s, respectively. Of course, this assumes no change in the secondary path during the sys id period.

Figure 15 shows the time history of the error signal for this test. For the initial secondary path modeling, $W_A(z)=0$ and $W_B(z)=0.1$ as was done for the previous tests. For all subsequent modeling, $W_B(z)$ was set equal to zero for the second snapshot. In the figure, this is reflected in the sudden decrease in amplitude during each sys id period as only the primary source is contributing to the error signal during the latter half. This was done to improve performance of the cancellation system where less noise is preferred. Imposing linear independence in a manner similar to that in the other tests would sustain the high level of noise in the control field during the relatively lengthy sys id period of 3.264 s. Upon

FIG. 15. Cancellation performance of random noise in changing secondary path.

modeling the secondary path, the FXLMS algorithm is iterated 256 times before the secondary path is interchanged. The amount of cancellation ranges from 30 to 40 dB across the passband.

Figure 16 reveals the secondary path impulse response estimates and also the primary path impulse response estimate as determined by the linear independence method. The primary path impulse response estimate indeed reflects the 100-sample delay yielded by the filter in DSP **P**; the secondary path impulse response estimates reflect the pure delays of 10, 25, 50, and 75 samples imposed by DSP **S**.

B. Multi-channel

The two-channel transducerless ANC test circuit pictured in Fig. 6 was used for this final test. The primary paths for both channels were such that $P_1(z)=P_2(z)=1$ (no delay or gain) making $d(n)=x(n)$. The resistor values which determine the magnitudes of the paths were set as follows: $R=R_{|S11|}=R_{|S22|}=10$ kΩ, $R_{|S12|}=20$ kΩ, and $R_{|S21|}=50$ kΩ. This translates to a magnitude response of unity for the direct secondary paths (control source 1 to error sensor 1, control source 2 to error sensor 2); a magnitude of 0.5 for the coupled path from control source 2 to error sensor 1; and a magnitude of 0.2 for the coupled path from control source 1 to error sensor 2. These values were chosen to fall within the secondary path magnitude constraints of the multiple-error FXLMS algorithm.[6] They yield a small eigenvalue spread for the matrix $\mathbf{S}^H\mathbf{S}$ (H denotes Hermitian or conjugate transpose) that would allow for relatively straightforward convergence (large eigenvalue spreads are generally indicative of ill-conditioned systems). For the specified paths the eigenvalue set is {0.9, 1.1}. An inverter was used to set the phase of S_{21}
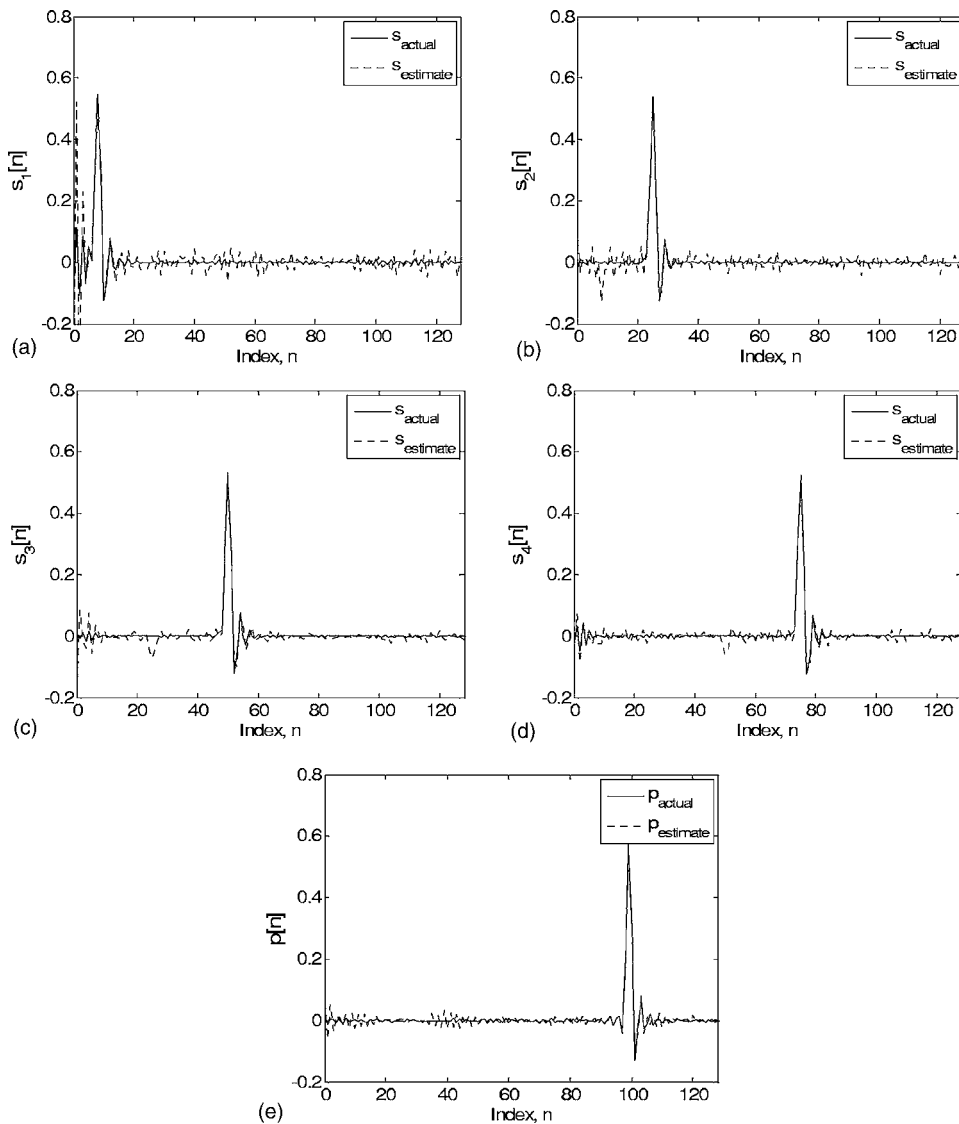
FIG. 16. Secondary path model comparison: (a) 10-sample pure delay, (b) 25, (c) 50, and (d) 75; primary path model comparison: (e) 100-sample pure delay.

to 180°; the phases of S_{11}, S_{12}, and S_{22} were set at random. The reference signal was sinusoidal with a frequency of 1 kHz and amplitude 1 V_p.

Figure 17 shows the cancellation performance using the
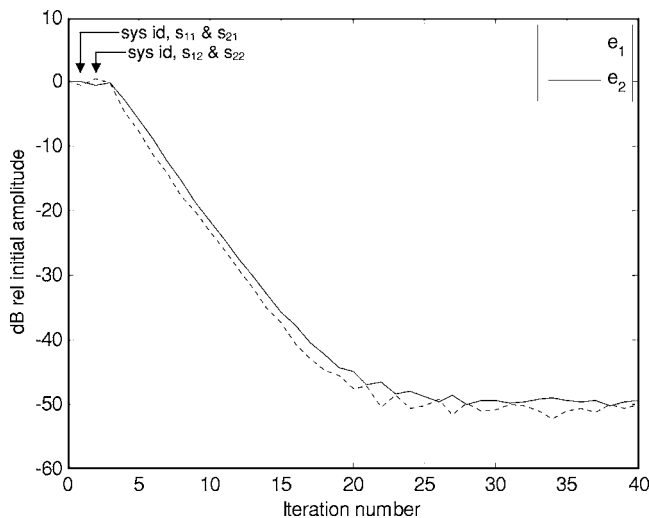
FIG. 17. Two-channel cancellation performance.

proposed secondary path modeling method. The points at which linear independence was imposed are indicated by the arrows. The control output for channel 1 was adjusted first to model S_{11} and S_{21}, then channel 2 to model S_{12} and S_{22}. The secondary path response estimates scaled by power (instead of the peak denominator value) were used for the cancellation shown in the figure. Only 15 iterations or so of the LMS algorithm were necessary to achieve 50 dB of cancellation for both channels. This again is attributable to the accuracy of the secondary path models revealed in Fig. 18. Their magnitudes can be further verified from the resistor values used. The amplitudes of the coupled paths 2-1 and 1-2 are respectively $\frac{1}{5}$ and $\frac{1}{2}$ that of the direct paths, which corresponds exactly to the secondary path magnitudes imposed by the resistors (amplitude of 0.0078 corresponds to unity gain).

It is deemed critical that in coupled multi-channel, multiple-error FXLMS-based ANC systems the secondary paths be identified as accurately as possible. As stated previously, the performance of the multiple-error FXLMS algorithm depends greatly on the relative magnitudes of the secondary paths as the eigenvalue spread of the matrix $\mathbf{S}^H\mathbf{S}$ determines the convergence. In addition, the rate of conver-
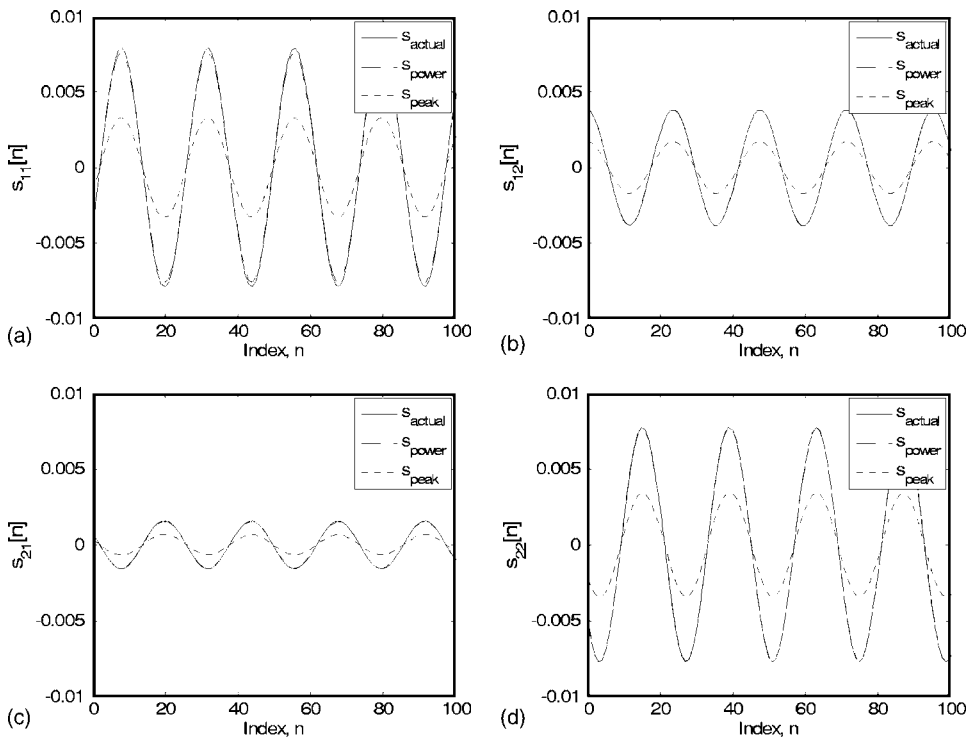
FIG. 18. Secondary path model comparison for direct and coupled paths: (a) $\hat{s}_{11}$, (b) $\hat{s}_{12}$, (c) $\hat{s}_{21}$, and (d) $\hat{s}_{22}$. Models reveal phase delays of (a) 8, (b) 23, (c) 20, and (d) 15 samples. Amplitude of 0.0078 corresponds to unity gain.

gence is limited by the step size which is governed by the maximum eigencomponent of $\mathbf{S}^H\mathbf{S}$.[4] The relative phases between the secondary paths also play a significant role in the convergence of the algorithm.[46–48] In short, accurate secondary path response estimates would aid tremendously in the analysis of robustness of multi-channel ANC systems.

It is interesting to note the phase of path S_{21} which subtly exposes a delay contributed by the DSP system board inputs and outputs (D/A and A/D converters, reconstruction and antialiasing filters). The left-most trough should be centered at index 0 which would correspond to the 180° phase shift imposed by the inverter. Instead it is centered at index 8 indicating an eight-sample phase delay in the control path. Measurement of the DSP hardware revealed a linear phase delay of about 7.6 samples, or 317 μs, relative to the 24-kHz sampling rate.

V. CONCLUSIONS

A system identification methodology has been introduced that reduces the linear dependence problem associated with on-line noninvasive secondary path modeling to a solvable system of two equations and two unknowns. Achieving linear independence of such a system is the critical feature of the method, hence the proposed designation. The procedure requires two consecutive snapshots of $\{x(n), e(n), y(n)\}$ data, both made linearly independent by a deliberate perturbation of the control output in between snapshots. Physically, the perturbation makes the control source discernable relative to the primary noise and other noise sources in the control system. Narrowband tests reveal secondary path models that are exact in phase and closely approximated in magnitude compared to the true models. Impulse response estimates obtained from the broadband tests were nearly identical to the true impulse responses. The results demonstrate the efficacy of the proposed linear independence method in removing the encumbering measurement bias due to primary noise and other control field contributions. Its appeal and value lie in its mathematical simplicity and physical intuitiveness.

APPENDIX: IMPLEMENTING THE MODELING EQUATION

The following is a detailed procedure for implementing the linear independence secondary path modeling equation $\hat{S}(k)$

$$= \frac{[X_A(k)E_B(k) - E_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*}{[X_A(k)Y_B(k) - Y_A(k)X_B(k)][X_A(k)Y_B(k) - Y_A(k)X_B(k)]^*},$$

(A1)

where $X(k)$, $Y(k)$, and $E(k)$ are transforms of time domain vectors $x(n)$, $y(n)$, and $e(n)$, respectively, and subscripts A and B correspond to time-aligned data snapshots reflecting control filters w_A and w_B, respectively. Of particular concern is how to effectively utilize the overlap-save method to avoid circular convolution/correlation errors due to wraparound when performing the numerous spectral multiplications in Eq. (A1).[22,40] The procedure assumes that time data vectors are arranged such that the oldest sample is first and the newest sample is last (from left to right). It also assumes a control filter of length L. Step 9 is a relatively key step.

(1) Take 8L-length snapshot of aligned time data. This first snapshot is the set $\{x_{A,8L}, y_{A,8L}, e_{A,8L}\}$ which contains reference, filter output, and error signal vectors, respectively, and corresponds to control filter w_A.

(2) Adjust control filter such that $w_B \neq w_A$.

(3) Wait for transients to decay out, i.e., for the system to return to steady state.

(4) Take another snapshot of data of length 4L. Second snapshot is the set $\{x_{B,4L}, y_{B,4L}, e_{B,4L}\}$ (corresponds to control filter w_B).

(5) Take 8L-point discrete Fourier transform (DFT) of first data set. This set is now $\{X_{A,8L}, Y_{A,8L}, E_{A,8L}\}$. It may be necessary to zero 0-Hz frequency bins to eliminate any DC offset in DFTs.

(6) *Post*-zeropad each vector in second data set with 4L zeros. Take 8L-point DFT. This set is now $\{X_{B,8L(zp)}, Y_{B,8L(zp)}, E_{B,8L(zp)}\}$ where "zp" stands for "zeropadded." Again, it may be necessary to zero 0-Hz frequency bins to eliminate any DC offset in DFTs.

(7) Compute $X_{A,8L}E_{B,8L(zp)} - E_{A,8L}X_{B,8L(zp)}$. The resultant quantity will be referred to as U_{8L}.

(8) Compute $X_{A,8L}Y_{B,8L(zp)} - Y_{A,8L}X_{B,8L(zp)}$. The resultant quantity will be referred to as V_{8L}.

(9) Take inverse DFT of U_{8L} to give u_{8L}. Keep *last* **2L** points of u_{8L}, pre-pad with 2L zeros, and take 4L-point DFT to yield $U_{4L(zp)}$.

(10) Take inverse DFT of V_{8L} to give v_{8L}. Discard first 4L points, or left half of v_{8L}, and take 4L-point DFT to yield V_{4L}. Take complex conjugate of V_{4L} to get V_{4L}^{*}.

(11) The expression for the secondary path model is now

$$S_{4L} = \frac{U_{4L(zp)}V_{4L}^{*}}{V_{4L}V_{4L}^{*}}.$$

For single-frequency narrowband noise the numerator can be scaled by either the power or peak value of the spectrum in the denominator. Evaluate expression.

(12) Take inverse DFT of S_{4L} to give s_{4L}. Discard last 2L points of s_{4L} to give s_{2L}. s_{2L} is the 2L-point time-domain secondary path response estimate.

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1985).

[2] P. A. Nelson and S. J. Elliott, *Active Control of Sound* (Academic, London, 1992).

[3] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, London, 1996).

[4] S. M. Kuo and D. R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations* (Wiley, New York, 1996).

[5] C. H. Hansen and S. D. Snyder, *Active Control of Noise and Vibration* (E & FN Spon, London, 1997).

[6] S. J. Elliott, *Signal Processing for Active Control* (Academic, London, 2001).

[7] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," IRE WESCON Conv. Rec., pt. 4, pp. 96–104 (1960).

[8] B. Widrow, J. R. Glover, J. M. Cool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive noise canceling principles and applications," Proc. IEEE **63**, 1962–1716 (1975).

[9] D. R. Morgan, "An analysis of multiple correlation cancellation loops in a filter with an auxiliary path," IEEE Trans. Acoust., Speech, Signal Process. **28**, 454–467 (1980).

[10] B. Widrow, D. Shur, and S. Shaffer, "On adaptive inverse control," in Proc. 15th ASILOMAR Conference on Circuits, Systems and Computers (1981), pp. 185–195.

[11] J. C. Burgess, "Active adaptive sound control in a duct: A computer simulation," J. Acoust. Soc. Am. **70**, 715–726 (1981).

[12] S. D. Snyder and C. H. Hansen, "The influence of transducer transfer functions and acoustic time delay on the LMS algorithm in active noise control systems," J. Sound Vib. **140**(3), 409–424 (1990).

[13] C. C. Boucher, S. J. Elliott, and P. A. Nelson, "Effect of errors in the plant model on the performance of algorithms for adaptive feedforward control," Proc. Inst. Electr. Eng. **138**, 313–319 (1991).

[14] S. D. Snyder and C. H. Hansen, "The effect of transfer function estimation errors on the filtered-x LMS algorithm," IEEE Trans. Signal Process. **42**, 950–953 (1994).

[15] L. J. Eriksson and M. C. Allie, "Use of random noise for on-line transducer modeling in an adaptive active attenuation system," J. Acoust. Soc. Am. **85**, 797–802 (1989).

[16] C. Bao, P. Sas, and H. Van Brussel, "Adaptive active control of noise in 3-D reverberant enclosures," J. Sound Vib. **161**(3), 501–514 (1993).

[17] S. M. Kuo and D. Vijayan, "A secondary path modeling technique for active noise control systems," IEEE Trans. Speech Audio Process. **5**(4), 374–377 (1997).

[18] M. Zhang, H. Lan, and W. Ser, "A robust online secondary path modeling method with auxiliary noise power scheduling strategy and norm constraint manipulation," IEEE Trans. Speech Audio Process. **11**(1), 45–53 (2003).

[19] S. D. Sommerfeldt and J. Tichy, "Adaptive control of a two-stage vibration isolation mount," J. Acoust. Soc. Am. **88**, 938–944 (1990).

[20] J. Tapia and S. M. Kuo, "New adaptive on-line modeling techniques for active noise control systems," Proc. IEEE Int. Conf. on Sys. Eng. (1990), pp. 280–283.

[21] L. Eriksson, "Active acoustic attenuation system with overall modeling," U.S. Patent 4,987,598 (1991).

[22] B. J. Kim, "A method for noninvasive on-line secondary path modeling for the filtered-x LMS algorithm for active control of periodic noise," Ph.D. thesis, The Pennsylvania State University, University Park, PA, 2002.

[23] C. Bao, P. Sas, and H. Van Brussel, "Comparison of two on-line identification algorithms for active noise control," Proc. Recent Adv. in Active Control of Sound Vib. (1993), pp. 38–51.

[24] J. Tichy, private communication, 2001.

[25] S. M. Kuo and X. H. Jiang, "Secondary path modeling technique for transformer active noise control," Proc. InterNoise (1995), pp. 505–508.

[26] F. Jiang, N. Ojiro, H. Ohmori, and A. Sao, "Fully adaptive noise control dealing with changes of secondary path response," Proc. IEEE Int. Conf. Control App. (1996), pp. 934–941.

[27] T. Usagawa, Y. Shimada, Y. Nishimura, and M. Ebata, "Adaptive algorithm for active control of harmonic signal with online error path modelling," Proc. Active (1997), pp. 837–848.

[28] H. S. Kim and Y. Park, "Delayed-X LMS algorithm: An efficient ANC algorithm utilizing robustness of cancellation path model," J. Sound Vib. **212**(5), 875–887 (1998).

[29] S. M. Lee, H. J. Lee, C. H. Yoo, D. H. Youn, and I. W. Cha, "An active noise control algorithm for controlling multiple sinusoids," J. Acoust. Soc. Am. **104**, 248–254 (1998).

[30] W. C. Nowlin, G. S. Guthart, and G. K. Toth, "Noninvasive system identification for multichannel broadband active noise control," J. Acoust. Soc. Am. **107**, 2049–2060 (2000).

[31] X. Qiu and C. H. Hansen, "An algorithm for active control of transformer noise with on-line cancellation path modelling based on the perturbation method," J. Sound Vib. **240**(4), 647–665 (2001).

[32] T. K. Yeung and S. F. Yau, "A modified on-line modeling algorithm for the feedforward multiple-point ANC system," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2001), pp. 3205–3208.

[33] R. L. Clark and G. P. Gibbs, "A novel approach to feedforward higher harmonic control," J. Acoust. Soc. Am. **96**, 926–936 (1994).

[34] D. L. Kewley, R. L. Clark, and S. C. Southward, "Feedforward control using the higher-harmonic, time-averaged gradient descent algorithm," J. Acoust. Soc. Am. **97**, 2892–2905 (1995).

[35] K. Fujii, M. Muneyasu, and J. Ohga, "Simultaneous equations method not requiring the secondary path filter," Proc. Active (1999), pp. 941–948.

[36] Y. Maeda and T. Yoshida, "An active noise control without estimation of secondary path—ANC using simultaneous perturbation," Proc. Active (1999), pp. 985–994.

[37] Y. Ohta, T. Kohno, H. Ohmori, and A. Sano, "Direct fully adaptive active noise control algorithms without identification of secondary path dynamics," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2002), pp. 453–458.

[38] Y. Kajikawa and Y. Nomura, "Active noise control without a secondary path model by using a frequency-domain simultaneous perturbation method with variable perturbation," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2003), pp. 580–583.

[39] A. Sano and Y. Ohta, "Adaptive active control without secondary path identification," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (2003), pp. 213–216.

[40] B. J. Kim, "A method for accurate noninvasive on-line secondary path modeling for the filtered-x LMS algorithm for active control of periodic noise," U.S. Patent Application 10/619,801 and Intl. Patent Application PCT/US03/22707 (2003).

[41] D. C. Swanson, *Signal Processing for Intelligent Sensor Systems* (Marcel Dekker, New York, 2000).

[42] J. J. Shynk, "Frequency domain and multirate adaptive filtering," IEEE Signal Process. Mag. **1**, 14–37 (1992).

[43] K. M. Reichard and D. C. Swanson, "Frequency-domain implementation of the filtered-x algorithm with on-line system identification," Proc. Recent Adv. in Active Control of Sound Vib. (1993), pp. 562–573.

[44] S. D. Snyder, "Active control—a bigger microprocessor is not always enough," Noise Control Eng. J. **49**(1), 21–29 (2001).

[45] A. S. Sedra and K. C. Smith, *Microelectronic Circuits* (Saunders College, Philadelphia, PA, 1991).

[46] D. B. Ward and G. W. Elko, "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," IEEE Signal Process. Lett. **6**(5), 106–108 (1999).

[47] S. E. Wright and B. Yuksanovic, "Active control of environmental noise III: Implementation of theory into practice," J. Sound Vib. **220**(3), 469–496 (1999).

[48] J. Yang, S. E. Tan, and W. S. Gan, "Robustness analysis of an environmental active noise control system," J. Sound Vib. **249**(3), 611–621 (2002).

Control strategies for active noise barriers using near-field error sensing

Arthur P. Berkhoff[a]

TNO Science and Industry, PO Box 155, 2600AD Delft, The Netherlands, and University of Twente,
PO Box 217, 7500AE Enschede, The Netherlands

In this paper active noise control strategies for noise barriers are presented which are based on the use of sensors near the noise barrier. Virtual error signals are derived from these near-field sensor signals such that reductions of the far-field sound pressure are obtained with the active system. The performance of the control algorithm is compared for far-field error signals, near-field error signals, and virtual far-field error signals, with and without additional reference sensors. The virtual error signals are obtained by using separate transfer functions for the primary sources and secondary sources. These separate transfer functions are determined in such a way that the necessity of the separation of the virtual sensors for the primary field and for the secondary field can be judged by direct comparison. The systems are evaluated for independent broadband and randomly positioned primary sources, changing source spectra, moving sources, and configurations involving a nonvanishing wind speed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992787]

I. INTRODUCTION

Recently, active noise barriers for reduction of traffic noise have been considered.[1,2] The goal of the active noise barriers is to reduce the sound-pressure level at a distance from the noise barrier, such as on the facades of houses. Microphones at these positions can be used in order to create error signals to be minimized by the controller.[2] In another configuration, microphones were placed directly near the active noise barrier.[1] This leads to obvious advantages due to the absence of sensors at inconvenient positions. Another advantage is that the primary sources are essentially diffraction sources on top of the noise barrier. If the secondary sources are placed near the top of the noise barrier, then the influence of wind is more or less the same for the primary sources and the secondary sources. This can lead to a possibly more robust system with respect to fluctuations of the wind. However, minimization of the near-field sound pressure does not necessarily lead to a reduction of the far-field sound pressure (e.g., Fig. 7 in Ref. 3). Another concern is the possibility to implement the systems in real time. A system where all sensors and loudspeakers of the complete noise barrier are coupled and combined in a single multi-input multioutput controller is clearly impossible. Still another aspect is the influence of moving sources on the performance of the system.[4–6] Considering the above, the main objectives for the design of the active noise barrier are (1) no microphones near buildings should be used but only microphones near the active noise barrier; (2) the system has to be suitable for moving noise sources; (3) the system should consist of modules that work more or less independently by using a limited amount of information from other parts in the system in order to reduce communication bandwidth and hardware

resources. In this paper the first two items will be addressed; only a single module will be considered, which can have multiple inputs and multiple outputs.

An important aspect of the present work is the application of a virtual sensor. Virtual sensors have been discussed in a general control framework by Albertos and Goodwin.[7] A virtual sensor for active noise control has been described by Munn *et al.*[8] An important observation is that for many active control applications the contributions from the primary source have to be separated from the contributions of the secondary source. For the near-field sensor, this separation has been described by Garcia-Bonito *et al.*[9] If the virtual sensor and the near-field error sensor are at very different positions, then this separation should be done for both the near-field sensor and the virtual sensor, as described by Popovich[10] and Roure.[11] The basic idea is to subtract the contribution of the actuator from the near-field sensor, leading to an estimate of the contribution of the primary source on the near-field sensor, and then to predict the primary farfield signal by a suitable transfer function. The estimate of the total far-field signal is then obtained as the sum of the estimated far-field contributions of the primary and secondary sources. Yuan[12] added a second near-field sensor in order to make it less sensitive to transmission zeros at the sensor location in resonating environments.

The present results are for a multi-input, multioutput, multireference system for random, broadband noise. Although the description of Ref. 10 is principally suitable for multichannel systems, no multichannel results are known to the present author. As compared to Ref. 11, the present results give an extension for random, broadband noise instead of harmonic noise. A further difference is that the transfer function between the secondary source and the far-field sensor is split in two parts, in contrast with Refs. 10 and 11, where a single transfer function is used. This was done for two reasons: first, the transfer functions which make an esti-

---
[a]Electronic mail: arthur.berkhoff@tno.nl
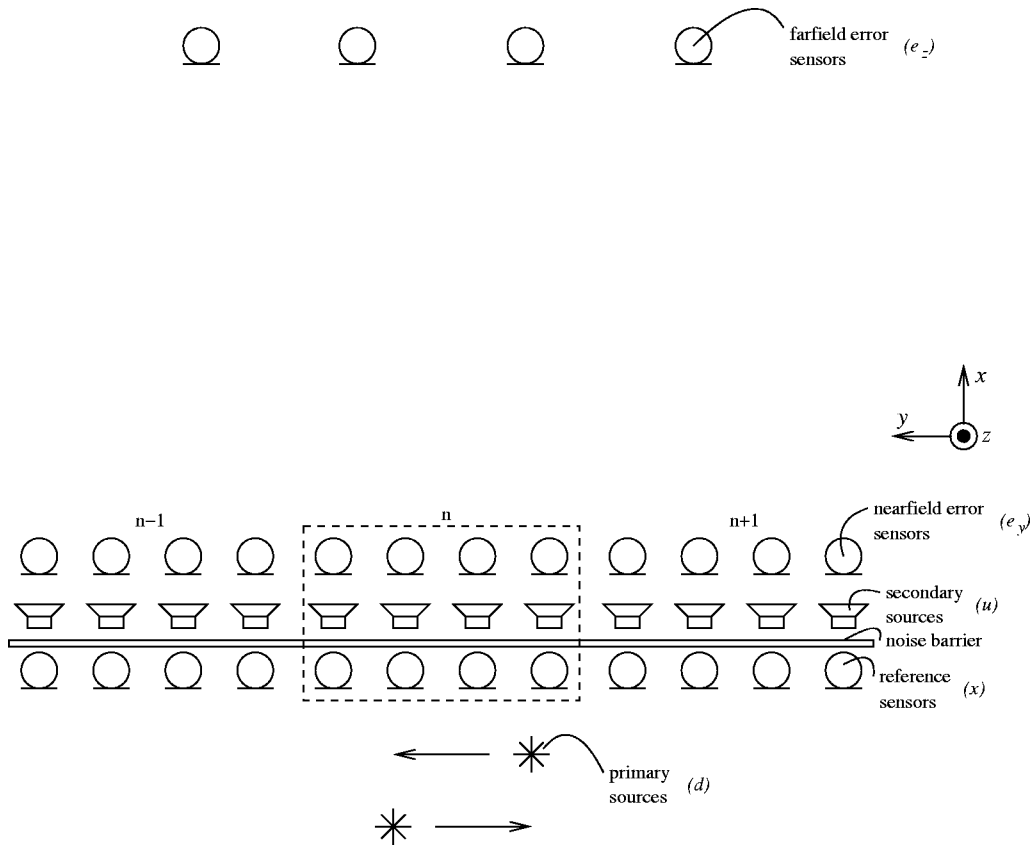
FIG. 1. Configuration of an active noise control system for reduction of noise at the position of the far-field sensors by using information derived from near-field error sensors.

mate of the far-field-from the near-field signals are then treated in a similar way for the primary sources as for the secondary sources. Second, the virtual sensors are based on the assumption that the path from the near field to the far field does not change. It is interesting to have an idea of the differences between the virtual sensors for the primary source and the secondary source. If, for example, these virtual sensors would have some similarity, then we could modify the primary virtual sensor and secondary virtual sensor by the same amount, for example adapting for a changing wind direction or speed. In this work the two multichannel virtual sensors will be compared and some conclusions will be drawn. In order to assess the performance of the proposed virtual sensor scheme, it will be compared with some other active control schemes.

II. DESCRIPTION OF THE SYSTEM

A proposed configuration of the system can be found in Fig. 1. The system consists of reference sensors (microphones), secondary sources (loudspeakers), and near-field error sensors (microphones). In an initial calibration phase far-field error sensors (microphones) also can be used in order to determine the transfer functions for the virtual sensors. A single section n of the active noise barrier is considered (Fig. 1). In order to assess the stability and performance of the active noise barrier, including the interactions between different subsystems, the sensors and sources of adjacent sections also should be taken into account.

A. Reduction of the far-field sound pressure using measurement of near-field quantities

One of the aspects of the system is that virtual error signals are derived from the near-field error sensors. These virtual error signals should represent the far-field pressure at a sufficient number of discrete positions in the angular sector of interest, or more generally, the acoustic degrees of freedom in that angular sector.[13] In the configuration, we distinguish the following components: the primary disturbance d, reference signals x, actuator control signals u, near-field error signals e_y, and far-field error signals e_z such that, in the z-transform domain

$$\begin{bmatrix} x(z) \\ e_y(z) \\ e_z(z) \end{bmatrix} = \begin{bmatrix} G_{dx}(z) & G_{ux}(z) \\ G_{dy}(z) & G_{uy}(z) \\ G_{dz}(z) & G_{uz}(z) \end{bmatrix} \begin{bmatrix} d(z) \\ u(z) \end{bmatrix}. \tag{1}$$

The action of the control operator W is described by

$$u(z) = W(z)x(z). \tag{2}$$

We seek to minimize the cost function

$$J = \|e_z\|_2^2 + \beta\|W\|_2^2, \tag{3}$$

where the term with β is used for regularization by penalizing the magnitude of W. The norm for a general causal transfer function matrix $A(z)$ which, using time τ, corresponds to an impulse response matrix $A(\tau)$, is defined as[14]

1470   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Arthur P. Berkhoff: Near-field sensing for active noise barriers
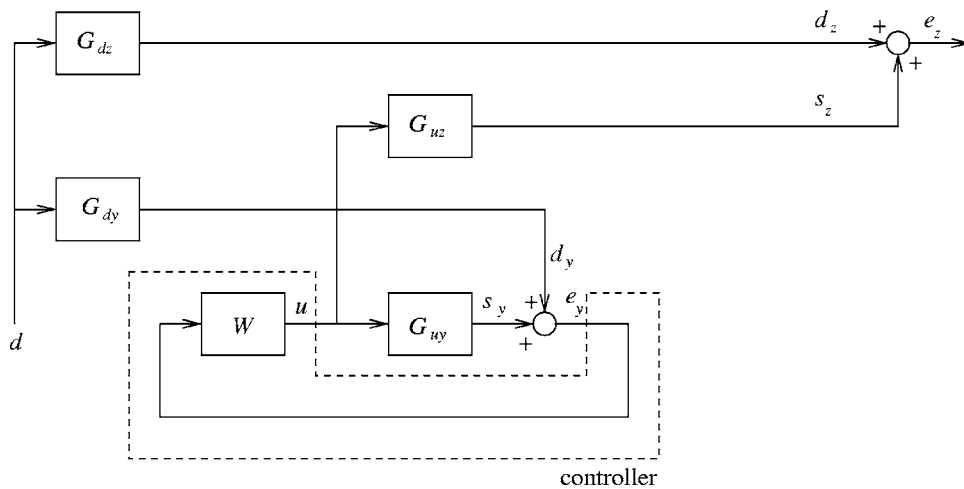
FIG. 2. Schematic representation of a system which directly minimizes e_y in a feedback configuration.

$$\|A\|_2 = \left[\int_0^\infty \mathrm{tr}\, A^T(\tau)A(\tau)\mathrm{d}\tau \right]^{1/2}, \qquad (4)$$

where tr is the trace of a matrix. The norm for a stochastic signal vector $a(\tau)$ is defined as

$$\|a\|_2 = [\mathrm{tr}\, \mathsf{E}\{a^T(\tau)a(\tau)\}]^{1/2}, \qquad (5)$$

where E is the expected value. Lower-case variables denote vectorial quantities, whereas upper-case variables denote matrices. The assumption is that G_{ux} is compensated in the controller, leading to a control structure based on internal model control.[15]

A relatively simple approach for the reduction of far-field sound pressure is to try to reduce the sound pressure near the noise barrier.[1] A disadvantage of this system is that a reduction of the near-field pressure does not necessarily lead to a reduction of the far-field pressure. This holds for both feedback configurations (Fig. 2) and feedforward configurations (Fig. 3). An improved technique could be based on the estimation of the far-field error signals e_z (which are assumed to be unavailable when the system is operational) from the near-field error signals e_y. A rather straightforward approach is to design a single transfer function H_{yz} which makes a real-time estimate of e_z from measured data e_y (see Fig. 4). From the simulations it will be seen that the maximum performance with such a transfer function is limited by

the fact that we are dealing with two sets of sources, the primary sources and the secondary sources. These two sets of sources have a very different nature and lead to rather different requirements for the transfer function H_{yz}. One significant difference, for example, is that the near-field sensors are, located at a distance from the primary sources, while the near-field sensors are, by definition, in the near field of the secondary sources. A compromised H_{yz} designed for the two sets of sources simultaneously did not give satisfactory results because of the relatively large prediction errors. The prediction errors for the compromised design were of similar magnitude as the off-diagonal prediction errors in Table I. So, although a single transfer function H_{yz} may have acceptable performance in some systems, such as active panels,[16] it does not lead to sufficient performance in the present systems for active noise barriers.

Therefore, a new strategy was devised in which the primary sources and the secondary sources were treated separately. The actual contributions on e_y and e_z consist of a part due to the primary signal and a part due to the secondary signal. As the transfer functions from the primary source and secondary source to both e_y and e_z are quite different, these contributions are taken into account separately. Therefore, in accordance with Eq. (1), we define G_{dy} as the transfer function between primary source and near-field error sensor, G_{dz} as the transfer function between primary source and far-field
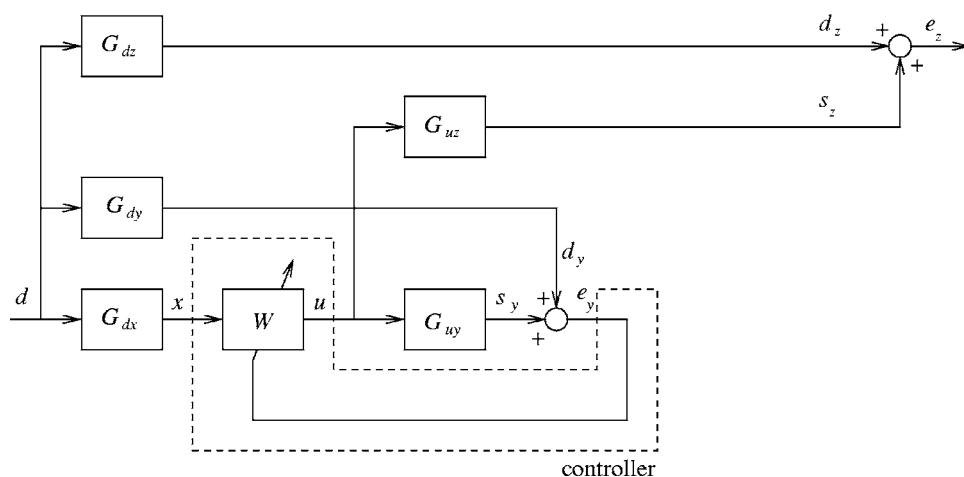
FIG. 3. Schematic representation of a system which directly minimizes e_y in a feedforward configuration.
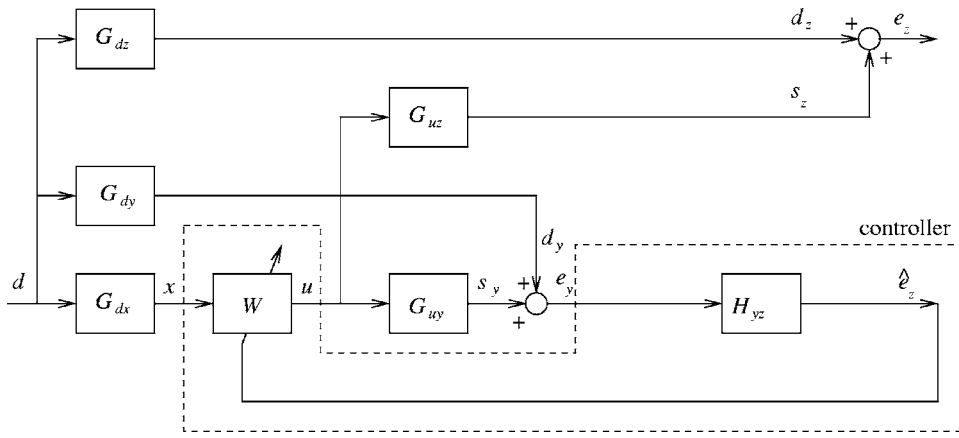
FIG. 4. Schematic representation of a system which makes an estimate $\hat{e}_z$ of the far-field error signals e_z based on a single transfer function H_{yz}.

error sensor, G_{uy} as the transfer function between secondary source and near-field error sensor, and G_{uz} as the transfer function between secondary source and far-field error sensor. The far-field signal e_z can be expressed as

$$e_z(z) = d_z(z) + s_z(z), \tag{6}$$

in which the primary signal d_z is given by

$$d_z(z) = G_{dz}(z)d(z), \tag{7}$$

and the secondary signal s_z is given by

$$s_z(z) = G_{uz}(z)u(z). \tag{8}$$

The estimate $\hat{d}_z$ of the signal d_z is derived from an estimate $\hat{d}_y$ of the primary near-field error signals as follows:

$$\hat{d}_z(z) = H_{yz}^d(z)\hat{d}_y(z), \tag{9}$$

where H_{yz}^d is obtained from system identification. The estimate $\hat{d}_y$ is obtained by subtracting the secondary signal from the measured signal e_y

$$\hat{d}_y(z) = e_y(z) - \hat{s}_y(z), \tag{10}$$

where $\hat{s}_y$ is an estimate of the secondary signal on the near-field error sensors, which is given by

$$\hat{s}_y(z) = \hat{G}_{uy}(z)u(z). \tag{11}$$

An estimate $\hat{s}_z$ of the contribution of the secondary sources on the far-field sensors is given by

$$\hat{s}_z(z) = H_{yz}^s(z)\hat{s}_y(z), \tag{12}$$

where H_{yz}^s is also obtained from system identification. Finally, the estimate $\hat{e}_z$ of the far-field sensor can be obtained from the summation of the estimates of the primary sources and secondary sources

$$\hat{e}_z(z) = \hat{d}_z(z) + \hat{s}_z(z). \tag{13}$$

The estimate $\hat{e}_z$ can be written as

$$\hat{e}_z(z) = H_{yz}^d(z)\hat{d}_y(z) + H_{yz}^s(z)\hat{s}_y(z)$$
$$= H_{yz}^d(z)e_y(z) + [H_{yz}^s(z) - H_{yz}^d(z)]\hat{G}_{uy}(z)u(z). \tag{14}$$

The near-field error signal is given by

$$e_y(z) = d_y(z) + s_y(z). \tag{15}$$

Therefore, using

$$d_y(z) = G_{dy}(z)d(z), \tag{16}$$

and

$$s_y(z) = G_{uy}(z)u(z), \tag{17}$$

the estimate $\hat{e}_z$ can be written as

$$\hat{e}_z(z) = H_{yz}^d(z)[G_{dy}(z)d(z) + G_{uy}(z)u(z)]$$
$$+ [H_{yz}^s(z) - H_{yz}^d(z)]\hat{G}_{uy}(z)u(z). \tag{18}$$

In the case of perfect modeling, we have $G_{uy}(z) = \hat{G}_{uy}(z)$. Then, the above expression reduces to

$$\hat{e}_z(z) = H_{yz}^d(z)G_{dy}(z)d(z) + H_{yz}^s(z)G_{uy}(z)u(z). \tag{19}$$

If the secondary sources are close to the diffraction sources, and if they radiate in a similar way, then we have $H_{yz}^d = H_{yz}^s = H_{yz}$, leading to

$$\hat{e}_z(z) = H_{yz}(z)[G_{dy}(z)d(z) + G_{uy}(z)u(z)] = H_{yz}(z)e_y(z). \tag{20}$$

B. System diagram and comparison with other systems

The complete block diagram of the system is shown in Fig. 5. It can be seen that two separate transfer functions H_{yz}^d and H_{yz}^s are used, which together are able to reconstruct the pressure e_z in the far field. For purposes of comparison, some existing systems will be described. Figure 4 shows a system using a single transfer function, H_{yz}. Figure 3 shows a system where the basic strategy is to minimize the near-field pressure, using a feedforward controller, as before. Figure 2 shows a system in which the near-field pressure is directly

TABLE I. Estimation accuracy of the near-field–far-field predictors H_{yz}^d and H_{yz}^s on far-field sensor 1 for different calibration sources.

Calibration sources	H_{yz}^d, estimation error [dB]	H_{yz}^s, estimation error [dB]
Sources in primary source region	−14.6	10.5
Secondary sources	−2.6	−36.3
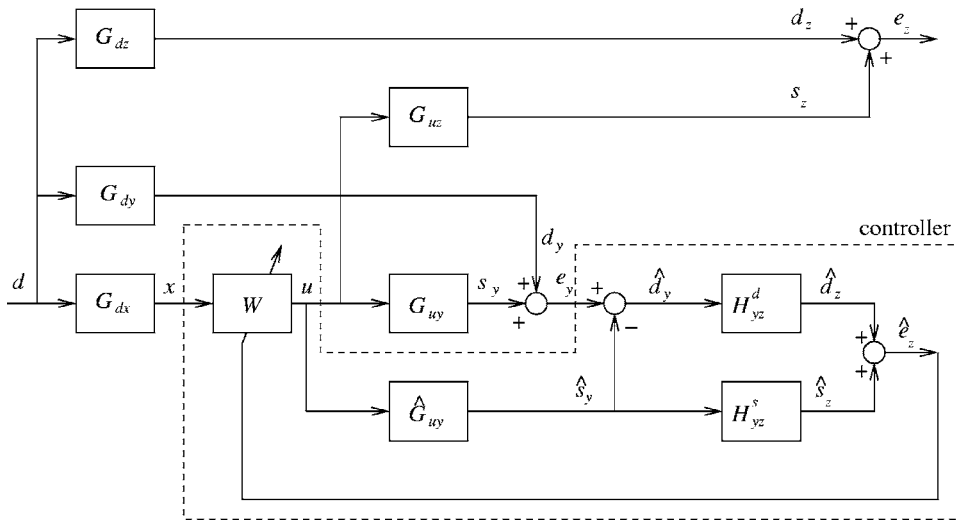
FIG. 5. Schematic representation of a system which makes an estimate $\hat{e}_z$ of the far-field error signals e_z.

fed back to the controller, where the basic strategy is also to minimize the near-field pressure. In the present application, reduction of the near-field pressure is very different from minimizing the far-field pressure. The simulations in subsequent section will show that the configuration of Fig. 5 performs substantially better than the configurations of Figs. 2 and 3. The configuration of Fig. 5 also performs substantially better than the configuration of Fig. 4 because of the reason mentioned in Sec. II A.

Changes in the secondary path are critical for stability robustness and performance robustness. The performance robustness for single-channel active noise control systems including the effects of wind has been analyzed in Ref. 17. In the configuration of the latter paper, the error sensor should be considered to be in the far field of the secondary source due to the relatively large distance of 20 m. The propagation distances involved in the secondary path are considerably reduced if near-field sensors are used instead of the far-field sensors. Therefore, it is expected that the reduced change of the propagation path has a positive effect on the stability robustness of the system. Furthermore, if the primary sources are effectively diffraction sources on top of a noise barrier,

and if the secondary sources are close to these diffraction sources, then we obtain a system in which a large part of the possibly varying transmission path is common to both the primary signals and the secondary signals. This results in an expected improved performance robustness with respect to changes in wind, temperature, etc.

Equation (20) suggests that, if H_{yz} has a sufficiently flat frequency response, minimizing e_y in principle would be sufficient to reduce $\hat{e}_z$. This only holds if the primary sources are coincident with the secondary sources. In practical situations it is very difficult to exactly realize the condition $H_{yz}^d = H_{yz}^s = H_{yz}$. Differences are caused by, for example, the distributed nature of the primary (diffraction) sources, while the secondary sources are concentrated sources. These different behaviors are especially seen on the near-field sensors. Also, the exact position of the primary sources may be difficult to determine. Therefore, for a practical implementation a more general approach is required in which the estimation of the primary path and the estimation of the secondary path are treated separately. For the estimation of the primary transfer functions we define a region with identification sources.
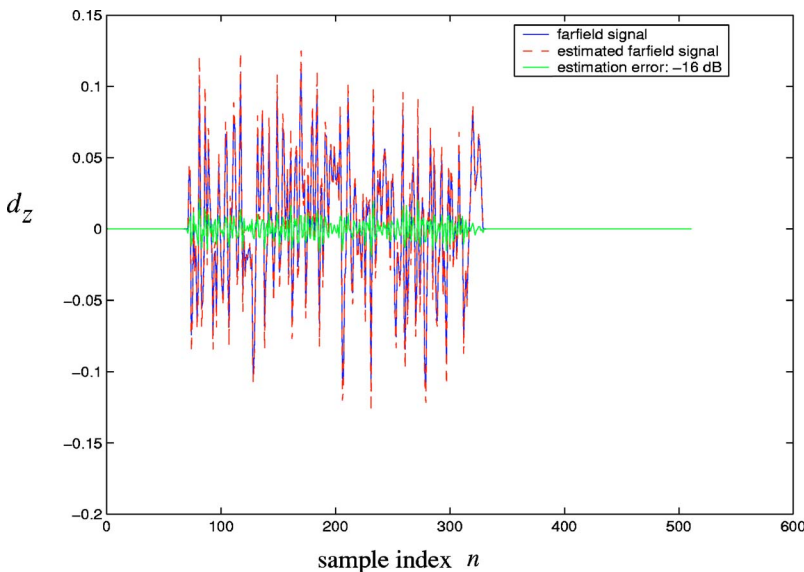
FIG. 6. (Color online) Comparison of a primary far-field signal with an estimate obtained from the near-field sensor signals.
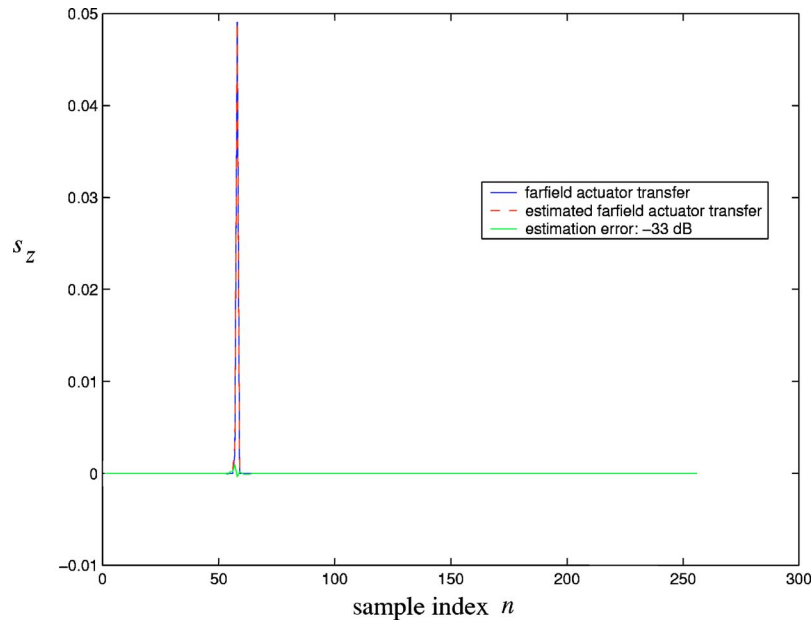
FIG. 7. (Color online) Comparison of a secondary transfer function to the far field with an estimate obtained from the near-field sensor signals.

Furthermore, only those near-field pressure distributions that contribute to the far field, as governed by the transfer functions H_{yz}, should be controlled. If e_y is minimized directly then usually only sound-pressure reductions in the near field are obtained. An additional reduction of the order of H_{yz} can lead to systems that are potentially more robust, have less spillover, and require less actuator effort.

In order to be able to compute the virtual error signals, an initial calibration phase is assumed in which sources are placed at random positions in the region where the primary noise sources are to be expected. The estimate of the primary

signal on one of the far-field sensors is shown in Fig. 6. The estimate of one of the secondary source transfer functions to the far field as estimated from the secondary transfer functions to the near-field sensors is shown in Fig. 7. In the present case 50 calibration sources have been used and three independent primary sources (an example of this can be found in Sec. III A, Fig. 13). The identification procedure is based on the solution of the equations resulting from an assumed multi-input, multioutput finite impulse response model leading to a block-Toeplitz structure.[18] In order to improve the stability of the solution, the mean-square value
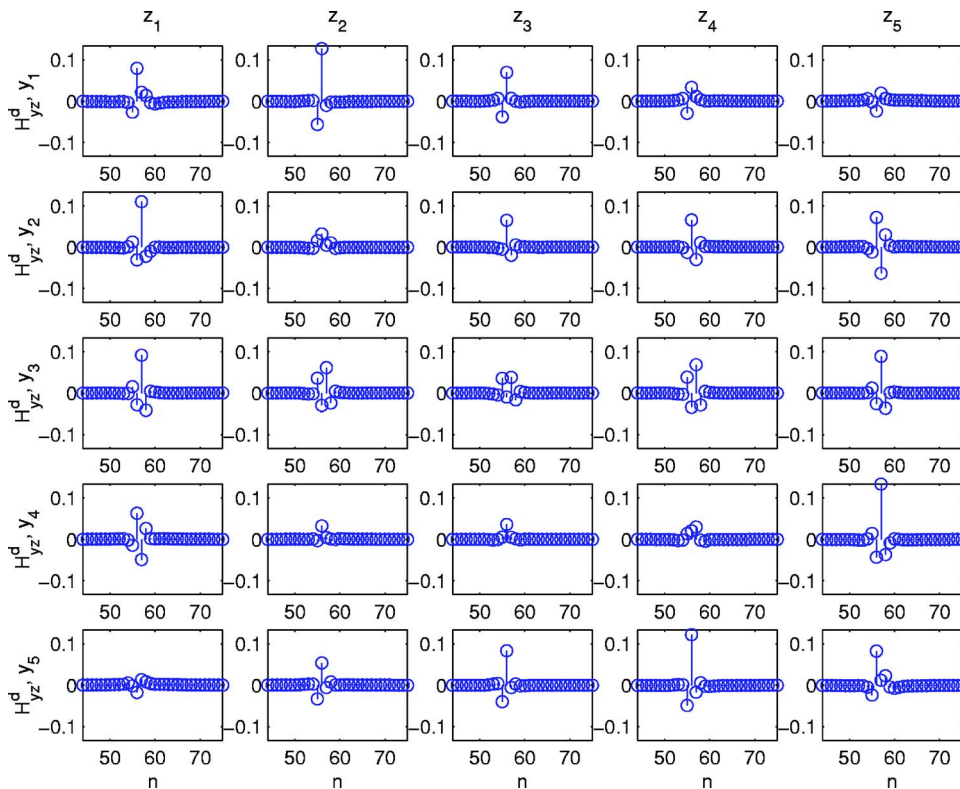
FIG. 8. (Color online) Impulse response matrix H_{yz}^d as a function of sample index n; a row represents the impulse response of a particular near-field sensor and a column denotes the impulse response of a particular farfield sensor.
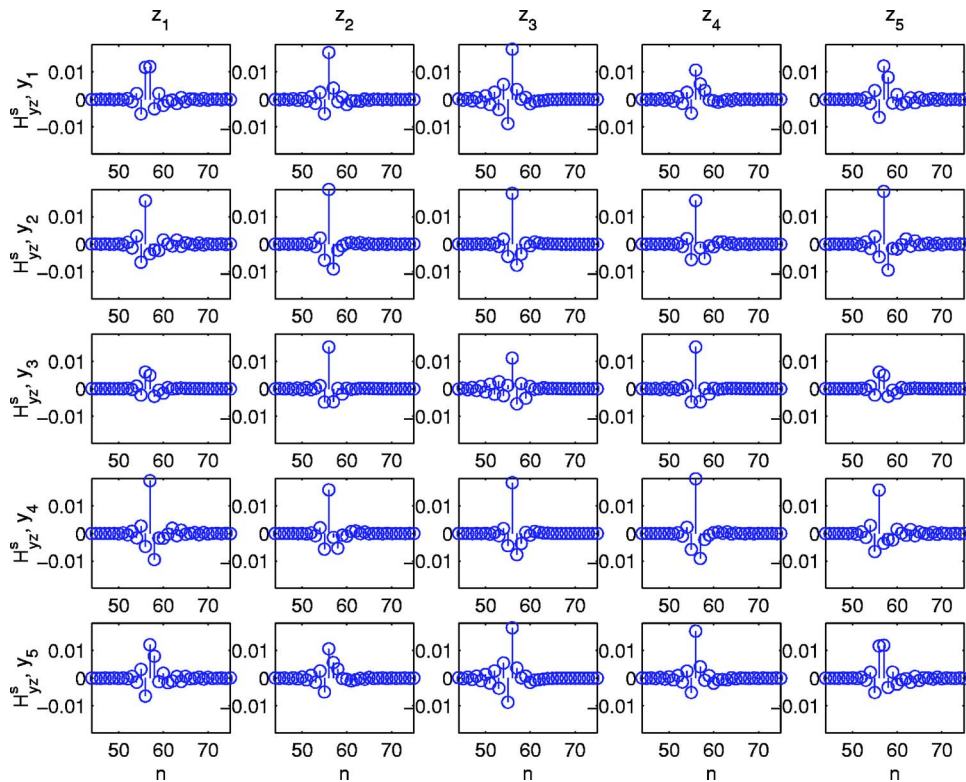
FIG. 9. (Color online) As Fig. 8, except that H_{yz}^s is shown.

of the coefficients of the FIR model are weighted with a normalized coefficient weighting factor $\beta' = 10^{-2}$ as compared to the mean-square prediction error of the far-field pressure. This value of β' was obtained by reducing the value of β' until the prediction error for primary signal validation data did not reduce anymore. The prediction error was obtained with validation data from an independent set of primary sources at random positions in the same region as

FIG. 10. Reduction of the broadband sound pressure for three independent broadband primary noise sources in which the error signal equals the near-field pressure; the primary sources are at $x=-4.5$ m (black asterisks), the reference sensors at $x=-0.5$ m (black circles), the secondary sources at $x=0.5$ m (black asterisks), the near-field error sensors at $x=1$ m (black circles), and evaluation microphones in the far field at $x=20$ m (blue circles).

for the solution of the system of equations. It has been found that the estimation errors did not significantly change with the number of primary sources, provided the primary sources were within the region used for calibration.

Although the transfer functions H_{yz}^d and H_{yz}^s relate the signals from identical near-field sensors and identical far-field sensors, they should be obtained from separate identification procedures, as mentioned previously. The necessity of this is demonstrated in Table I, which gives the estimation results on far-field sensor 1 for different combinations of calibration sources and the two transfer functions, H_{yz}^d and H_{yz}^s. It can be seen that the estimator H_{yz}^d only gives accurate results if the calibration sources are placed in the same region as where the primary sources are to be expected. The same holds for H_{yz}^s, which is only accurate if the calibration sources are taken to be the secondary sources. The differences between H_{yz}^d and H_{yz}^s can be seen in Figs. 8 and 9. One particular difference is that the magnitude of H_{yz}^d is higher than the magnitude of H_{yz}^s. This is caused by the strong near-field behavior and corresponding reactive components of the sound pressure which do not propagate to the far field.

The virtual sensor descriptions are based on a multi-input, multioutput model, and therefore allow for the prediction of far-field signals due to multiple uncorrelated sources in the pertinent source region. However, uncorrelated sources outside the pertinent source region could reduce the accuracy of the far-field estimate. In principle, because of the multi-channel description, the virtual sensors could be designed in such a way that the error of the far-field estimate inside the source region would be minimized while, simultaneously, minimizing the sensitivity for disturbing sources outside that source region. Although numerical tests have shown that this

FIG. 11. As Fig. 10, except that the error signal equals the far-field pressure.

is a useful approach, the design procedure for the virtual sensors in the rest of the paper will be limited to the minimization of the far-field prediction error, either for the sources inside the primary source region, resulting in H_{yz}^d, or for the secondary sources, resulting in H_{yz}^s.

III. RESULTS

Examples will be given in the form of a simulated section of an active noise barrier.

A. Comparison of different configurations

The first example is a purely active noise barrier, as in Ref. 19 (i.e., without passive noise barrier), in which the error signal equals the pressure as measured on microphones near the secondary sources, assuming the control structure of Fig. 3. In the present case the distance (in the x direction) between the error sensors and the secondary sources in 0.5 m. In addition, five reference sensors are used, which are positioned, as seen from the secondary sources, 1 m toward
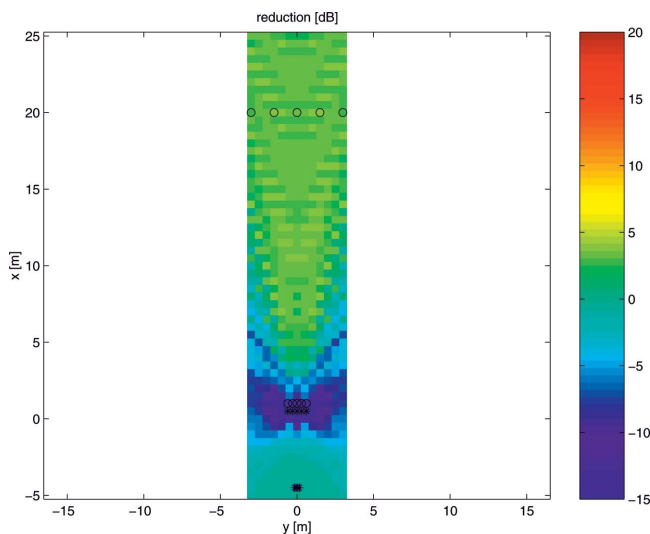
FIG. 12. As Fig. 11, except that no separate reference sensors are used.
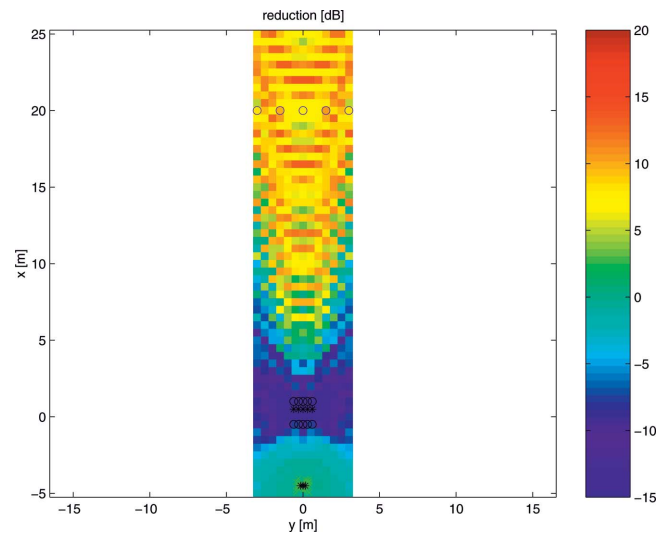
FIG. 13. As Fig. 10, except that the error signal equals the virtual far-field pressure; the sources as used for the identification of H_{yz}^d and H_{yz}^s are indicated by green asterisks.

the primary sources. The primary noise sources are three independent broadband noise sources which are at a distance of 5 m from the secondary sources and at 0.15 m from each other. The sound pressure in the far field is evaluated with five microphones at $x=20$ m. The horizontal distance between these microphones, in the y direction, is 1.5 m. The method for designing the broadband controller of Eq. (2) which minimizes the cost function of Eq. (3) has been described previously.[18] The sampling frequency is $f_s=1$ kHz, the number of controller coefficients for each control filter equals 128, and the normalized regularization parameter β equals 0; the distance between the secondary sources is 0.3 m, as are the distances between the reference sensors and the error sensors. All transducers and sources are at a height $z=2$ m. Figure 10 shows the configuration and the resulting sound pressure in the x,y plane. It can be seen that sound-pressure reductions are mainly obtained near the error sensors. The reduction at the error sensors is 24.2 dB, but the reduction in the far field is much less, being 1.9 dB.

If the error sensors are moved to the far field, *viz.*, to the positions of the evaluation microphones, then the reductions in the far field become much higher. The average reduction of the error signal and consequently the reduction at the evaluation microphones now becomes 11.3 dB, which is limited by the causality constraint. Again, the control structure of Fig. 3 applies, but with the additional simplification $e_y=e_z$. The results are shown in Fig. 11.

TABLE II. Average reductions of the error signals and the far-field pressure signals for different configurations.

Error signal	Error signal, av. reduction [dB]	Far-field pressure, av. reduction [dB]
Near-field pressure (Fig. 10)	24.2	1.9
Far-field pressure (Fig. 11)	11.3	11.3
Far-field pressure, no reference sensors (Fig. 12)	3.5	3.5
Virtual far-field pressure (Fig. 13)	11.6	9.0

TABLE III. Average reductions of the error signals and the far-field pressure signals for three different primary source configurations with the same controller coefficients, using 32 coefficients, showing the influence of modified positions combined with a modified spectrum for the primary sources as well as the influence of a moving source.

Error signal	Error signal, av. reduction [dB]	Far-field pressure, av. reduction [dB]
Nominal primary sources	12.8	10.6
New primary sources, modified positions, modified spectrum	11.0	11.4
Moving source	9.4	11.0

In the third simulation the effect of a removal of separate reference sensors is demonstrated. In order to demonstrate this, the reference sensors are now shifted to the position of the near-field error microphones; the error signal equals the far-field pressure. The result is shown in Fig. 12. It can be seen that the reduction of the error signal (3.5 dB) and consequently the reduction of the sound-pressure level in the far field (3.5 dB) is considerably less than in Fig. 11. Hence, the use of separate reference sensors giving time-advanced information of the noise sources is necessary if significant noise reductions are to be obtained.

In the last simulation we show the result in case the error signal equals virtual far-field signals, using the same microphone positions as in Fig. 10 and using the control scheme of Fig. 5. The transfer functions H_{yz}^d and H_{yz}^s, as well as $\hat{G}_{uy}$, are determined from off-line system identification. The sources as used in the identification phase are positioned in the region where the primary noise sources are to be expected (see Fig. 13). It can be seen that the resulting far-field sound-pressure reductions are slightly less than in Fig. 11, being 9.0 dB. However, the far-field sound-pressure reductions are considerably higher than obtained with the near-field error

sensors as used in Fig. 10. The reduction of the error signals is approximately equal in case of minimizing the virtual error signals, being 11.6 dB, and in case of minimizing the true far-field error signals (11.3 dB). The numerical results for these four cases are collected in Table II. It can be seen that the sound-pressure reductions have a ripple in the x direction. It has been found that this ripple can almost be eliminated while keeping the same noise reductions by a proper regularization, e.g., by increasing β from 0 to 0.01.

B. Robustness

Robustness of the controllers was evaluated for the following cases: for a different order of the controller, for a new set of primary sources in the same region as for the nominal case, for a change of the primary spectrum as compared to the nominal case, and for a moving source. Robustness was also checked for situations with wind, which is described in Sec. III C.

1. Change of primary source spectrum

The influence of a simultaneous change of the positions of the primary sources and a changing primary spectrum is shown in Table III. Here, the distance between the secondary sources is 0.5 m, as are the distances between the reference sensors and the error sensors. The modified primary spectrum was obtained by applying a fourth-order bandpass filter with Butterworth characteristic with lower and upper cutoff frequencies of 0.15 and 0.35 f_s, respectively, to the primary source signals. It can be seen that, with the modified primary source positions and spectrum, there is even a slight improvement of the performance on the far-field evaluation microphones from 10.6 to 11.4 dB.
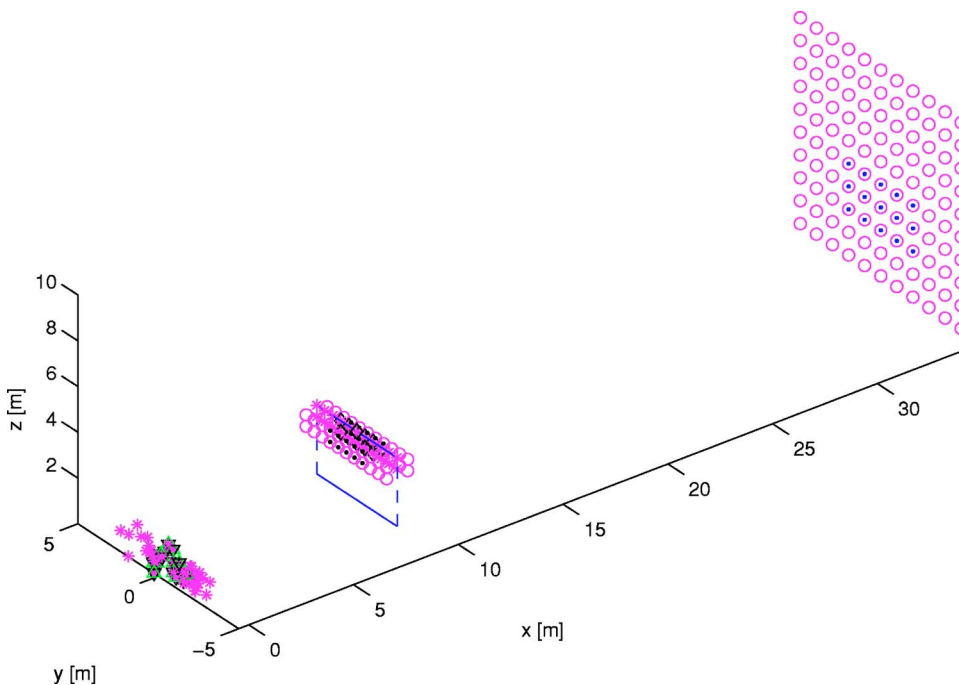
FIG. 14. (Color online) Configuration for the simulation of the performance of a combined passive and active noise barrier; the passive noise barrier is positioned at $x=10$ m and has a height of 3 m; the active system consists of double rows of reference microphones (dots at $x=9.5$ m, $z=2$ m, and $z=3$ m), double rows of loudspeakers (diamonds at $x=10$ m, $z=2.5$ m, and $z=3.5$ m) and double rows of near-field error microphones (dots at $x=10.5$ m, $z=2$ m, and $z=3$ m). Of the total of 10×11 microphone positions in the far field at $x=35$ m, 3×5 positions are used for the virtual sensors (indicated by dots). The primary sources are randomly positioned in the region -1.5 m$\leq y\leq1.5$ m, $0\leq z \leq1$ m $-0.5\leq x\leq0.5$ m (indicated by triangles).
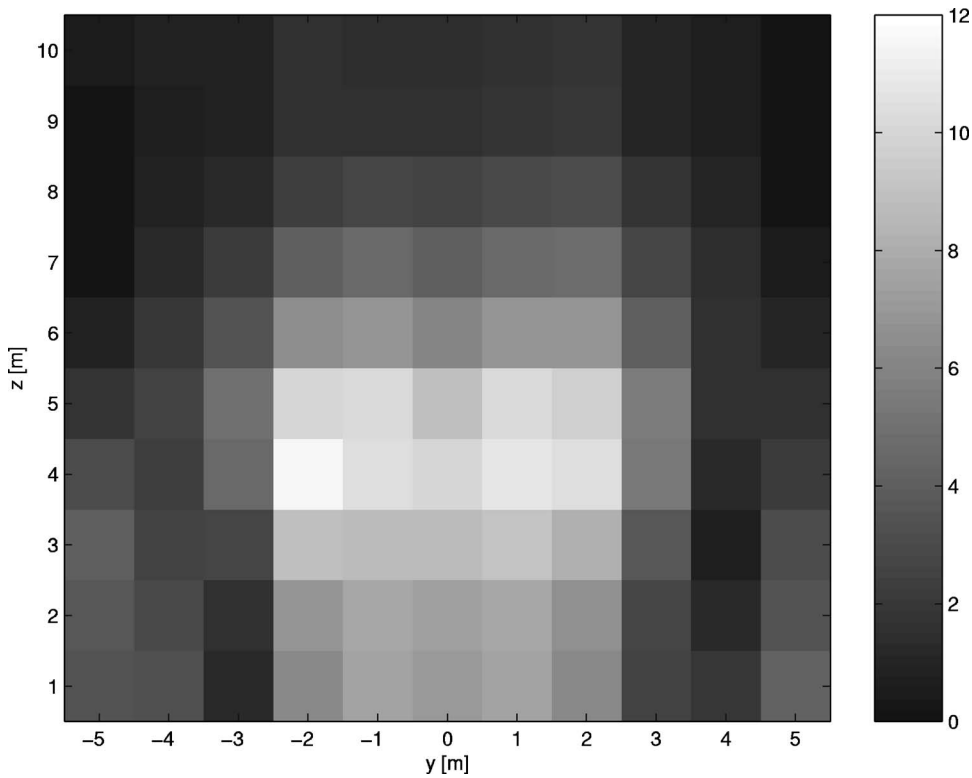
FIG. 15. Noise reduction in dB(A) at the 10×11 microphones at $x=35$ m (Fig. 14), without wind.

2. Moving sources

In this case a single source emitting random broadband noise was moving from $x=1$ m to $x=-0.91$ m during a simulation period of 512 samples. The performance for this moving source can also be found in Table III. The same controller coefficients were used as for the other two cases in Table III. The radiation characteristics of the moving sources were simulated by taking into account the Doppler shift as well as the changing radiation characteristics depending on the direction of the movement.[20]

C. Influence of wind for a simulated noise barrier

In a subsequent simulation study, several features were added in order to be able to analyze somewhat more realistic
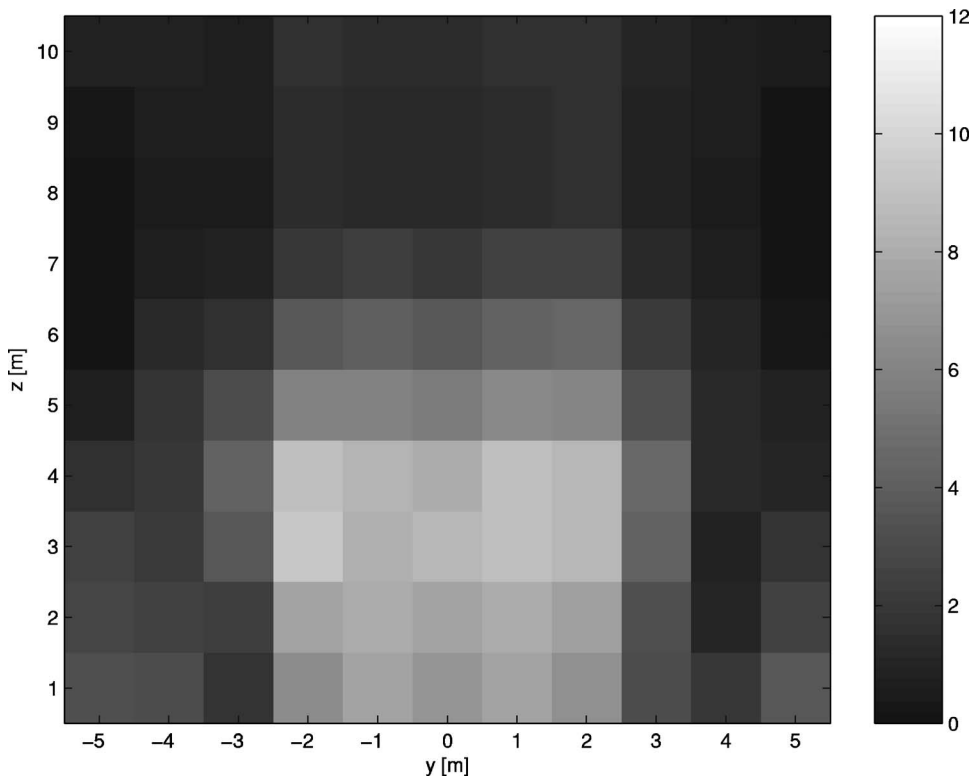
FIG. 16. Noise reduction in dB(A) at the 10×11 microphones at $x=35$ m (Fig. 14), with wind of 5m/s in the x direction.

situations involving a passive noise barrier having a height of 3 m. The components of the active reduction system were placed near the top of the noise barrier (Fig. 14). In this case double rows were used for the reference microphones, the loudspeakers, and the near-field microphones. This was done to have more control in the height direction z, leading to effective noise reductions over extended heights. The identification procedure was as before, where the calibration sources in the primary source region were statistically independent from the real primary noise sources; they also had different positions. The simulation of the passive noise barrier was based on a ray-tracing model which includes the effects of wind[21] for $x \geq 10$ m. Also, the transfer functions from sources to microphones were based on this ray-tracing model. The effect of the wind can be seen in the transfer functions by shorter impulse responses due to the reduced propagation speed (not shown). The source spectrum was obtained from measured traffic noise, where the traffic consisted of 30 percent trucks and 70 percent cars. The sampling frequency was 3.6 kHz. The performance for the case of no wind can be seen in Fig. 15. In view of Fig. 14, it can be seen that the noise reduction is maximum for the region where the virtual microphones are positioned. If a wind of 5 m/s in the positive-x direction was used, then the performance of the system, using the same transfer functions H_{yz}^d and H_{yz}^s as with no wind, is as shown in Fig. 16. It can be seen that the noise reductions at the virtual microphones in Fig. 16 are somewhat lower as compared to Fig. 15, but still significant. So, although the virtual sensors for the primary sources and the secondary sources are very different (see, for example, Figs. 8 and 9), the system seems to be quite robust with respect to changes in the transfer function between the noise barrier and the far-field microphones. In this paper only wind directions perpendicular to the noise barrier were taken into account. Subsequent robustness studies should also consider wind directions parallel to the noise barrier.

IV. CONCLUSIONS

Simulations have shown that additional reference sensors at the incidence side of an active noise barrier are required if large sound-pressure reductions for broadband noise are to be obtained at relatively large distances from the noise barrier. Furthermore, the control system should be based on the minimization of the far-field sound pressure, which can be either a measured far-field sound pressure or a virtual far-field sound pressure. If virtual sensors are used then the contributions of the primary sources and the secondary sources should be treated separately. Finally, although the virtual sensors for the primary sources and the secondary sources are very different, a system in which all transducers are positioned near a noise barrier seems to be relatively robust with respect to changes in the propagation characteristics between noise barrier and the region where reductions are to be obtained.

[1] K. Ohnishi, S. Teranishi, M. Nishimura, K. Uesaka, K. Hachimine, and H. Ohnishi, "Development of the noise barrier using active controlled acoustical soft edge," in *Proc. Active 99*, edited by S. Douglas (INCE USA, Washington, DC, 1999), pp. 595–606.

[2] D. Duhamel, P. Sergent, C. Hua, and D. Cintra, "Measurement of active control efficiency around noise barriers," Appl. Acoust. **55**, 217–241 (1998).

[3] J. Guo and J. Pan, "Analysis of active noise control in a free space," in *Proc. Active 95*, edited by S. D. Sommerfeldt and H. Hamada (Noise Control Foundation, New York, 1995), pp. 649–660.

[4] V. Martin, "Active control of moving sound source radiation—Numerical modelling in the space-frequency and space-time domains," J. Sound Vib. **228**, 477–509 (1999).

[5] J. Guo, J. Pan, and M. Hodgson, "Active control of a moving noise source—Effect of off-axis source position," J. Sound Vib. **251**, 457–475 (2002).

[6] A. Omoto, D. Morie, and K. Fujiwara, "Behavior of adaptive algorithms in active noise control systems with moving noise sources," Acoust. Sci. and Tech. **23**, 85–89 (2002).

[7] P. Albertos and G. C. Goodwin, "Virtual sensors for control applications," Annu. Rev. Control **26**, 101–112 (2002).

[8] J. M. Munn, C. D. Kestell, B. S. Cazzolato, and C. H. Hansen, "Real time feedforward active noise control using virtual sensors," in *Proc. Internoise 2001*, edited by R. Boone [Nederlands Akoestisch Genootschap (NAG), Maastricht, 2001], pp. 781–784.

[9] J. Garcia-Bonito, S. J. Elliott, and C. C. Boucher, "Generation of zones of quiet using a virtual microphone arrangement," J. Acoust. Soc. Am. **101**, 3498–3516 (1997).

[10] S. R. Popovich, "Active acoustic control in remote regions," Patent No. US5701350, Appl. No. 657295, assigned to Digisonix.

[11] A. Roure and A. Albarrazin, "The remote microphone technique for active noise control," in *Active 99*, edited by S. Douglas (INCE USA, Washington, DC, 1999), pp. 1233–1244.

[12] J. Yuan, "Virtual sensing for broadband noise control in a lightly damped enclosure," J. Acoust. Soc. Am. **116**, 934–941 (2004).

[13] G. V. Borgiotti, "The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements," J. Acoust. Soc. Am. **88**, 1884–1893 (1990).

[14] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control* (Wiley, Chichester, 1996).

[15] M. Morari and E. Zafiriou, *Robust Process Control* (Prentice-Hall, London, 1989).

[16] A. P. Berkhoff, "Weight reduction and transmission loss tradeoffs for active/passive panels with miniaturized electronics," in *Proc. Active 04*, edited by R. H. Cabell and G. C. Maling (INCE USA, Washington DC, 2004), pp. 1–12.

[17] A. Nakashima and M. Hodgson, "Effect of realistic grounds and atmospheric conditions on single-channel active control of outdoor sound propagation," J. Acoust. Soc. Am. **117**, 1080–1087 (2005).

[18] A. P. Berkhoff, "Broadband radiation modes: Estimation and active control," J. Acoust. Soc. Am. **111**, 1295–1305 (2002).

[19] S. E. Wright and B. Vuksanovic, "Active control of environmental noise II. Noncompact source regions," J. Sound Vib. **202**, 313–359 (1997).

[20] A. P. Dowling and J. E. F. Williams, *Sound and Sources of Sound* (Ellis Horwood, Chichester, 1983).

[21] E. M. Salomons, "Reduction of the performance of a noise screen due to screen-induced wind-speed gradients. Numerical computations and wind-tunnel experiments," J. Acoust. Soc. Am. **105**, 2287–2293 (1999).

The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay

David T. Bradley[a] and Lily M. Wang

University of Nebraska-Lincoln, Architectural Engineering Program, Peter Kiewit Institute, Omaha, Nebraska 68182-0681

This project focuses on the individual and interactive acoustic effects of three architectural parameters on the double slope profile from a simple coupled volume system created in the computer modeling program ODEON. The three variables studied are the volume ratio between the main and secondary spaces, the absorption ratio between the two spaces, and aperture size. The resulting energy decay profiles are analyzed using T_{30}/T_{15} coupling coefficient ratios and Bayesian analysis. Coupling coefficient results show general trends in the effects of the three architectural parameters that match previous research results and the predominant interactive effect between the three variables involving a large coupled volume. Similar results are suggested by the Bayesian analysis based on the newly developed quantifiers, decay ratio and ΔdB, although a more complex relationship among the variables may exist at larger volume ratios. A subset of the simulated sound fields have been auralized and used to conduct psychoacoustic testing. The subjective testing results indicate that perceived reverberation increases with greater coupled volume size and aperture size, correlating well with objective results. The outcomes also suggest that higher perceived reverberation coincides with larger decay ratios and smaller ΔdB values. Subjective results based on clarity showed no significant effects. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1984892]

I. INTRODUCTION

Coupled volume geometries have been used in several modern concert halls, including Festival Hall in Tampa, FL; the Great Hall in Hamilton, Ontario; Lucerne Concert Hall in Lucerne, Switzerland; the Myerson-McDermott Hall in Dallas, TX; and Verizon Hall in Philadelphia, PA. Architectural acousticians have been particularly interested in halls that utilize coupled volume systems because of their potential for creating a nonexponential energy decay. This paper focuses on the effort to articulate more clearly the effect of certain architectural parameters on the nonexponential decay in coupled volume spaces, both from objective and subjective viewpoints. Objectively the authors explore new ways of quantifying the phenomenon of nonexponential decay, while subjective psychoacoustic tests have been conducted to learn more about its perception.

A system utilizing coupled volumes typically consists of two or more spaces that are connected through an acoustically transparent opening known as a coupling aperture. When the secondary space exhibits a decay time that is longer than that of the main space, sound energy will be fed back into the main space by the auxiliary volume. This late-arriving energy results in the phenomenon known as the double slope decay.

A coupled volume system consisting of two spaces is shown in Fig. 1, where V_i is the volume of the ith space, A_i is the total surface area of the ith space, S is the surface area of the coupling aperture separating the two spaces, and P is the power of the sound source in room 1. Also, α_i is defined as the average absorption coefficient of the ith space, c as the speed of sound, and E_i as the energy density in the ith space. The power balance equations for the two rooms can be written as

$$P - \frac{A_1\alpha_1 cE_1}{4} - \frac{ScE_1}{4} + \frac{ScE_2}{4} = 0 \qquad (1)$$

$$-\frac{A_2\alpha_2 cE_2}{4} - \frac{ScE_2}{4} + \frac{ScE_1}{4} = 0. \qquad (2)$$

These equations give us the steady state response of the system. By setting $P=0$, we can obtain the differential equations representing the reverberant sound decay in the two spaces:

$$\frac{c}{4}(A_{1S}E_1 - SE_2) = -V_1\frac{dE_1}{dt}, \qquad (3)$$

$$\frac{c}{4}(-SE_1 + A_{2S}E_2) = -V_2\frac{dE_2}{dt}, \qquad (4)$$

where $A_{iS}=A_i\alpha_i+S$, and exponential decay is assumed such that

$$E_i = E_{i0}e^{-2\delta t}, \qquad (5)$$

where E_{i0} is the initial energy density in the ith space and δ is the decay constant of the space. By substituting Eq. (5) into Eqs. (3) and (4), and simplifying, we can rewrite the system of equations in matrix form:

[a]Electronic mail: dbradley@mail.unomaha.edu
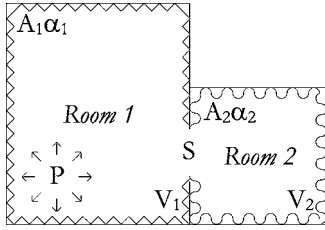
FIG. 1. Plan view of a coupled volume system used in theoretical development.

$$\begin{bmatrix} \left[\left(\frac{c}{4}\right)A_{1S} - 2\delta V_1\right] & -\left(\frac{c}{4}\right)S \\ -\left(\frac{c}{4}\right)S & \left[\left(\frac{c}{4}\right)A_{2S} - 2\delta V_2\right] \end{bmatrix} \cdot \begin{bmatrix} E_{10} \\ E_{20} \end{bmatrix} = 0. \quad (6)$$

This system can only be valid for a single value of δ if the ratio E_{10}/E_{20} is the same for both equations, which requires the determinant of the coefficients of E_{10} and E_{20} to equal zero, producing the following polynomial equation:

$$4V_1V_2\delta^2 - \left(\frac{c}{2}\right)(A_{1S}V_2 + A_{2S}V_1)\delta + \left(\frac{c^2}{16}\right)(A_{1S}A_{2S} - S) = 0. \quad (7)$$

The eigenvalues of this quadratic equation give the decay constants for each room, δ_1 and δ_2, as a function of the volumes, surface areas, and absorptions of the two rooms, and the speed of sound. The decay time (T_i) for room i alone can then be calculated from δ_i using

$$T_i = \frac{6.9}{\delta_i} \quad (8)$$

in S.I. units.

To visualize the reverberant process as a temporal phenomenon, it is helpful to plot the sound level as a function of time. Cremer et al.[1] give the relationship between T_i, E_{i0}, and a reference sound energy density, E_{ref}, as

$$L_i(t) = -\left(\frac{60}{T_i}\right)t + 10 \log_{10}\left(\frac{E_{i0}}{E_{ref}}\right), \quad (9)$$

where $L_i(t)$ is the sound pressure level in the ith space. The early and late temporal components of the decay in the first room can be derived from this equation. For the case where $\delta_1 > \delta_2$, the early portion is dominated by the autonomous effects of room 1 and the late portion is driven by the effects of room 2 on room 1, given by the following:

$$L_{1early}(t) = -\left(\frac{60}{T_1}\right)t + 10 \log\left(\frac{E_{10}}{E_{ref}}\right), \quad (10)$$

$$L_{1late}(t) = -\left(\frac{60}{T_2}\right)t + 10 \log\left(\frac{E_{21}}{E_{ref}}\right), \quad (11)$$

where the initial energy density in room 1 caused by the energy transferred from room 2 to room 1 is given as

$$E_{21} = k_1 k_2 E_{10} \quad (12)$$

and k_1 and k_2 are coupling factors characterizing the transfer of energy between the rooms based on their geometry and absorption, defined as
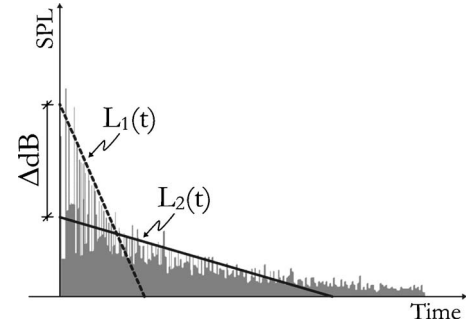
FIG. 2. Representative energy decay curve of a space exhibiting a double slope profile.

$$k_1 = \frac{S}{A_{1S}}, \quad k_2 = \frac{S}{A_{2S}}. \quad (13)$$

The term E_{ref} in Eqs. (10) and (11) may be chosen to be equal to E_{10}, thereby normalizing the sound level in room 1, $L_1(t)=0$ dB at $t=0$. Then substitution of Eqs. (12) and (13) into Eq. (11) and further simplification yields the following early and late decay equations for room 1:

$$L_{1early}(t) = -\left(\frac{60}{T_1}\right)t, \quad (14)$$

$$L_{1late}(t) = -\left(\frac{60}{T_2}\right)t - 10 \log\left(\frac{A_{1S}A_{2S}}{S^2}\right). \quad (15)$$

These equations are in $y=-mx+b$ form, from which we can see that the early and late portions of the decay are two lines, each with a slope and y intercept that can be calculated from geometrical parameters of the coupled volumes. The above-mentioned theoretical development is a modified version of that given in Cremer et al.[1] For a more detailed description of this analysis, the reader is referred to Refs. 1–5.

The above-mentioned results allow for the decay slope of the main room in a coupled volume system to be analyzed based on its two temporal components, early decay and late decay. Many room acousticians believe that a steep early decay slope allows for a high level of perceived clarity, whereas a shallow late decay slope leads to a more lingering reverberance.[4] Clarity describes the definition of a sound in a space, or how easily different parts of a sound signal can be differentiated from one another. Reverberance describes the fullness of tone, and refers to sound energy that persists in a room after a sound is suddenly stopped. Both clarity and reverberance are desired quantities in concert hall acoustics, but seem to have contradictory decay slope criteria. This apparent contradiction may be alleviated in coupled volume systems because $L_1(t)$ and $L_2(t)$, which can exhibit differing slope characteristics, dominate the decay at different times. Figure 2 shows a representative decay outline for this situation, where $L_1(t)$ is shown by the dashed line and $L_2(t)$ by the solid line.

This so-called double slope decay, or double slope effect (DSE), shown by the bold lines in Fig. 2, may result in high clarity and reverberation. The term DSE will be used in this paper to refer to the phenomenon in which the secondary decay is longer than the first, and in which it dominates the

decay profile during the latter portion of time. Although coupled volume systems can be designed to exhibit a variety of acoustic effects, both with and without DSE, this work will focus on those which incorporate auxiliary volumes external to a main space connected through apertures specifically designed to obtain DSE. The authors refer to such systems as dedicated coupled volume systems.

II. PREVIOUS RESEARCH

As can be seen from the above-noted theoretical development, the DSE in a coupled volume system is affected by several variables, namely main volume size, coupled volume size, aperture size, and level of absorption in the two spaces. The effects of these variables, as well as source-receiver positions and the relationship between theoretical and computational analyses of coupled systems, have been explored to a limited degree in previous studies.

Eyring conducted a study of several coupled room configurations in an attempt to develop empirical modifications to theoretical formulas.[6] Eyring's configurations examined three distinct variables: absorption in both rooms, aperture size, and the source-receiver positions. Specifically, Eyring found that the double slope effect is most noticeable when a large absorptive room is coupled with a smaller live room. Additionally, for his particular configurations, Eyring posited that the coupled volume effect occurs only for absorption coefficients of less than 0.07 in the coupled room. Finally, he suggested that DSE only occurs when the source and receiver are placed in the same room, particularly, the less reverberant room. This work began to shed light on the relationship between the architectural parameters of the system and the DSE; however, more systematic analysis of these variables was needed.

Harrison and Madaras utilized computer modeling of coupled volumes to study the effects of aperture size and coupled volume size.[7] They quantified the coupled volume effect with a ratio of T_{30}/T_{15}, referred to as the "coupling coefficient." This ratio gives the relationship between two different portions of the reverberant decay. T_{30} is defined as the decay time from -5 to -35 dB in the energy decay function, multiplied by a factor of 2, such that T_{30} is on the same order of magnitude as T_{60}. Similarly, T_{15} is given as the decay time from -5 to -20 dB, multiplied by a factor of 4. Although T_{30}/T_{15} is an adequate indicator of DSE, it does not rigorously differentiate between different double slope profiles. Particularly, a specific coupling coefficient can be used to describe several different double slope profiles. This ambiguity compromises the effectiveness of this descriptor as a metric of the DSE.

Harrison and Madaras showed that T_{30}/T_{15} is maximized for smaller aperture opening sizes and generally increases exponentially with increasing coupled volume sizes. The researchers also pointed out that the shape, location, and configuration of the aperture openings had noticeable, but less significant effects on the acoustic performance than did the volume of the coupled space and aperture size.

Ermann and Johnson studied how the architectural variables, aperture size, and coupled volume absorption affect the decay slope of a coupled system.[8] To quantify the coupled volume effect, the investigators used the ratio of T_{60}/T_{15}, referred to as the "coupling constant," which is a slight variation on Harrison and Madaras' coupling coefficient. This ratio has problematic issues, similar to the coupling coefficient, namely that it is not a rigorous measure of DSE. Additionally, this method obtains a T_{60} value from a straight line fit, which is a poor representation of the actual double sloped decay function and could lead to great inaccuracies. Thus, coupling constant is only a rough indicator of DSE.

The Ermann and Johnson study showed that having an average absorption coefficient in the coupled space less than 0.02 ($\bar{\alpha} < 0.02$) produced significantly higher coupling constants than $0.02 < \bar{\alpha} < 0.05$. The coupling constant approached infinity as $\bar{\alpha}$ decreased to zero, and asymptotically approached unity for $\bar{\alpha} \geqslant 0.10$. The results regarding the effect of aperture size indicate that the coupling constant peaks when the aperture size is 1% of the total main volume surface area, while the coupled volume effect declines dramatically for aperture sizes larger than 1% of the total main volume surface area.

Summers extended the earlier theoretical work into the computational domain using CATT Acoustic computer models of coupled spaces.[9] His work analyzed the limits of statistical models by comparing them to computer modeled results, subsequently making refinements to the basic formulas of statistical coupling. Results showed that high levels of coupling caused large deviations between statistical predictions and computational values. Additionally, the results showed that all assumptions for single rooms, such as a diffuse sound field with even distribution of absorption, must hold true in the coupled system so that theoretical and computational methods may agree.

Research on subjective perception of nonexponential decays has also been undertaken. Atal *et al.* conducted an investigation to determine the correlation between the subjective feeling of reverberation and nonexponential decays.[10] In this study, nonexponential decays were produced by manipulating reverberated signals obtained from computer-simulated reverberators using combfilters connected in parallel. The decays exhibited a flat response across frequency, indicating that each normal mode had an equal decay rate. The researchers carried out pairwise comparison subjective tests between classic exponential decays and the artificially generated nonexponential decays. The reverberation time corresponding to the exponential decay chosen as most similar to a particular nonexponential decay was designated as the subjective reverberation (T_s). The investigators found that this T_s was highly correlated with a new reverberation time metric $T(160 \text{ ms})$. $T(160 \text{ ms})$ was found by fitting a straight line through the first 160 ms of the nonexponential decay and extrapolating to 60 dB down. This result indicates that the most important factor in determining the subjective evaluation of reverberation in nonexponential decays is the early temporal content of the decay. Further investigation, though, found that introducing the complexity of frequency dependence found in typical concert halls confounded the

correlation between the subjective impression of reverberation and the early portion of non-exponential decay.

Picard also constructed a series of artificial nonexponential decays by splicing together pairs of impulse responses exhibiting exponential decay profiles.[11] Subjective testing was conducted in an effort to determine just noticeable differences between exponential and nonexponential decays. The results showed that subjects could more easily perceive a difference between the decays when the y intercepts of the two slopes were minimized. Additionally, difference perception increased as the difference between the two slopes in the nonexponential decays was increased. In the study, Picard suggested a quantitative descriptor for nonexponential decay based on the amount of energy from the latter secondary decay that is added above the original exponential decay.

Ermann conducted a similar subjective perception study, using computational geometric acoustics modeling to generate exponential and nonexponential decays.[12] The results from subjective tests utilizing these decays corroborated those found by Picard, showing that subjects are more likely to recognize a difference between two decay curves when the difference between the first and second slopes in the nonexponential decay is increased.

The research in this paper uses room acoustic computer modeling techniques to extend the work on quantifying the effect of coupled room geometry by more comprehensively analyzing the individual and interactive effects of three architectural parameters found to be significant in previous work: the volume ratio between the main and secondary spaces, the absorption ratio between the two spaces, and the aperture size. In this paper, the coupling coefficient is used to compare against previous research, and new DSE quantifiers are additionally introduced to differentiate the nonexponential decays with greater precision. The study then continues to extend the previous work by correlating subjective responses with the coupled geometry itself. Together, the objective and subjective analyses help to provide a better understanding of how architectural parameters affect DSE in coupled volume systems.

III. DEVELOPMENT OF NEW DSE QUANTIFIERS

A. Decay ratio and ΔdB

As discussed previously, the coupling coefficient obtained from the ratio of T_{30}/T_{15} does not uniquely represent the double slope effect apparent in an energy decay curve associated with a coupled volume system, and thus may not accurately indicate its perception. A more accurate method of characterizing the DSE may be achieved by analyzing the decay as a composite of two distinct slopes. The energy decay curve shown in Fig. 2 will be used as an illustrative example. The steep slope characterizing the early decay, shown as a dashed line in Fig. 2, will be referred to as slope 1. The shallower slope representative of the late decay, depicted as a solid line in Fig. 2, will be referred to as slope 2 (noise floor is disregarded during slope calculation). A time decay, which is inversely proportional to the slope, can be

calculated for each line, giving Decay1 and Decay2. The quotient of the two time decay quantities will be defined as a parameter called decay ratio:

$$\text{decay ratio} = \frac{\text{Decay 2}}{\text{Decay 1}} = \frac{T_{60_2}}{T_{60_1}}. \tag{16}$$

A second parameter, ΔdB, can be defined by the difference between the y intercepts of each of the two slopes. The dB level of slope 1 is given by the starting level of its energy decay curve. The dB level of slope 2 is found by extending the slope to the ordinate and determining the level at time zero of its energy decay curve. ΔdB is found by subtracting the second level from the first. In the earlier theory section, the early and late decay functions were normalized such that

$$\Delta dB = 10 \log\left(\frac{A_{1S}A_{2S}}{S^2}\right). \tag{17}$$

The decay ratio and ΔdB provide more information about the double slope characteristic of a particular energy decay curve than the coupling coefficient; thus, they can be used to more distinctly describe the DSE produced by a coupled volume system. Additionally, if one of the decay rates is specified, then the DSE may be uniquely defined.

Previous subjective research indicated that the difference between exponential and nonexponential decays was more easily perceived when the two slopes of the nonexponential decays were more different,[11,12] and when the y intercepts of the two slopes were minimized.[11] Note that these two trends correspond to the two proposed quantifiers: more perceptible DSE with larger decay ratio and smaller ΔdB.

The calculation of decay ratio and ΔdB requires four quantities from an energy decay function: the two slopes of the decay, and the y intercepts of those slopes. Obtaining these data from virtual and real coupled volume rooms can prove to be problematic. Particularly, determining the presence of nonexponential decay from visual inspection of the energy decay curve resulting from the Schroeder backwards integration method can be difficult. This is because background noise can alter the shape of the Schroeder curve, depending on the upper time limit of integration.[13] An analytical method is necessary to obtain accurate slope and relative dB level data for the new DSE quantifiers. One such method can be found through the use of Bayesian statistics.

B. Bayesian analysis

Xiang and Goggans developed Bayesian analysis for use in room acoustics and specifically for studying multirate decay impulse responses.[13,14] Bayesian model-based parameter estimation is used to produce an algorithm for the evaluation of multirate decay functions. The method allows for the estimation of the number of decay rates present in a multirate decay curve without requiring an initial guess on the number of slopes inherent in the decay. Additionally, this analysis can be used to determine the parameters of the decay profile, namely the slopes of the decays and ordinate intercepts of those slopes. Bayesian analysis also provides a major improvement over the classical least-squares approach. In the least-squares method, the initial estimations of parameters

TABLE I. Dimensions of main and coupled volumes with volume ratio percentage values used in computer model configurations.

	l (m)	w (m)	h (m)	V(m³)	Ratio level	Percentage of main volume
Main	34	28	26	24752		
Coupled 1	21	18	13	4914	1	20%
Coupled 2	23	22	17	8602	2	35%
Coupled 3	26	24	20	12480	3	50%

that describe an impulse response must be relatively close to actual values for iteration convergence to occur. The Bayesian method avoids this problem, as careful estimates of the initial values are not required.

In Xiang and Goggans' work, the Bayesian analysis method is shown to be a suitable approach to predicting the decay rates from sound energy decay functions. There is found to be little dependence on the upper time limit of integration or signal to noise ratio. Bayesian analysis can therefore be used to determine the slopes and intercepts in a multirate decay function, and when applied to double slope decay data, it yields the values necessary to calculate the parameters of decay ratio and ΔdB.

IV. COMPUTATIONAL ANALYSIS

This research project uses computer modeling of room acoustics to determine the individual and interactive effects of three architectural parameters on DSE: the volume ratio between the main and secondary spaces, the absorption ratio between the two spaces, and the aperture size. These three variables were modified in a simplified coupled volume system composed of two rectangular boxes connected through one acoustically transparent aperture. The coupled system was realized as a three-dimensional space and studied using the ODEON v5 computational modeling program.[15,16] Although it is beyond the scope of this paper to verify computer modeling as a tool to study DSE, ongoing work by the authors has validated this methodology.[17]

The ranges of values for the architectural parameters in question were developed based on data from existing halls that utilize dedicated coupled volume systems in their design. Additionally, consideration was taken from the values given by the spaces studied in the previous research mentioned earlier.

The first architectural parameter, volume ratio, quantifies the size of the coupled volume with respect to that of the main volume, and is represented as a percentage in this research. Existing dedicated coupled volume halls reviewed for this study have volume ratios ranging from 28% to 50%. Therefore, three volume ratios closely matching this range have been used in this study: 20%, 35%, and 50%. In the computer models, the main volume was held constant at 24 752 m³, and the coupled volume was varied to obtain the respective volume ratio percentage. Details are provided in Table I.

The second architectural parameter is the absorption ratio. This ratio depicts the equivalent absorption area in the coupled volume as a percentage of the equivalent absorption area in the main volume. The equivalent absorption area in each space, with units of m²-sabins, is calculated by multiplying its total surface area by its average absorption coefficient, which in this study has been applied uniformly to all surfaces in the space. Absorption ratio percentage values range from 3% to 14% in the existing halls reviewed for this study; four levels across this range have been modeled. The main volume's absorption coefficient in the computer models was held constant at 0.25. Since the surface area in the coupled space changes as its volume changes, the resulting absorption coefficients in the coupled space varied across volume ratio in order to achieve the desired absorption ratio. Values for the four levels of absorption ratio used in this study, labeled as a, b, c, and d, and the respective coupled volume absorption coefficients are shown in Table II.

The final varying parameter is aperture opening size. In this study, aperture opening size is defined as the surface area of the opening between the main and coupled volumes, given as a percentage of the available aperture area. The available aperture area is defined by the surface area of the intersection of the two volumes, which equals the surface area of one face of the coupled volume, and thus varies with each volume ratio. Five percentages of this available area have been used for the aperture opening size, as shown in Table III. The aperture opening size can also be given as a percentage of the main volume surface area, a quantity commonly used in earlier research; those corresponding percentages are also listed in Table III.

Varying the three parameters resulted in a total of 60 different combinations. Each combination was created as one of 60 model configurations in the ODEON program. Figure 3 shows a representative model configuration.

In addition to the previously listed absorption coefficients, the surfaces in each computer model were assigned a uniform scattering coefficient of 0.3. This value is usually a

TABLE II. Absorption ratio values used in computer model configurations.

				Absorption ratio percentages			
				(a) 3%	(b) 7%	(c) 11%	(d) 15%
Volume ratio	Main surface area (m²)	Main absorption coefficient	Coupled surface area (m²)	Coupled absorption coefficients			
(1) 20%	5128	0.25	1770	0.02	0.05	0.08	0.11
(2) 35%	5128	0.25	2542	0.01	0.03	0.05	0.07
(3) 50%	5128	0.25	3248	0.01	0.03	0.04	0.02

TABLE III. Aperture opening size as a percentage of available surface area used in computer model configurations.

| | | Percentage of available aperture | | | | |
| | | 2% | 5% | 10% | 20% | 40% |
Volume ratio	Available aperture area m^2	Percentage of main surface area				
(1) 20%	234	0.09%	0.23%	0.46%	0.91%	1.83%
(2) 35%	374	0.15%	0.36%	0.73%	1.46%	2.92%
(3) 50%	480	0.19%	0.47%	0.94%	1.87%	3.74%

satisfactory, average scattering coefficient for use in room acoustic computer modeling,[15,16] although with a low level of geometric model detail, as used in this study, it may result in a less diffuse sound field.[18] Each of the 60 models was processed in ODEON to produce an impulse response for each configuration. An omni-directional source was used with an impulse response length of 4000 ms and 1705 rays. Generally, a higher number of rays will lead to denser reflection information in the impulse response, since more rays are able to travel in and out of the secondary space. Current research incorporating computer modeling of coupled volume spaces shows a tendency to use a higher number of rays,[19] and future research should further follow this issue in a more systematic way. For the study discussed here, though, inspection of ray-tracing graphics in ODEON produced satifactory results showing that a sufficient amount of rays had traveled between the main and secondary spaces.

Objective measures, such as T_{15} and T_{30} values, were subsequently extracted from each impulse response. Finally each impulse response was evaluated with Bayesian analysis to obtain its decay slopes and y intercepts.

V. COMPUTATIONAL RESULTS

A. Coupling coefficient results

The coupling coefficients for the 60 models were found by calculating the respective T_{30}/T_{15} ratios. These coupling coefficients are plotted as a function of absorption ratio and aperture size for the three volume ratios in Fig. 4. A few general trends can be noted regarding the individual effects of the architectural parameters on the quantifier, coupling coefficient.
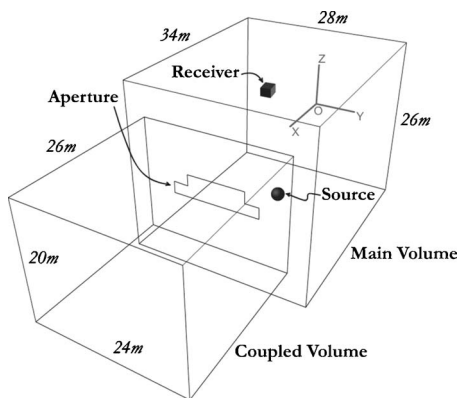
Looking across volume ratio levels, we see that coupling coefficient generally increases as the size of the coupled volume increases. This effect is in accordance with physical expectations. A larger coupled volume would allow sound energy to dissipate more slowly, feeding back into the main volume over a longer period of time, increasing the late energy in the main volume.

Across absorption ratio, coupling coefficient is shown to decrease as the level of absorption in the coupled volume increases. Again, this result is corroborated by the behavior of sound in the physical world. For larger values of α in the coupled volume, sound energy would be absorbed more rapidly. Therefore, the disparity between the early and late energy would decrease due to the smaller amount of late energy being fed back into the main hall.

Finally, in observing the general effect of aperture size, note that coupling coefficient peaks at a particular opening size for each volume/absorption combination, which is typically a small percentage of the total surface area of the main volume. In the physical world, smaller aperture sizes would force sound energy from a relatively reverberant coupled volume to leak back into the main volume more slowly, thus increasing the late decay time. Aperture sizes that are too small, however, would reduce the probability of the sound reentering the space before dissipation, and would therefore decrease the chances of obtaining a double slope decay profile. These observations are in line with the results found in previous research.

FIG. 3. Representative model configuration: volume ratio 3 and aperture size 10% of available area, with source at coordinates (25, 5, −5) and receiver at (5, −5, 5).
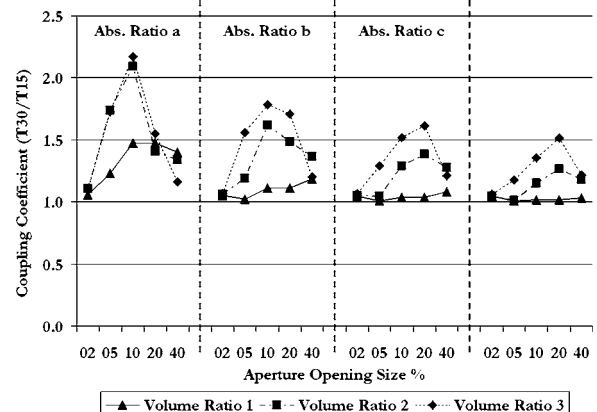
FIG. 4. Coupling coefficient data for all 60 configurations.
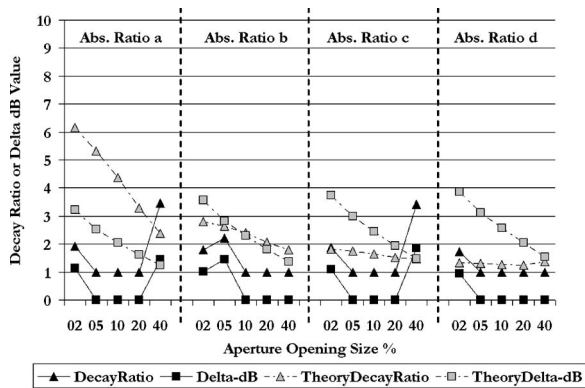
FIG. 5. Computer modeled and theoretical decay ratio and ΔdB data for volume ratio 1.

The results in Fig. 4 also provide insight into the interactions between the three parameters studied. Note that larger coupled volumes with aperture sizes approximately 10%–20% of the available aperture area (or roughly 1% of main volume's surface area) produce high coupling coefficient values. This effect is maximized with lower absorption ratios. A small volume ratio, though, only results in higher coupling coefficients if the absorption ratio is also quite low. In summary, the coupled volume size seems to have the largest effect on coupling coefficient; if it is not large enough, then high coupling coefficient can only be achieved with very low absorption ratios.

B. Decay ratio and ΔdB results

The results based on the newly suggested parameters of decay ratio and ΔdB, determined from Bayesian analysis, are more difficult to interpret than the coupling coefficient results because they provide more information about the double slope decay curves that has not been well-linked to subjective perception of DSE yet. The black data points and lines in Figs. 5–7 show the computer modeled results, with each graph representing a different volume ratio. The ΔdB values shown are normalized by dividing each value by the average ΔdB value across all configurations: 10.5 dB. This normalization was conducted in order to display the ΔdB data in the same range as the decay ratio data; this procedure has no influence on the overall trend of the data.
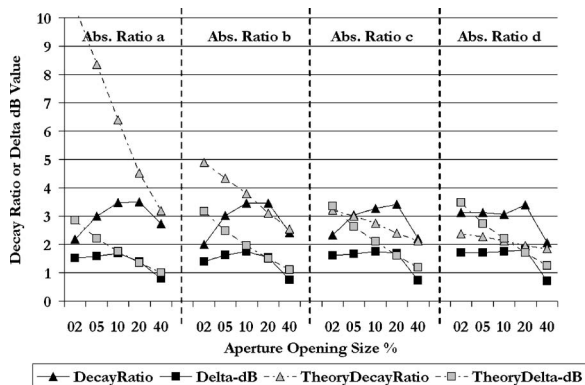
FIG. 6. Computer modeled and theoretical decay ratio and ΔdB data for volume ratio 2.
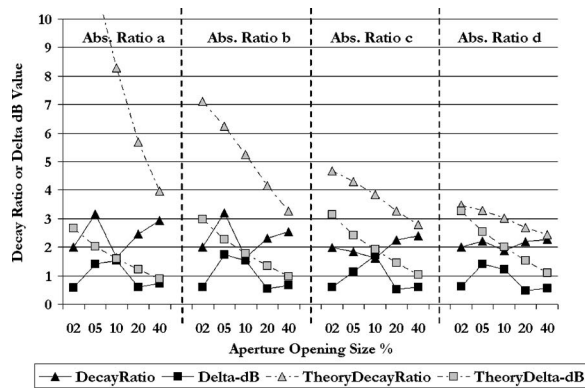
FIG. 7. Computer modeled and theoretical decay ratio and ΔdB data for volume ratio 3.

For the smallest volume ratio 1 (Fig. 5), decay ratio and ΔdB follow the same pattern across configurations with the majority of cases demonstrating decay ratio values of one and ΔdB values of zero. Such decay ratio and ΔdB values indicate an energy decay function from which the Bayesian analysis detected no double slope. In general, the results from this volume ratio suggest that smaller coupled volume sizes do not readily produce DSE. A nonexponential decay curve seems to be found at lower and higher aperture sizes, but not in a consistent or expected fashion across the cases. It is possible that the simplified geometry or selection of calculation parameters have compromised the results in these cases.

For volume ratio 2 (Fig. 6), decay ratio and ΔdB follow the same general pattern with all combinations of aperture size and absorption ratio producing some degree of nonexponential decay. Both decay ratio and ΔdB values peak at 10%–20% of the available aperture area (or roughly 1% of the main volume's surface area) for each absorption ratio. These results match the coupling coefficient trends in Fig. 4 for this volume ratio and those obtained from previous research.

As with volume ratio 2, the results for largest volume ratio 3 (Fig. 7) show all combinations of aperture size and absorption ratio producing some degree of nonexponential decay; however, the decay ratio and ΔdB do not behave similarly across configurations for this large coupled volume size. For example, there is a distinct difference at 10% aperture opening for the lower absorption ratios. Decay ratio and ΔdB do not necessarily have to show the same trends, though; as a matter of fact, that is why decay ratio and ΔdB are being suggested as more accurate quantifiers in this paper. Having both of these descriptors provides a more distinct description of nonexponential decays within coupled volume systems. Figure 8 demonstrates a number of cases of different combinations of decay ratios and ΔdB, where decay ratio may be large while ΔdB small and vice versa. What remains unclear at this point is how these different nonexponential decays are perceived. Subjective testing is required to study further how decay ratio and ΔdB may be linked to perception of DSE.

The theoretical derivation given earlier in this paper showed how decay ratio and ΔdB can be calculated from

SPL

Time

Large Decay Ratio
Large ΔdB

SPL

Time

Large Decay Ratio
Small ΔdB

SPL

Time

Small Decay Ratio
Large ΔdB

SPL

Time

Small Decay Ratio
Small ΔdB
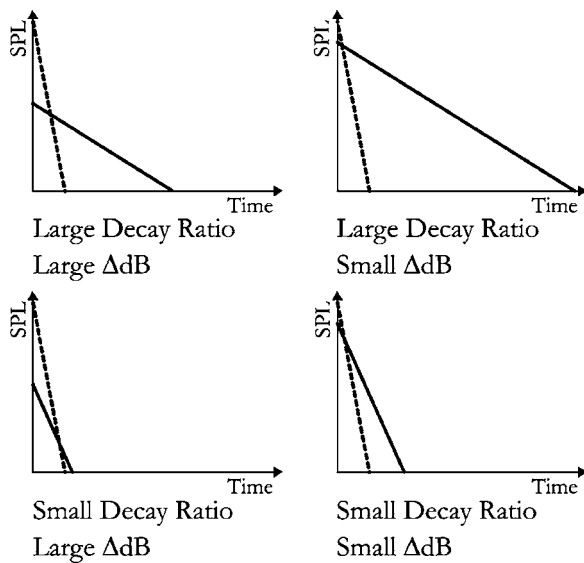
FIG. 8. Four examples of double slope decay outlines showing different combinations of decay ratio and ΔdB.

knowledge of the geometrical parameters. These theoretical values have been calculated for each of the 60 configurations used in the computer simulations, and are shown by the gray data points and dashed lines in Figs. 5–7. The theory expects decay ratio to decrease as aperture size increases, decrease as absorption ratio increases, and increase as the coupled volume increases. Recall that previous research found that non-exponential decays with larger decay ratios were more easily differentiated from exponential decays; the theory finds that larger decay ratios are obtained with smaller aperture openings, smaller absorption ratios, and larger volumes (with some values off the y-axis range in Figs. 6 and 7). The computer modeled decay ratio data generally matches these trends, but not completely across all configurations. The main differences are that the computer modeled behavior at the smallest aperture openings always demonstrates smaller decay ratios. Also the computer modeled results do not show a clear increase in decay ratio when increasing from volume ratio 2 to 3.

The theoretical ΔdB values decrease as aperture opening increases, increase as absorption ratio increases, and decrease as the coupled volume increases. Recall that previous research found that nonexponential decays with smaller ΔdB were more easily differentiated from exponential decays; the theory finds that smaller ΔdB are obtained with larger aperture openings, smaller absorption ratios, and larger volumes. The computer modeled ΔdB results show similar tendencies for aperture size and volume, but the trend with absorption ratio is not found.

The differences between the theoretically expected decay ratio and ΔdB values and the computer modeled results obtained from Bayesian analysis could be due to the following factors. Summers' work indicated that results between computer models and statistical theory deviate when diffuse fields are not achieved, and also when there is a high level of coupling.[9] The computer modeling parameters, particularly the scattering coefficient and number of rays used, may have detracted from the diffuseness of the generated sound fields.

Another limitation in both the theory and the computer modeling analysis involves the behavior at small aperture openings. In the physical world, impedance effects at the boundary and considerations of wavelength compared to opening size could reduce the decay ratio values, so that the high values of decay ratio predicted theoretically would be invalid. Furthermore, high decay ratios are usually accompanied by higher uncertainty values in the Bayesian analysis, as explained by Xiang et al.[20]

Although the decay ratio and ΔdB results from theory and from Bayesian analysis of the computer modeled impulse responses do not perfectly match, it is still of interest to use the simulated results to learn more about the subjective perception of DSE.

VI. SUBJECTIVE ANALYSIS

To help reach the goal of having an objective measure of DSE that correlates to subjective response, psychoacoustic testing has been conducted. Another purpose of this testing was to determine the effect of the architectural parameter variation on subjective response to DSE. A subset of the impulse responses generated in the computational phase of this study was used in the subjective testing phase. The subset consisted of the sound fields simulated in the combinations of the volume ratio and aperture opening size variables, producing 15 impulse responses in total. The absorption ratio for these combinations was held constant at the lowest level of 3%, since lower levels are most likely to produce DSE according to theory and previous research. Subjects were asked to rate the perceived reverberation and clarity from a series of sound tracks. These two acoustic qualities were chosen as the subjective variables because DSE is allegedly related to having both high reverberance and clarity. Combining the results from subjective perception of these two qualities would indicate the broader psychoacoustic response to DSE.

Auralizations were produced in ODEON by convolving the impulse responses for each configuration with an anechoic music sample. The anechoic piece used in this study was a selection from Beethoven's 9th symphony performed by the Osaka Philharmonic Orchestra.[21] A KEMAR head related transfer function was applied to the convolved impulse responses in ODEON. A series of 15 sound tracks were created from the 15 configurations. The tracks were presented over headphones to 30 human test subjects, and the presentation order was randomized to reduce bias error. Subjects were members of the University of Nebraska community, with the majority being between 20 and 30 years of age. Each subject was determined to have normal hearing thresholds in both ears. Most of the subject pool indicated that they had limited exposure to classical music, such as that played during the testing.

Subjects were given a brief training session that covered the definitions of the acoustical qualities of reverberation and clarity. After the training period, subjects were asked to rate each track on a 9 point scale for reverberation (dead to live) and clarity (unclear to clear).
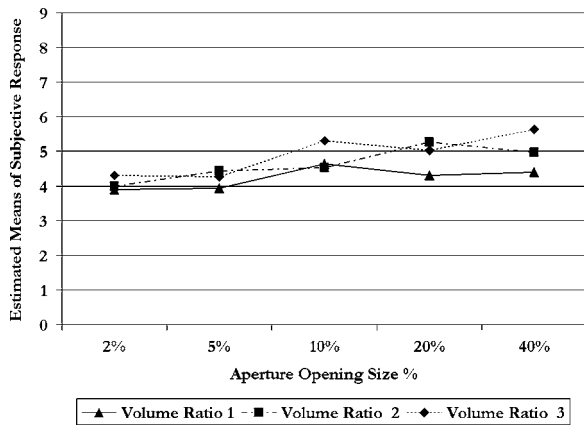
FIG. 9. Statistically estimated means of subjective ratings for perceived reverberance.

FIG. 11. Objective early decay time (EDT) values for the subset of configurations used in subjective testing.

VII. SUBJECTIVE RESULTS

This study used a repeated measures design with two independent variables: volume ratio with three levels, and aperture size with five levels. There were two dependent variables: perceived reverberation and perceived clarity.

A two-way repeated measures ANOVA was performed on these data. Results indicate that there was a significant effect of volume ($p < 0.0001$) and aperture size ($p < 0.0001$) on perceived reverberation. Additionally, perceived reverberation showed a significant effect due to the interaction between volume and aperture size ($p = 0.028$). There were no observed significant effects on perceived clarity. Figures 9 and 10 show the statistically estimated mean values of perceived reverberation and perceived clarity for each combination of volume and aperture size.

The nonparallel lines in Fig. 9 do indicate a reaction between the independent variables, volume ratio, and aperture size. In other words, as both independent variables change, the perceived reverberation results change in a manner different from when the independent variables are changed separately. It appears that the main cause of the nonparallel lines is the volume ratio 2 data in the midrange of aperture opening sizes tested, implying that this is a region where the subjects may find it difficult to differentiate between their perceptions of reverberation.

Post-hoc pairwise Bonferroni comparison tests revealed that, for perceived reverberation, volume ratios 2 and 3 were significantly different from volume ratio 1. Additionally, for perceived reverberation, the 10%, 20%, and 40% openings significantly differed from the 2% opening. Also, the 20% and 40% openings significantly differed from the 5% opening ($p < 0.05$ for all significant tests).

In summary, the volume ratio and aperture size have been found to affect perceived reverberation in this study. Listeners generally perceived a higher reverberation as volume and aperture size were increased independently. Figure 11 shows the objective early decay time (EDT) results from the computational analysis of the impulse responses of each model. These data show a clear tendency for EDT to increase as volume and aperture size are increased. The early portion of the reverberant decay has long been regarded as a good indicator of subjective impression of reverberance.[22,23] Therefore, comparative analysis between the EDT values and the subjective response to reverberance shows good agreement between the objective and subjective results.

The lack of significant effects with clarity is most likely due to the nature of the simplified forms of the modeled coupled volume systems. Clarity may be objectively quantified by the clarity index, a ratio of early to late sound energy received at a given position in a space. The clarity index is given in dB as

$$C_{80} = 10 \log\left(\frac{\int_{t=0}^{80 \text{ ms}} g^2(t)dt}{\int_{t=0}^{\infty} g^2(t)dt} \right), \tag{18}$$

where $g(t)$ is the impulse response of the space in time. This measure is affected greatly by early reflections. In this study, the models were comprised of two simple rectangular boxes, a geometry that does not produce a high number of early reflections that are distinguishable from other reflections. As the coupled space's volume and the aperture size are changed, the early reflections are unfortunately not directly influenced. Objective clarity (C_{80}) results from the 60 configurations studied in this paper are shown in Fig. 12. The greatest differences in the C_{80} values were found when comparing across volume ratios, giving an average difference of 0.68 dB, which is on the order of the just noticeable difference: $0.67 \pm 0.13 \ dB$.[24] These objective results

FIG. 10. Statistically estimated means of subjective ratings for perceived clarity.

Bradley and Wang: Effects of simple coupled volume geometry
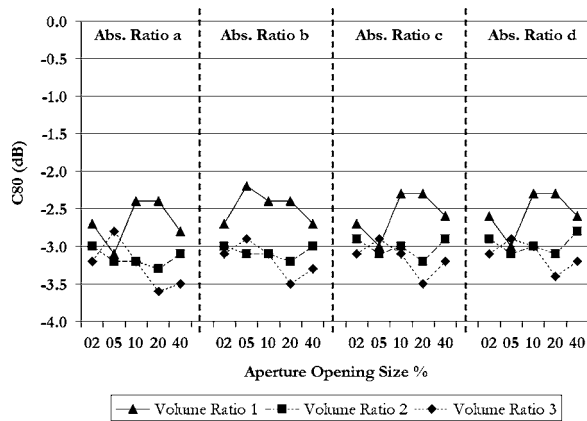
FIG. 12. Objective clarity (C_{80}) values for the subset of configurations used in subjective testing.

TABLE IV. Computer modeled decay ratio and ΔdB data, and statistically estimated perceived reverberation data for volume ratio 3 and five aperture opening size percentages.

Aperture size (%)	Decay ratio	ΔdB	Statistically estimated reverb
2	2.00	0.58	4.30
5	3.16	1.40	4.27
10	1.64	1.53	5.30
20	2.46	0.60	5.03
40	2.94	0.72	5.63

suggest no perceptual change in clarity across configurations, agreeing with the subjective results obtained.

The original hypothesis of the authors was that higher perceived reverberation and perceived clarity resulted from nonexponential decay curves that demonstrated greater degrees of nonexponential decay, and that this degree of nonexponential decay could be objectively quantified by coupling coefficient or more precisely with the new combination of decay ratio and ΔdB. Unfortunately, since perceived clarity did not result in any significant results in this study, it is not possible at this time to delve further into the prospect of nonexponential decays producing both high reverberation and high clarity.

Even though the perceived clarity results are not obvious, comparison of the perceived reverberation results to the coupling coefficient, decay ratio, and ΔdB results show very good correlations. As indicated previously, volume ratios 2 and 3 produced statistically higher perceived reverberation values than volume ratio 1. These larger volume ratios also produced higher values of coupling coefficient, decay ratio, and ΔdB than volume ratio 1. The same is true for the subjective results of aperture size; aperture sizes around 20% produced significantly higher perceived reverberation values than those from the lower apertures of 2% and 5%. For volume ratio 2, these larger aperture sizes generally produced higher values of coupling coefficient, while the decay ratio and ΔdB results were more varied depending on the volume ratio.

Recall that the decay ratio and ΔdB data from volume ratio 3 of the computer simulations exhibited different trends across aperture opening size (Fig. 7). To analyze how each of the new quantifiers may relate to subjective perception of reverberation, a comparison of the volume ratio 3 data is made against the statistically estimated means of reverberation from the subjective tests (Table IV). Comparing cases that were found to be significantly different from the Bonferroni tests, we see that in comparing the 2% and 20% cases, decay ratio increases while ΔdB remains about the same, correlating to higher perceived reverberation. Similarly in comparing the 2% and 40% cases, decay ratio increases while ΔdB increases slightly, again correlating to higher perceived reverberation. Comparing the 5% and 20% cases, de-

cay ratio and ΔdB both decrease, resulting in higher perceived reverberation. A comparison of the 5% and 40% cases shows decay ratio decreasing slightly and ΔdB decreasing a larger amount, producing higher perceived reverberation. Comparison between 2% and 10%, however, shows a different trend not expected from previous research; decay ratio decreases and ΔdB increases, producing higher perceived reverberation. The authors believe this anomaly may be due to the issues discussed earlier with the computer modeling, concerning the diffuseness and density of the reflection diagram.

In summary, the results mostly confirm findings from previous research: larger decay ratios and smaller ΔdB values generally lead to greater perceived reverberation. It still remains unclear, though, which of the two parameters has the more severe influence on DSE overall, since no significant relationships were found with perceived clarity ratings.

VIII. CONCLUSIONS AND FUTURE WORK

Computer modeling results for 60 coupled volume configurations with varying volume ratio, absorption ratio, and aperture size have been studied objectively using coupling coefficient and the new parameters of decay ratio and ΔdB. There is a general increase in coupling coefficient as coupled volume size increases, a general increase in coupling coefficient as coupled volume absorption decreases, and coupling coefficient peaks at a particular opening size for each volume/absorption combination. Coupling coefficient results also suggest that among the three parameters, having larger coupled volume sizes is the most significant in producing higher coupling coefficient values. Smaller coupled volume sizes can only achieve high coupling coefficients if the absorption ratio is also very low. The proposed parameters of decay ratio and ΔdB provide more specific information about each nonexponential decay curve, and have been found to show similar trends as coupling coefficient. However, the new parameter data for the highest volume ratio tested were harder to interpret without further information about subjective perception.

Subjective testing using the results from 15 of the configurations showed a general increase in perceived reverberance with increasing volume and aperture size. Higher perceived reverberance also matched well with objective coupling coefficient results, and generally with higher decay ratios and lower ΔdB values. No significant effects on perceived clarity were found, which may be due to the lack of distinguishable early reflections in the simplified geometry studied. Due to the lack of this significant effect, further

extrapolation by combining the subjective results of reverberation and clarity could not be carried out to determine overall subjective response to DSE.

Further study in this area could focus on advancing a metric for describing DSE that will allow for the effects of the architectural parameters on coupled volume systems to be more clearly understood. Several issues can be improved in ongoing work. The simplified geometry of the models may have adversely affected the results due to the lack of strong early reflections. Consequently, models with more realistic geometries can be built to produce results that better articulate the clarity of the systems. Additionally, certain computer modeling calculation parameters may have had an adverse effect on the results. The number of rays used, for example, could have been significantly higher, allowing for a more accurate exploration of the coupled volume geometry. Future work could incorporate these modifications, and more concrete conclusions regarding DSE prediction and perception may be made with further research.

ACKNOWLEDGMENTS

[1] L. Cremer, H. A. Müller, and T. J. Schultz, *Principles and Applications of Room Acoustics* (Applied Science, New York 1982), Vol. 1, pp. 189–193, 261–292.

[2] H. Kuttruff, *Room Acoustics* (Spon, New York 2000), pp. 142–145.

[3] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York 1989), pp. 281–283.

[4] A. Munro and R. Johnson, "Calculating reverberation time in coupled rooms," *Workshop on Design of Concert Chambers, Tanglewood '99 Concert Hall Research Group Summer Institute, 1999.*

[5] A. H. Davis, "Reverberation equations for two adjacent rooms connected by an incompletely soundproof partition," Philos. Mag. **50**, 75–80 (1925).

[6] C. F. Eyring, "Reverberation time measurements in coupled rooms," J. Acoust. Soc. Am. **3**, 181–206 (1931).

[7] B. Harrison and G. Madaras, "Computer modeling and prediction in the design of coupled volumes for a 1000-seat concert hall at Goshen College, Indiana," J. Acoust. Soc. Am. **109**, 2388(A) (2001).

[8] M. Ermann and M. Johnson, "Pilot study: Exposure and materiality of the secondary room and its impact on the impulse response of coupled-volume concert halls," J. Acoust. Soc. Am. **111**, 2331(A) (2002).

[9] J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**, 958–969 (2004).

[10] B. S. Atal, M. R. Schroeder, and G. M. Sessler, "Subjective reverberation time and its relation to sound decay," *Proceedings of the Fifth ICA, Liege, Paper G 32, 1965.*

[11] D. Picard, "Audibility of non-exponential reverberation decays," *Rapport de Stage D'Option Scientifique* (Rensselear Polytechnic Institute, Troy, NY, 2003).

[12] M. Ermann, "Coupled volume/double slope subjective listening test," J. Acoust. Soc. Am. **114**, 2331(A) (2003).

[13] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1424 (2001).

[14] N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," J. Acoust. Soc. Am. **113**, 2685–2697 (2003).

[15] C. L. Christensen, *ODEON Room Acoustics Program User Manual V6.5 Industrial*, Auditorium and Combined Editions, Odeon A/S, Lyngby, Denmark, 2003.

[16] C. L. Christensen, "ODEON-a design tool for auditorium acoustics, noise control and loudspeaker systems," *Proceedings of the Institute of Acoustics: Reproduced Sound 17, Measuring, Modeling or Muddling, 2004*, Vol. 23, pp. 137–144.

[17] D. T. Bradley and L. M. Wang, "Comparison of real world measurements and computer model results for a dedicated coupled volume system," J. Acoust. Soc. Am. **116**, 2552(A) (2004).

[18] H. Kuttruff and Th. Strassen, "Zur abhängigkeit des raumnachhalls von der Wanddiffusität und von der raumform," Acustica **45**, 246–255 (1980).

[19] L. Nijs, G. Jansens, G. Venmeir, and M. van der Voorden, "Absorbing surfaces in ray-tracing programs for coupled spaces," Appl. Acoust. **63**, 611–626 (2002).

[20] N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayesian decay time estimation," J. Acoust. Soc. Am. (to be published).

[21] Osaka Philharmonic Orchestra, *Anechoic Orchestral Music Recording* (Denon Records, 1995)

[22] R. H. Bolt and P. E. Doak, "Tentative criterion for the short-term transient response of auditoria," J. Acoust. Soc. Am. **22**, 507–509 (1950).

[23] M. R. Schroeder, B. S. Atal, G. M. Sessler, and J. E. West, "Acoustical measurements in Philharmonic Hall (New York)," J. Acoust. Soc. Am. **40**, 434–440 (1966).

[24] T. J. Cox, W. J. Davies, and Y. M. Lam, "The sensitivity of listeners to early sound field changes in auditoria," Acustica **79**, 27–40 (1993).

Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing[a)]

Jason E. Summers,[b)] Rendell R. Torres, and Y. Shimizu[c)]
Program in Architectural Acoustics, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180

Bengt-Inge L. Dalenbäck
CATT, Mariagatan 16A, SE-41471 Göteborg, Sweden

The ability of computational geometrical acoustics to accurately model energy decay in systems of coupled rooms is investigated both theoretically and experimentally. Unlike single-volume rooms, coupled rooms have reflection density that is not described by a single quadratic function of time. It is shown that tail-correction procedures used by beam-axis-tracing algorithms, which assume quadratic growth of reflection density, can lead to inaccurate predictions in coupled rooms. Further, beam-axis tracing implemented as ray tracing with a growing detection sphere is susceptible to error in coupled rooms when the detection sphere extends into adjacent subrooms. Marked error is anticipated in those cases for which the source and receiver are in the less reverberant of two rooms and is expected to be most severe for (1) small coupling apertures and (2) receiver positions near boundaries between subrooms. Errors are demonstrated by comparison of computational geometrical acoustics predictions with scale-model measurements made in a two-room coupled system. A revised beam-axis/ray-tracing algorithm is investigated that circumvents possible error mechanisms by switching to ray tracing for the late part of the decay. Comparisons with scale-model measurements indicate that the revised algorithm is able to predict energy decay accurately in coupled rooms. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2000772]

PACS number(s): 43.55.Ka [NX]   Pages: 1491–1502

I. INTRODUCTION

Computational modeling of sound fields in rooms using software based on geometrical acoustics (GA) is now commonplace among researchers and practitioners in the field of architectural acoustics. As a result, GA software is being applied to an ever-increasing variety of room types—many of which were never anticipated when the software algorithms were first developed. Of the nontraditional room types being modeled, systems of coupled rooms are of particular interest because of growing use of such geometries in the design of auditoria (see, e.g., Chaps. 2 and 3 of Ref. 1). However, the accuracy of certain computational GA algorithms is compromised when applied to coupled rooms.[2–5] Based on comparisons between predictions of GA models and acoustical measurements of coupled-room systems, prior studies have suggested that ray-tracing and beam-axis-tracing algorithms require many more rays or beam axes, relative to the numbers required for single-volume rooms, to accurately predict sound fields in systems of coupled rooms.[4,5] Authors have also reported that methods that algorithmically generate the late portion of the decay curve in order to extend the duration beyond the truncation time of a computational GA algorithm (a technique that is distinct from the tail-correction procedures discussed below) can lead to errors when applied to coupled rooms.[5,6]

This paper presents an investigation of the ability of the previously developed commercial GA software CATT-Acoustic (v. 7.0 and above), which is based on a randomized form of the widely used beam-axis paradigm,[7,8] to accurately predict sound fields in coupled rooms in the high-frequency region for which wave effects primarily consist of surface scattering. (Recognizing the severe limitation of pure-specular methods, randomized beam-axis tracing stochastically redirects the direction of beam axes on reflection to account for frequency-dependent surface scattering, in the same manner as ray tracing.) Systems of coupled rooms composed of large, reverberant enclosures, such as those encountered in concert halls equipped with reverberation chambers, are primarily addressed. Assumptions inherent to the algorithm are described together with the conditions under which they can be violated. Though the algorithm has been validated in single-volume rooms by comparison of predictions with measurements,[9] comparisons of predictions with scale-model measurements indicate that the algorithm displays systematic error when applied to certain systems of coupled rooms. To address this, a modified algorithm is in-

---

vestigated, which, by comparison with the same scale-model measurements, is shown to accurately model sound fields in systems of coupled rooms.

II. BEAM-AXIS TRACING

A. Background

As stated in the Introduction, the work described is this paper is largely directed toward the high-frequency region for which surface scattering is the only significant wave effect and sound fields can be accurately modeled by GA. Two primary methods have been traditionally used for the computational modeling of sound fields in rooms according to the laws of GA. Of these, the image-source model[10,11] is a deterministic pure-GA model that, when applied to coupled rooms, displays no limitations unique to that room configuration.[6] (It is, however, flawed in all geometries by an inability to model nonspecular reflection. It is further limited by its computational complexity to computing only low-order reflections or reflections in simple geometries.) The second method, ray tracing,[12] is computationally less demanding, but subject to systematic geometrical errors.[13] Beam-axis tracing,[7,8] a variant of ray tracing that approximately models the propagation of divergent beams (either having triangular or circular faces), offers certain advantages, while still being subject to similar errors.

Hybrid methods have been developed that combine aspects of ray-tracing and image-source models.[2,9,14–16] In these hybrid models, rays or beam axes are used either to transport energy, as in the conventional ray-tracing approach, or to efficiently probe the geometry and locate potential image sources. Because of their foundation in (specular only) ray tracing and beam-axis tracing, these techniques are also subject to systematic geometrical errors.

Recently, interest has renewed in beam-tracing algorithms in which the intersections of beam faces with multiple surfaces are modeled exactly.[17–19] Because these models are formally identical in their predictions to the image-source model, they likewise have no limitations unique to coupled rooms. At present, such models are limited to rooms that can be modeled by a moderate number of polyhedra [i.e., those without highly faceted surfaces (see p. 753 of Ref. 19)], and also suffer from inability to model nonspecular reflection. Moreover, the computational complexity of even the most recent of these models[19] remains exponential in reflection order, suggesting that the high-order-reflection calculations required to model reverberation may be impractical in some cases.

B. Mechanisms of energy-detection error

Beam tracing is most often applied approximately, by tracing only the central axes of the beams—so-called beam-axis tracing. This approach offers the advantage of decreased computing time by covering the 4π steradians outwardly visible from the source more efficiently than ray tracing (i.e., fewer beam axes are needed than rays due to the increased detection probability of a beam face relative to a ray). However, it can result in incorrect detections of energy arrivals at receivers in the late part of the decay, which must be ad-

dressed by a tail-correction procedure. The two primary mechanisms by which energy is incorrectly detected are discussed below.

First, if only the axes of beams are traced, it is possible that a surface that would interact with a portion of the outgoing wavefront represented by one of the beam faces will not be accounted for because it will not interact with the axis of the beam. The area of the beam face grows over time and gives a measure of the spatial-detection uncertainty. When the beam face is on the order of the area of representative room surfaces and larger, significant numbers of geometrically valid reflections are unaccounted for. This error mechanism can be viewed as spatial undersampling due to the divergence of the rays representing the axes of the cones.[13] As a result of this, an increasing number of geometrically valid reflections are undetected as time increases.[15,16] In beam-axis tracing, unlike ray tracing, this results in an increasing loss of detected energy arrivals over time due to the explicit reduction of the energy of each beam axis to account for spherical divergence, which it is not balanced by geometrically invalid detections. However, these undetected reflections can be naturally compensated for by relaxing the geometrical detection criteria with time to allow an increasing number of reflections to be detected that are *almost* geometrically valid.[9,16]

A second source of error results if the size of the beam face at late times exceeds the size of the total room such that each reflection is detected by all receivers, regardless of whether the arrival is physically possible or not. The particular form that this error takes depends on the algorithm used to model beam-axis tracing. In a single-volume room, the observed reflection density $n(t)$ asymptotically approaches a constant value[15]

$$\lim_{t \to \infty} n(t) = \frac{\pi c M \beta \langle \overline{\ell} \rangle^2}{4V}, \tag{1}$$

where t is time, c is the speed of sound, M is the number of beam axes, β is an empirically derived factor depending on the shape of the beam faces and their overlap (which has an analytical expression for Gaussian shaded beams of circular cross section[15]), and $\langle \overline{\ell} \rangle$ is the mean-free path length.

It should be noted that if beam axes are used to find image sources in the hybrid method and the geometrical validity of these sources is checked, some valid image sources (the amount increasing over time) will be missed, and a tail-correction procedure will likewise be needed.

In contrast with beam-axis tracing, pure ray tracing needs no correction of the energy levels of the arrivals in the computed echogram (energy impulse response) because the effect of undetected geometrically valid reflections is balanced by the effect of multiple and geometrically invalid arrivals,[13,15] though the consequence of this is that the tail of the decay has unnaturally low reflection density growth and, unless the receiver sphere is large, shows greater run-to-run variation.

C. Tail correction

Tail-correction procedures required by beam-axis tracing have been typically developed for use in single-volume rooms. Consequently, the assumptions of these procedures need not be valid in systems of coupled rooms. Three approaches presented in the literature are considered here. Each of these approaches relies on applying an energy correction to the arrivals that are detected. As such, these tail-correction algorithms assume that the detections are representative of all correct arrivals, including those that are not detected. Detection is essentially a random process and, for numbers of beam axes that would be adequate in a single-volume room, but are not in a system of coupled rooms, the actual number of detections in the late part can be small enough to produce large run-to-run fluctuation in the predicted decay rate.

1. Reflection growth

Following an image-source argument,[20] the observed reflection growth (or *reflection density*), $n(t)$, in a single-volume rectangular room can be shown to be, asymptotically, equal to the late-time expected value $\tilde{n}(t)$, which is quadratic in time following the expression

$$\tilde{n}(t) = \frac{dN}{dt} \approx \frac{4\pi c^3 t^2}{V}, \tag{2}$$

where N is the number of reflections received up to a time t and V is the volume of the room. Vorländer[21] has plausibly argued that this result for the true asymptotic behavior of $\tilde{n}(t)$ is general for any room shape under the assumptions of exponential decay from steady state and isotropic incidence of sound on the receiver. A number of authors[16,22] have further suggested that a quadratic growth of reflection density occurs, at least approximately, in an even broader class of rooms (see Appendix G of Ref. 1 for further discussion). Failure to detect late-time arrivals can therefore be addressed by assuming quadratic growth of reflections and accordingly increasing the energy of those late reflections that *are* detected according to the growth rate determined by a least-squares fit to the early portion of the decay.[16] This is the approach to tail correction used by CATT-Acoustic (v. 7.0 and above).

2. Analytic expressions

Lehnert and Blauert[22] have proposed a method in which the energy of the arrivals is increased by a factor inversely proportional to the detection probability after the growth rate of the observed reflection density drops below the rate expected and becomes approximately constant in time. A similar approach[23] assumes a smooth transition from the behavior of the algorithm, described by Eq. (2), to the late behavior, described by Eq. (1), and multiplies the energy of arrivals by the reciprocal of the resulting expression for the expected growth rate,

$$n(t) = \frac{4\pi c^3 t^2}{V} \left[1 - \exp\left(-\frac{t_c^2}{t^2} \right) \right], \tag{3}$$

where $t_c \propto \sqrt{M/\beta}$ is the intersection point in time of the expected reflection growth rate with the constant asymptotic rate.[15]

3. Sources of error

All of these methods of tail correction require the assumption that Eq. (2) is correct or, at least, $\tilde{n}(t) \propto t^2$. Decay of energy in a rectangular parallelepiped room with nonuniform absorption and insufficient surface scattering is nonexponential[24] (i.e., yields a nonlinear decay curve) yet displays a quadratic growth in reflection density.[20] Thus, assuming quadratic growth of reflection density is not equivalent to assuming exponential decay of energy. But it is not always accurate to make the assumption of quadratic reflection growth in other cases. In particular, the true asymptotic value of reflection growth $\tilde{n}(t)$ in a system of coupled rooms is not necessarily described by Eq. (2).

It is primarily quadratic but time-variant reflection growth, rather than nonquadratic reflection growth, that, by violating the assumptions of tail-correction procedures, leads to errors in predictions using GA algorithms in coupled rooms. In single-volume rooms the method-dependent observed $n(t)$ deviates from the predicted value given by Eq. (2) initially, but later in time is well described by the quadratic function. In coupled rooms, this behavior is not observed. Instead, for the computational GA algorithm, room geometry, and aperture sizes studied here, the following behavior is observed. If source and receiver are in the same subroom, the initial behavior of $n(t)$ is approximately what it would be in an uncoupled room of the same volume. As time progresses, $n(t)$ gradually transitions to approximately behave as though in a single-volume room having the same volume as the entire coupled system. (See Fig. 7 for an illustration.) If the source and receiver are in different subrooms, $n(t)$ initially increases very slowly in time. As in the prior case, $n(t)$ transitions at later times to approximate the behavior of a large single-volume room, though the curve observed in the system of coupled rooms is displaced laterally relative to the curve predicted for a single-volume room due to the slower initial buildup of reflection density in the system of coupled rooms. (See Fig. 12 for an illustration.)

The result of $n(t)$ that transitions between two quadratic functions of time, as in the case of the source and receiver in the same subroom, is that a tail-correction algorithm will attempt to adjust the energy of detected arrivals based on an estimated reflection density that may be either too high or too low. The consequences are essentially the same in the case of source and receiver in different subrooms: tail correction is based on an inaccurate estimate of $\tilde{n}(t)$. In this case it is a particular problem if the form of $\tilde{n}(t)$ is estimated from the early portion of the computed response, as this results in an estimate of $\tilde{n}(t)$ that will be much too small to yield useful tail correction.

D. Other considerations

A second error source unique to coupled-room geometries can occur for certain implementations of beam-axis tracing. In single-volume rooms, increasing size of circular beam faces is equivalent to increasing radii of detection spheres surrounding receivers. Therefore, certain computational models (such as CATT-Acoustic) implement beam-axis tracing as ray tracing with spherical receivers that increase in radius with time.[16] This presents no problem for single-volume rooms because the case in which the detection sphere exceeds the size of the room and $n(t)$ approaches a constant value corresponds to the case in which the beam face exceeds the size of the room. However, in coupled rooms, the growing detection sphere has two distinct points of transition. The first is its extending into one or more of the adjacent subrooms, resulting in a change in $n(t)$. The second is its exceeding the size of the entire coupled system, which yields constant $n(t)$. The first point of transition may be marked by an upward shift in the level of the decay curve, particularly in those cases for which the energy density of the adjacent subroom is significantly higher. Thus, the level increase of the error is expected to be greatest in systems that are weakly coupled (roughly speaking, those in which subrooms are coupled by small apertures), and should be earlier in time (higher in level) for receivers near boundaries between subrooms.

III. DESCRIPTIONS OF ALGORITHMS TESTED

A. Original algorithm

In this study, the GA software CATT-Acoustic is investigated. CATT-Acoustic (v. 7.0 and above) is based on the randomized tail-corrected cone-tracing (RTC) algorithm.[9] Reflections are treated by beam-axis tracing with a circular beam face (cone-axis tracing), but with the image-source method used to calculate all geometrically valid first- and second-order reflections. First-order nonspecular reflections are treated by the creation of secondary "diffuse reflection sources" radiating according to Lambert's law.[16] Higher-order reflections are treated by randomization of the cone direction according to the probabilistic interpretation of Lambert's law.[24] In the Lambert model, a scattering coefficient s is defined as the fraction of energy that is not specularly reflected. Surface absorption is modeled by the angle-independent power absorption coefficient α, which represents the fraction of the incident energy that is absorbed by the surface.

Tail-correction assumes a quadratic growth rate of reflection density, as described in Sec. II C 1. The algorithm assumes that $n(t)$ is proportional to t^2 and defines the constant of proportionality by fitting a straight line to the initial portion of $\sqrt{n(t)}$, the length of this portion being dependent on the number of cone axes used. Corrections are performed individually for each receiver so that varying rates of reflection growth throughout the room are accommodated. For single-volume rooms the corrections predict the decay curve rather accurately, even for a small number of cone axes, as shown in Fig. 1 (see also Fig. 3 of Ref. 16).
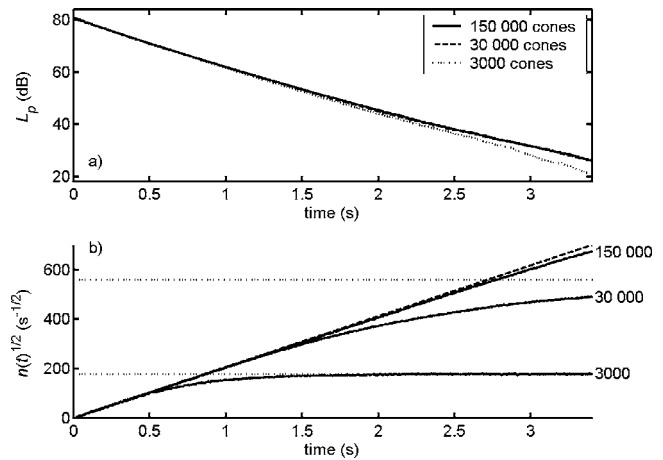
FIG. 1. Spatial ensemble averages of decay curves in the closed room 1 of the coupled system shown in Fig. 2 (configured as a single-volume room) as predicted by the RTC algorithm using 3000, 30 000, and 150 000 cones are shown in (a). Though the square root of the observed reflection growth rates $\sqrt{n(t)}$ (solid lines) shown in (b) tend toward constant slope predicted by Eq. (1) (dotted lines), the tail-correction procedure of RTC, based on a linear least-square estimate of $\sqrt{n(t)}$ (dashed line), is robust to changes in the number of cones.

B. Modified algorithm

The assumption of quadratic growth in reflection density can fail for certain systems of coupled rooms, as described in Sec. II C 3. To address this, a modified form of the RTC algorithm has been developed. The new algorithm behaves as RTC until any one of the expanding detector spheres contacts one of the surface boundaries. From that point on, rather than applying tail correction, the radius of each detection sphere is held fixed and the detection procedure at each receiver is altered to that of ray tracing (i.e., the propagation law of r^{-2} is addressed by changes in detection probability due to ray divergence, rather than explicit reduction of the energy associated with each ray).[13,22] No tail-correction procedure is necessary, but the number of rays traced must be large enough that the ray density, throughout the time that the size of the detection sphere is fixed, is sufficient to insure adequate probability of detection.[13,22] This new algorithm was first implemented as a user-selected option ("late part ray trace") in the commercial distribution of CATT-Acoustic with version 8.0b, and is also used in other cases (such as when a room includes semitransparent surfaces) for which the reflection growth cannot be assumed quadratic.

IV. SCALE-MODEL EXPERIMENTS

A. Apparatus and methods

In order to assess the accuracy of the algorithms under the most controlled conditions, measurements were carried out in a schematic 1:10 scale-model representation of a concert hall with an auxiliary coupled room, shown in Fig. 2.

The scale model was constructed from $\frac{5}{8}$-in.-thick medium-density fiberboard (MDF), which was first painted and then spray coated with two layers of polyurethane. Audience absorption was simulated by 2-in.-thick foil-backed fiberglass batting placed on the floor of the primary room. Wide-bandwidth surface scattering was provided by 30-cm
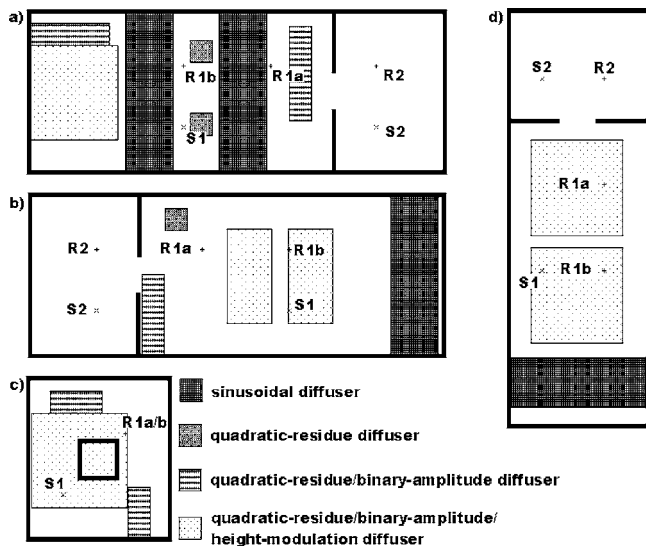
FIG. 2. Schematic representation of the scale model of two coupled rooms. Transverse (a, b) and lateral (c) sections are shown together with overhead plan (d). Source (x) and receiver (+) positions are indicated together with diffuser locations.

lengths of the commercial product RPG FlutterFree®, a seven-element quadratic-residue diffuser[25] with design frequency 5000 Hz model scale (500 Hz full-scale equivalent). The samples used were constructed of milled cedar and spray coated with two layers of varnish. In addition to being used in groups of three to form (nearly) square panels, the quadratic-residue diffusers were also used as elements of larger binary-amplitude diffusers[26] with design frequencies of 1114 and 1715 Hz model scale (111.4 and 171.5 Hz full-scale equivalent), based on 11- and 13-element Barker sequences. A photograph of a diffuser consisting of two 13-element Barker sequences is shown in Fig. 3. Binary amplitude diffusers were constructed on $\frac{1}{4}$-in.-thick plywood sheeting, which was also spray coated with two layers of varnish. Variation in the design frequency of the binary-amplitude diffusers was accomplished by using wooden spacers of two different thicknesses to adjust the depths of the wells formed between the quadratic-residue diffusers. By grouping binary-amplitude diffusers of different design frequencies together, some additional surface-height modulation with spatial period $\Lambda=60$ or 120 cm was obtained. Additional surface scattering was provided by sinusoidal surfaces of formed polyvinyl-chloride (PVC) plastic having $\Lambda=7.5$ cm and peak amplitude $\Xi=0.9$ cm.

A 1:10 scale dodecahedron loudspeaker (edge length 20 mm, driver diameter 13 mm) was used as the sound
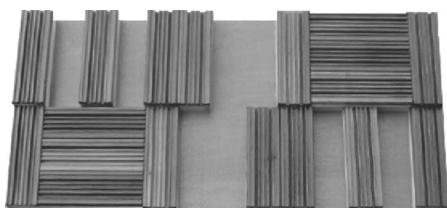
FIG. 3. Photograph of a diffuser panel consisting of two 13-element Barker sequences with design frequencies 1114 Hz (top) and 1715 Hz (bottom) with individual elements formed of quadratic-residue diffusers.

source. Measurements of far-field directivity indicate that the loudspeaker is omnidirectional to within ±5 dB over the entire frequency range of interest. Omnidirectional microphones with $\frac{1}{8}$-in. diaphragms were used as receivers. For measurements, they were mounted with the diaphragms oriented parallel to the floor. Modeling the microphones as a piston at the end of a long tube (see, e.g., Ref. 27) gives the prediction that they will differ from ideal omnidirectional response by less than ±2 dB over the entire frequency range of interest.

Computer-generated sweeps were used as excitation signals.[28] A combination of pre- and postequalization was used to address frequency dependence in the power response of the loudspeaker and maximize the signal-to-noise ratios (S/N) of the measured impulse responses by accounting for the spectrum of the background noise and the attenuation due to air absorption.[28] Coherent averaging of five sequentially repeated measurements provided an additional gain in S/N. Comparison between averaged and unaveraged decays indicated that the decay process was sufficiently repeatable that averaging did not suppress the decay.

Because the purpose of the model was to test the algorithms rather than to model a full-scale space, neither nitrogen nor air drying was used to compensate for the additional propagation loss. Rather, the increased dissipation of the air at scaled frequencies, as computed by ISO 9613-1,[29] was modeled in the software based on measurements of atmospheric conditions made at approximately 30-min intervals. For each octave band, the dissipation coefficient used was the mean of the coefficients calculated for each frequency in the band rather than the coefficient calculated for the center frequency—the difference being significant at high frequencies.

B. Characterization and calibration

1. Absorption coefficients

As suggested by Polack *et al.*,[30] the absorption coefficients $\alpha(f)$ of the wall surfaces were measured in a 1:10 scale-model reverberation room (with geometry modeled after a validated full-scale room) constructed entirely of the same material as the auditorium model. Instead of estimating the absorption coefficient from the measured reverberation time by inverting a statistical model of energy decay (such as Sabine's), the direct-inversion method based on the computational GA proposed in Ref. 31 was used because it was believed to be more accurate. The absorption coefficients of the fiberglass batting and the individual quadratic-residue diffusers were also measured by this method in conjunction with a deep well[32] of height 8 cm. Absorption coefficients of larger objects (diffusers) were measured in the concert-hall model configured as a single-volume room, following the procedure proposed in Ref. 31.

2. Scattering coefficients

Because scattering coefficients $s(f)$ often have little effect on the rate of decay in small, proportionate rooms that are sufficiently mixing,[33] such as the nonrectangular scale-model reverberation room used here, $s(f)$'s were estimated

TABLE I. Mean absorption coefficients $\bar{\alpha}$ and mean scattering coefficients $\bar{s}$ are reported for both subrooms of the scale model for each of four full-scale-equivalent octave bands measured. Values reported are geometrical averages, calculated with rooms configured as isolated single-volume rooms (both $\bar{\alpha}$ and $\bar{s}$ are functions of aperture area when the rooms are coupled).

	125 Hz	250 Hz	500 Hz	1000 Hz
$\bar{\alpha}_1$	0.244	0.214	0.209	0.199
$\bar{\alpha}_2$	0.055	0.061	0.057	0.031
$\bar{s}_1$	0.295	0.284	0.198	0.093
$\bar{s}_2$	0.120	0.120	0.080	0.060

from decay curves measured in the auditorium model following the procedure proposed in Ref. 31. This is a much more approximate approach than the method described in ISO/DIS 17497-1 (Ref. 34), but one necessitated because the diffusers were too large to be rotated (as required) within either the reverberation-room model or the concert-hall model.

3. Adjustments

Both $\alpha(f)$ and $s(f)$ were iteratively adjusted to match predictions of RTC to decay-curve measurements made in each room of the concert-hall model when isolated from the adjacent room. However, determining $\alpha(f)$ and $s(f)$ by minimizing the error between either the predicted and measured reverberation time or the gross shapes of the predicted and measured decay curves is an ill-conditioned problem. Thus, while this iterative-adjustment procedure does remove most effects of uncertainty in $\alpha(f)$ and $s(f)$ (as discussed below), it does not ensure that the parameters used are exactly those of the actual materials. While this could be problematic if the model were to be used to predict detailed aspects of the sound field or changes associated with fine variations in source and receiver position, for the assessment of gross characteristics of the decay curve it is sufficient. Mean absorption and scattering coefficients for each of the two rooms of the scale model, calculated from the values used in the geometrical acoustics computations, are reported in Table I.

C. Uncertainty

The primary source of uncertainty in the comparison between the computed and measured results are uncertainties in the input parameters of the computer model: absorption and scattering coefficients of surfaces, air-dissipation coefficient (as calculated from temperature, relative humidity, and barometric pressure), directionality of sources and receivers, source and receiver locations, and room dimensions. Monte Carlo error analysis is required to estimate the uncertainty in the GA computer-model predictions resulting from the uncertainty in the input parameters,[35] but would be prohibitive here given the number of input parameters. In Ref. 35, such computations were performed for the same geometry with simplified boundary conditions and give some indication of the expected uncertainty given uncertainty in the inputs.

Nonetheless, for the comparisons conducted here, the uncertainty discussed above is less germane than in a conventional prediction scenario because input parameters have been iteratively adjusted in order to match the predictions of

the algorithms to the measurements conducted in each of the subrooms when configured as an isolated room. The uncertainty which is germane is that resulting from stochastic variations between individual measurements or prediction runs. Measurement uncertainty is small: $\ll 1$ dB difference between averaged and unaveraged decay curves in the time region before noise begins to affect the shape of the decay curve. Likewise, run-to-run fluctuations of the GA algorithms were found to be small ($\ll 1$ dB) for the number of cones/rays used in these simulations.[35]

D. Results and discussion

Depending on source and receiver positions, a number of different cases for testing the prediction algorithms can be defined. The notation S1R2, for example, represents the case in which the source is in room 1 and the receiver is in room 2. Because of symmetry of Green's functions, S1R2 is formally identical to the S2R1, assuming that both source and receiver are omnidirectional. However, reciprocity need not hold for errors that may arise due to the computational GA algorithm. Source ($\times$) and receiver ($+$) positions are shown in Fig. 2. One representative source position was used in each room. Two receiver positions were used in room 1, 1a near the aperture, and 1b away from the aperture. A single receiver was used in room 2.

Figures 4, 6, 8, 9, 11, 13, and 15 compare the predictions of both the original RTC algorithm (denoted by RTC) and the new algorithm (here denoted by RAY) with measured decay curves for various source-receiver combinations in each of four full-scale-equivalent octave bands: 125, 250, 500, and 1000 Hz, full-scale equivalent. Geometrical acoustic computations were carried out using both RTC and RAY algorithms with 30 000, 150 000, and 300 000 cones/rays. Truncation times for the computations were chosen to exceed the maximum time value of interest.

To assist visual comparison between the computed decay curves and the measured decay curves, noise subtraction[36] was used when Schroeder integrating the measured decay curves. In some cases, and particularly for the highest frequency band considered, noise subtraction is unable to completely remove the effects of noise. Consequently, some of the measured decay curves that are displayed show a decreasing slope at late times that is not representative of the actual behavior of the sound field in the room.

For the particular apertures chosen, the wavelengths of the frequency bands studied are small enough that diffraction is not a significant concern ($ka > 2.8$ in all cases, where $k = 2\pi/\lambda$ and $2a$ is the edge length of the square apertures; see, e.g., Ref. 37).

1. S1R1

For the S1R1a case, using 30 000 cones/rays, the predictions of the RTC algorithm are poorly matched to the scale-model measurements, while the RAY algorithm accurately predicts measured decay curves, as shown in Fig. 4. The RTC algorithm predicts an unphysical break in the decay curve, occurring between 1 and 2 s. This break culminates in
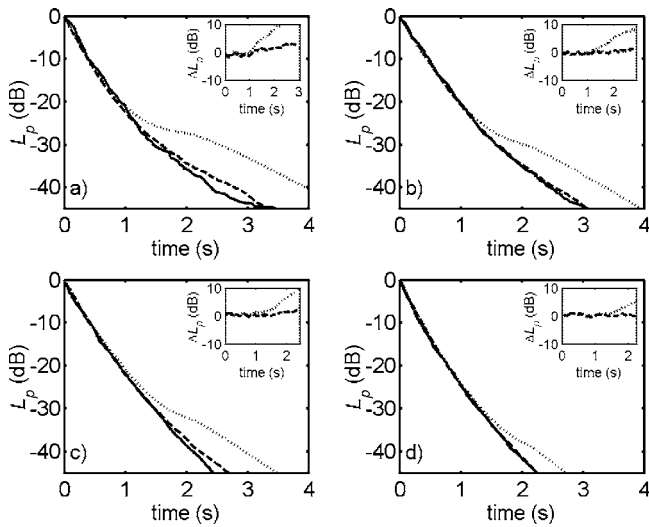
FIG. 4. For source-receiver condition S1R1a and coupling area of 50 m², decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

FIG. 6. For source-receiver condition S1R1a and coupling area of 50 m², decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 150 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

a decay curve of nearly constant decay rate for $t > 2$ s. Figure 5 indicates that predictions of the RTC algorithm first depart from scale-model measurements at $t \approx 0.6$ s, which corresponds to the onset of tail correction. Also shown in Fig. 5, as the detection sphere extends into room 2, there is an abrupt jump in detected energy due to falsely detected arrivals. The level of the error due to these false detections decreases with increasing frequency, as shown in Fig. 4. This is largely due to the increased influence of air dissipation at these higher frequencies.

Increasing the number of cones/rays decreases the initial size of the detection sphere such that the onset of the error

due to the detection sphere extending into room 2 is shifted to a later point in time, as shown in Figs. 6 and 7. Similarly, moving the receiver farther from the boundary between the subrooms shifts the onset of the error to a later point in time by increasing the size of the detection sphere necessary to extend into the adjacent subroom, as illustrated by comparison of Fig. 8 with Fig. 6.

While increasing the number of cones/rays or moving the receiver farther from the boundary between the subrooms can effectively eliminate errors in the predictions of the RTC algorithm due to the detection sphere extending into the ad-

FIG. 5. For source-receiver condition S1R1a and coupling area of 50 m², the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 30 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 30 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale-equivalent octave band.

FIG. 7. For source-receiver condition S1R1a and coupling area of 50 m², the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 150 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). Values of $\sqrt{n(t)}$ predicted by Eq. (2) are plotted assuming $V = V_1$ (triangles) and $V = V_1 + V_2$ (circles). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 150 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale-equivalent octave band.
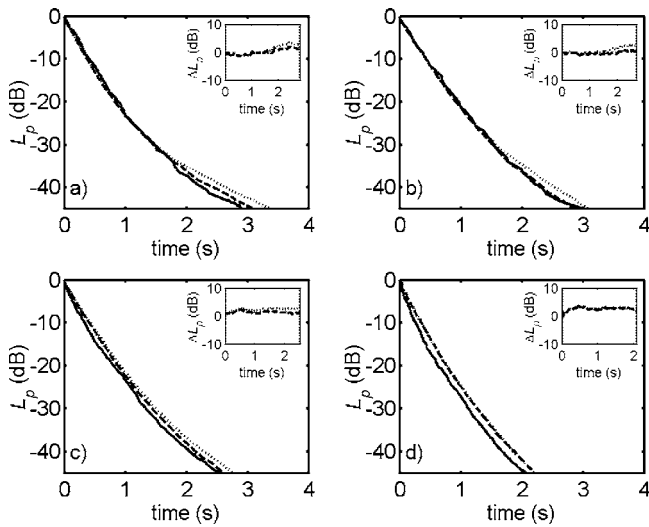
FIG. 8. For source-receiver condition S1R1b and coupling area of 50 m^2, decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

FIG. 9. For source-receiver condition S1R2 and coupling area of 50 m^2, decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$. The measured curve in the 1000-Hz octave band shows the effect of noise at late times.

jacent subroom, differences between the predictions of the RTC algorithm and the RAY algorithm remain, as shown in Figs. 6 and 8. In general, RAY predictions are better matched to the scale model measurements than RTC predictions. The reason for this is the faulty assumption of the RTC algorithm that $n(t)$ at late times in a subroom of a coupled system can be predicted from $n(t)$ at early times. (Conversely, it should be noted that, in single-volume rooms, this approach is more robust to common user errors, such as incorrect entry of geometry, than those based on the volume of the room because the effective acoustical volume is easily derived from the observed reflection growth rate whereas this volume can be difficult to compute numerically from a poorly drawn model.) Reflection density, as estimated by RTC, is too high at late times, as Fig. 8 shows. Consequently, the energy density is predicted to be too high. The least-square estimate of $n(t)$ calculated by RTC is nearly $4\pi c^3 t^2 / V_1$, where V_1 is the volume of room 1. However, at late times, $n(t)$ is better described by $4\pi c^3 t^2 / (V_1 + V_2)$, where V_2 is the volume of room 2.

2. S1R2

When the source and receiver are in different rooms, the observed error is of a different character. For the S1R2 case, using 30 000 cones/rays, the decay rate is consistently under-predicted by the RTC algorithm, as shown in Fig. 9. The early part of the decay, dominated by reflections associated with room 1, does not well predict the greater reflection density of the late part of the decay, which is dominated by reflections associated with room 2, as shown in Fig. 10. A break in the decay curve, similar to that observed in the S1R1 cases, occurs when the detection sphere extends into room 1 ($t > 1.25$ s). The level difference at this break is smaller than those observed in the S1R1 cases because the mean energy density of room 1 is lower than the mean en-

ergy density of room 2. Increasing the number of cones/rays from 30 000 to 150 000 decreases the detection sphere size, shifting the error to a later point in time. However, the estimate of reflection growth rate is still determined from the early part of the decay and, therefore, continues to underpredict the reflection density in the late decay. Consequently, the energy level of the decay curve is underpredicted by the RTC algorithm. In contrast, the RAY algorithm yields predictions of consistent accuracy.

3. S2R1

For the S2R1 case, the nature of the error differs from that observed for the converse S1R2 case, indicating that
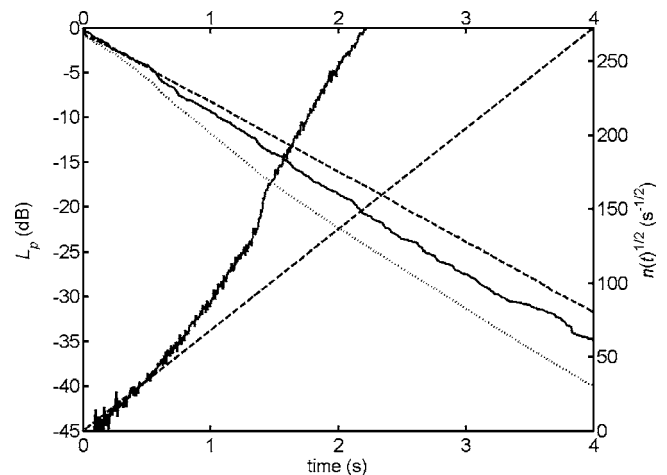
FIG. 10. For source-receiver condition S1R2 and coupling area of 50-m^2, the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 30 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 30 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale equivalent octave band.
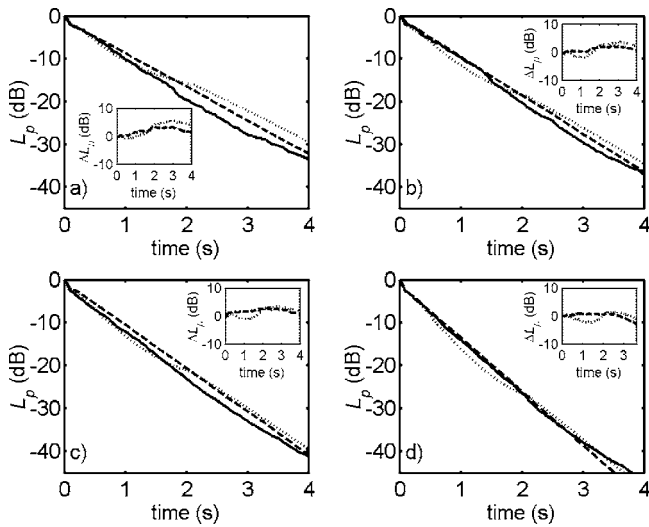
FIG. 11. For source-receiver condition S2R1a and coupling area of 50 m^2, decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

reciprocity does not hold for errors due to the algorithm. For 30 000 cones/rays, decay curves predicted by RTC display an error similar to that observed for the S1R1 cases; they have a break after which the curve is displaced upward in level due to the detection sphere extending into room 2. This is illustrated in Figs. 11 and 12 for the receiver position near the aperture (S2R1a). Increasing the number of rays shifts the point of the break later in time, as observed for the S1R2 case. For the S2R1b case, with the receiver further from the aperture, similar effects are found, but, as in the S1R1b case, they occur later in time.
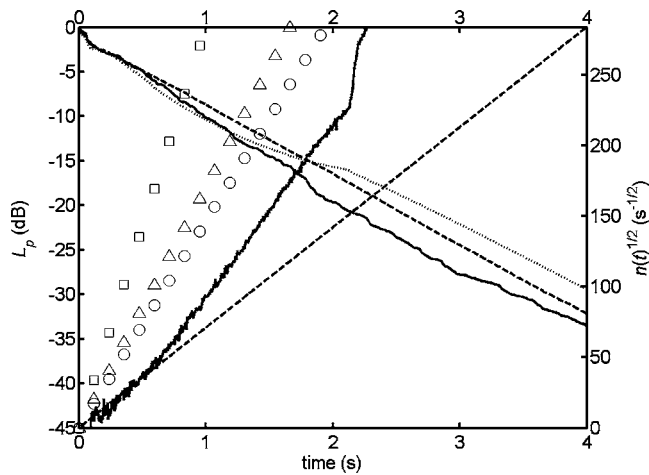
FIG. 12. For source-receiver condition S2R1a and coupling area of 50 m^2, the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 30 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). Values of $\sqrt{n(t)}$ predicted by Eq. (2) are plotted assuming $V=V_1$ (triangles), $V=V_2$ (squares), and $V=V_1+V_2$ (circles). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 30 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale-equivalent octave band.
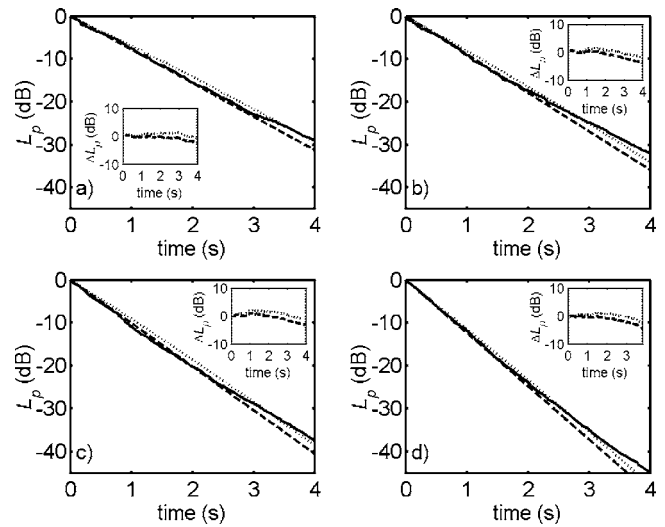
FIG. 13. For source-receiver condition S2R2 and coupling area of 50 m^2, decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t) = L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

4. S2R2

The S2R2 case differs from the prior cases in that both reflection density and decay level behavior are predominantly controlled by room 2. Using 30 000 cones/rays, the algorithms both predict the decay curve with little error, as shown in Fig. 13. For the first 20 dB of decay, the RAY algorithm is marginally more accurate than the RTC algorithm, which slightly overpredicts the level of the decay curve. Increasing the number of cones/rays to 150 000 alters the predicted decay curves very little. The observed reflection density, shown in Fig. 14, indicates that the RTC estimate of $n(t)$ is too high at large time. Reflection density
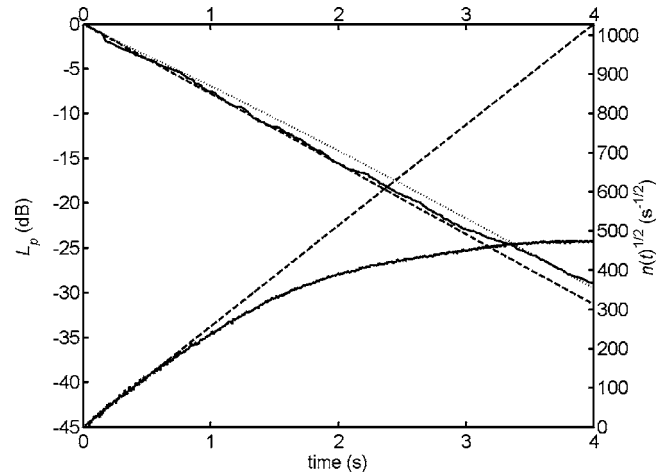
FIG. 14. For source-receiver condition S2R2 and coupling area of 50 m^2, the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 30 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 30 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale-equivalent octave band.
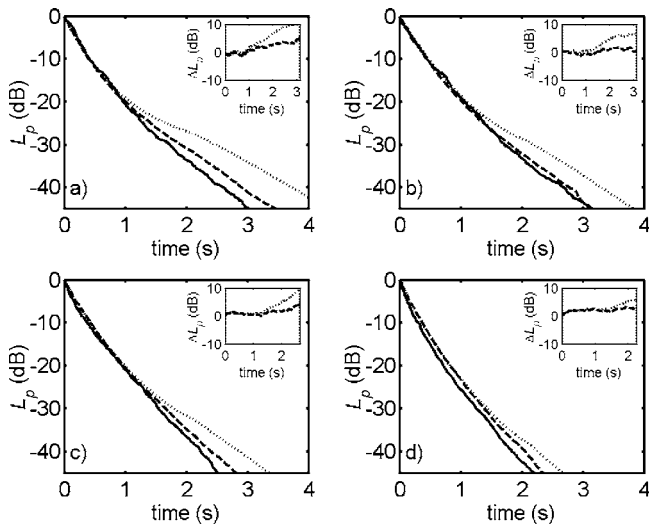
FIG. 15. For source-receiver condition S1R1a and coupling area of 70.7 m², decay-curve predictions of the RTC algorithm (dotted lines) and the revised (RAY) algorithm (dashed lines) using 30 000 cones/rays are compared with scale-model measurements (solid line) in full-scale-equivalent octave bands 125 Hz (a), 250 Hz (b), 500 Hz (c), and 1000 Hz (d). Insets show the differences in level between predictions and measurement: $\Delta L_p(t)$ $= L_p(t)_{\text{computed}} - L_p(t)_{\text{measured}}$.

FIG. 16. For source-receiver condition S1R1a and coupling area of 70.7 m², the square root of reflection density $\sqrt{n(t)}$ as computed by the RTC algorithm using 30 000 cones (solid line) is plotted on the right axis and compared with the linear estimate used by RTC for tail correction (dotted line). On the left axis, the decay-curve prediction of the RTC algorithm (dotted line) and the revised (RAY) algorithm (dashed line) using 30 000 cones are plotted together with the scale-model measurement (solid line) in the 125-Hz full-scale-equivalent octave band.

initially follows $4\pi c^3 t^2/V_2$ and then transitions to $4\pi c^3 t^2/(V_1+V_2)$. As a result of this, RTC increases the energy of late detections by too large a factor and marginally overpredicts the level of the decay curve. Increasing the number of cones cannot improve the error of the RTC algorithm in this case.

It is notable that at the point in time when the detection sphere extends into room 1, there is little influence on the decay curve predicted by RTC because both the reflection density and the energy density in room 1 are sufficiently lower than those of room 2.

5. Aperture area

Doubling the area of the aperture results in relations between computer-model predictions and measured decay curves similar to those observed for the smaller aperture, as shown in Fig. 15 for S1R1a case and 30 000 cones/rays. In those cases for which a break occurs in the RTC predictions, such as the case shown, the onset time of the break is (inevitably) the same as that observed for the small-aperture case though the level difference is smaller, confirming the speculation of Sec. II B 3. The subrooms are more strongly coupled by the larger aperture such that their respective energy densities are less disparate than they were in the more weakly coupled small-aperture condition. Thus, while the increase in detected reflections that occurs when the detection sphere extends into room 2 (as shown in Fig. 16) is similar to that observed for the small-aperture condition, the effect on the decay curve is smaller in level.

6. Discussion

The results discussed above suggest that prediction errors of the RTC algorithm observed in coupled rooms have two components. The first is the overprediction or underpre-

diction of decay energy/rate due to a changing reflection-growth rate, which violates the assumption of the tail-correction algorithm. The second is an erroneous break in the predicted decay rate due to the penetration of the detection sphere into the adjacent subroom that results in false detections and sometimes causes an abrupt upward shift in level of the decay curve. Comparisons of computational GA predictions with scale-model measurements have shown that the RAY algorithm, by circumventing the error mechanisms of the RTC algorithm, yields accurate predictions of decay curves in coupled rooms.

The accuracy of the RAY algorithm has been independently confirmed by comparison of predictions with reverberation-time measurements made in a cathedral comprising two large, reverberant rooms.[38] The comparisons, conducted by Martellotta et al.[38] (and reported in Chap. 6 of Ref. 1), show that RAY is able to predict spatial averages of T15 and T30 in the room not containing the source to within 5%, while RTC is not.

V. CONCLUDING REMARKS

Tail-correction procedures for beam-axis tracing can fail in certain cases when applied to the prediction of sound fields in systems of coupled rooms. This is partly due to the almost-universal assumption that reflection density is quadratic in time. In systems of coupled rooms reflection density is ultimately quadratic in time, but is not described by a single quadratic function through time. As a result, algorithms that assume reflection density is quadratic in time can inaccurately predict decay curves in systems of coupled rooms. These errors have been demonstrated by comparison of the predictions of a particular cone-axis-tracing algorithm (CATT-Acoustic v. 7.0 and above) with scale-model measurements of a two-room coupled system.

A second error, which has likewise been demonstrated by comparison with scale-model measurements, results from penetration of a detector sphere into an adjacent subroom. This error is unique to cone-axis tracing implemented as ray tracing with a growing detection sphere. It is most severe when the receiver is near a boundary between subrooms (causing error to occur earlier in time) or when the adjacent subroom has greater energy density than the subroom containing the receiver (causing greater error in decay-curve level).

A revised beam-axis-tracing algorithm, which addresses the problems of prediction in coupled rooms, has been investigated. The revised algorithm (introduced in CATT-Acoustic v 8.0b and above) uses a conventional ray-tracing algorithm for the late part of the decay, making tail-correction procedures unnecessary (because ray tracing inherently compensates for the failure to detect certain paths by detecting other, geometrically invalid paths) and preventing penetration of detector spheres into adjacent subrooms. This new algorithm has been shown to yield accurate predictions in the scale model used here and should have no intrinsic limitation when applied to any system of coupled rooms.

While the revised algorithm is more accurate than the original RTC algorithm when applied to systems of coupled rooms, the RTC algorithm remains preferable when applied to single-volume rooms. For single-volume rooms, both algorithms yield similar decay-curve predictions. However, the RTC algorithm yields results that are more stable and generally requires fewer beam axes. Further, RTC is preferable for auralization because it yields echograms that have a natural (quadratic) growth of reflection density, while the revised (RAY) algorithm does not at times after the radius of the detection sphere is held fixed.

ACKNOWLEDGMENTS

[1] J. E. Summers, "Reverberant acoustic energy in auditoria that comprise systems of coupled rooms," Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, 2003 [J. Acoust. Soc. Am. **114**, 2526(A) (2003)]. (Electronic copies are available from the author.)

[2] G. M. Naylor, "ODEON—Another hybrid room acoustical model," Appl. Acoust. **38**, 131–143 (1993).

[3] J. S. Anderson and M. Bratos-Anderson, "Acoustic coupling effects in St. Paul's Cathedral, London," J. Sound Vib. **236**, 209–225 (2000).

[4] U. Ayr, E. Cirillo, and F. Martellotta, "Theoretical and experimental analysis for coupled rooms: transient behavior and relevant acoustical parameters," in Proceedings of the 17th ICA, Rome (2001).

[5] L. Nijs, G. Jansens, G. Vermeir, and M. van der Voorden, "Absorbing surfaces in ray-tracing programs for coupled spaces," Appl. Acoust. **63**, 611–626 (2002); J. E. Summers, "Comments on 'Absorbing surfaces in ray-tracing programs for coupled spaces'," *ibid.* **64**, 825–831 (2003).

[6] C. D. Lyle, "Truncation error in computing reverberation by geometric acoustics," Acoust. Lett. **4**, 245–247 (1981); C. D. Lyle, "Sound decays in a semipartioned rectangular enclosure with non-uniform absorption, and comparison with geometric theory," *ibid.* **5**, 30–34 (1981).

[7] D. van Maercke, "Simulation of sound fields in time and frequency domain using a geometrical model," in Proceedings of the 12th ICA, Toronto, paper E11-7 (1986).

[8] T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics," Appl. Acoust. **38**, 161–178 (1993).

[9] B.-I. L. Dalenbäck, "Verification of prediction based on Randomized Tail-Corrected cone-tracing and array modeling," on the CD-ROM: *Berlin, March 14–19, Collected Papers, 137th Meeting of the Acoustical Society of America and the 2nd Convention of the European Acoustical Association* (ISBN 3–9804458–5–4, available from Deutsche Gesellschaft für Akustik, Fachbereich Physik, Universität Oldenburg, 26111 Oldenburg, Germany), paper 3pAAa1 (1999). [J. Acoust. Soc. Am. **105**, 1173(A) (1999).]

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," J. Acoust. Soc. Am. **65**, 943–951 (1979).

[11] J. Borish, "Extension of the image method to arbitrary polyhedra," J. Acoust. Soc. Am. **75**, 195–206 (1984).

[12] A. Krokstad, S. Strøm, and S. Sørsdal, "Calculating the acoustical room response by the use of a ray tracing technique," J. Sound Vib. **8**, 118–125 (1968).

[13] H. Lehnert, "Systematic errors of the ray-tracing algorithm," Appl. Acoust. **38**, 207–221 (1993).

[14] M. Vorländer, "Simulation of the transient and steady-state sound propagation in rooms using a new combined ray-tracing/image-source algorithm," J. Acoust. Soc. Am. **86**, 172–178 (1989).

[15] D. van Maercke and J. Martin, "The prediction of echograms and impulse responses within the Epidure software," Appl. Acoust. **38**, 93–114 (1993).

[16] B.-I. L. Dalenbäck, "Room acoustic prediction based on a unified treatment of diffuse and specular reflection," J. Acoust. Soc. Am. **100**, 899–909 (1996).

[17] J. P. Walsh and N. Dadoun, "What are we waiting for? The development of Godot, II," presented at the 103rd Meeting of the Acoustical Society of America, April 1982 [J. Acoust. Soc. Am. Suppl. 1 **82**, S5(A) (1982)].

[18] I. A. Drumm and Y. W. Lam, "The adaptive beam-tracing algorithm," J. Acoust. Soc. Am. **107**, 1405–1412 (2000).

[19] T. Funkhouser, N. Tsingos, I. Carlbom, G. Elko, M. Sondhi, J. E. West, G. Pirgali, P. Min, and A. Ngan, "A beam tracing method for interactive architectural acoustics," J. Acoust. Soc. Am. **115**, 739–756 (2004).

[20] R. H. Bolt, P. E. Doak, and P. J. Westervelt, "Pulse statistics in room acoustics," J. Acoust. Soc. Am. **22**, 328–340 (1950).

[21] M. Vorländer, "Revised relation between the sound power and the average sound pressure level in rooms and consequence for acoustics measurements," Acustica **81**, 332–343 (1995).

[22] H. Lehnert and J. Blauert, "Principles of binaural room simulation," Appl. Acoust. **36**, 259–291 (1992).

[23] A. Farina, "Ramsete- a new pyramid tracer for medium and large scale acoustic problems," in Proceedings of EURO-NOISE 95, Lyon (1995).

[24] H. Kuttruff and Th. Straßen, "Zur Abhängigkeit des Raumnachhalls von der Wanddiffusität und von der Raumform" ("On the dependence of reverberation time on the 'wall diffusion' and on room shape"), Acustica **45**, 246–255 (1980).

[25] M. R. Schroeder, "Binaural dissimilarity and optimum ceilings for concert halls: More lateral sound diffusion," J. Acoust. Soc. Am. **65**, 958–963 (1979).

[26] M. R. Schroeder, "Diffuse sound reflection by maximum-length sequences," J. Acoust. Soc. Am. **57**, 149–150 (1975).

[27] L. L. Beranek, *Acoustics* (McGraw-Hill, New York, 1954), p. 103.

[28] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**, 443–471 (2001).

[29] ISO 9613-1, Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere (1993).

[30] J. D. Polack, A. H. Marshall, and G. Dodd, "Digital evaluation of the acoustics of small models: The MIDAS package," J. Acoust. Soc. Am. **85**, 185–193 (1989).

[31] J. E. Summers, "Considerations for the modeling of seat absorption in geometrical-acoustics software," ARLO **4**, 77–82 (2003).

[32] F. Kawakami and T. Sakai, "Deep-well approach for canceling the edge effect in random incident absorption measurement," J. Acoust. Soc. Jpn. (E) **19**, 327–338 (1998).

[33] M. Hodgson, "Evidence of diffuse surface reflections in rooms," J. Acoust. Soc. Am. **89**, 765–771 (1991).

[34] ISO/DIS 17497-1, Acoustics—Sound-scattering properties of surfaces—Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room (2000).

[35] J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**, 958–969 (2004).

[36] W. T. Chu, "Comparison of reverberation measurements using Schroeder's impulse method and decay-curve averaging method," J. Acoust. Soc. Am. **63**, 1444–1450 (1978).

[37] J. E. Summers, R. R. Torres, and Y. Shimizu, "Estimating mid-frequency effects of aperture diffraction on reverberant-energy decay in coupled-room auditoria," Build. Acoust. **11**, 271–291 (2004).

[38] F. Martellotta, personal communication; see also Ref. 4.

Impedance and Brewster angle measurement for thick porous layers

Craig J. Hickey and Del Leary
National Center for Physical Acoustics, The University of Mississippi, Coliseum Drive, University, Mississippi 38677

Jean F. Allard and Michel Henry
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, 72085 Le Mans Cedex 9, France

For thin nonlocally reacting porous layers, a method derived from the work of [Chien and Soroka, J. Sound Vib. **43**, 9–20 (1975)] has been used to localize the pole of the reflection coefficient located at a complex angle close to $\pi/2$ and to measure the surface impedance at this angle. Measurements are performed with a small source/receiver separation. The method is used in the present work to measure the surface impedance of acoustically thick porous layers of high flow resistivity. Simulations show that the measured impedance, which is related to a complex angle close to $\pi/2$ like for thin porous layers, is close to the impedance at grazing incidence. It is also shown that for semi-infinite layers the method provides a measure of the Brewster angle of the medium. Measurements of the cosine of the complex angle and of the related impedance on two granular media of high flow resistivity, Ottawa sand and glass beads, are in a reasonable agreement with predictions for frequencies above 1 kHz for a source/receiver separation of 30 cm. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1979447]

I. INTRODUCTION

In the present work, a method previously developed to evaluate the surface impedance of thin porous layers[1,2] is used for acoustically thick porous layers of granular porous materials with a high flow resistivity. An acoustically thick porous layer is characterized by a strong damping of the waves when they propagate from one side to the opposite side of the layer. If the frame is motionless, this condition corresponds to Im $k_1 l \gg 1$, where l is the thickness of the layer and k_1 is the wave number in the air that saturates the frame. The quantity Im k_1 increases with frequency and with the flow resistivity of the medium. Many porous media, concretes, soils, and some sound absorbing media, exist in layers with large or semi-infinite thickness. The method developed here consists of a very simple experimental setup that allows for the *in situ* measurement of the surface impedance on samples with an area around 1 m^2, using frequencies larger than 1 kHz.

The method is based on the classical expression of the reflected monopole field close to grazing incidence. This expression is widely used for outdoor ground impedance measurements with large source receiver offsets.[3] This underlying equation is only valid if the reflection coefficient has a pole θ_p for some value of the angle of incidence θ that is close to $\pi/2$. Let Z and $Z_s(\theta)$ be the characteristic impedance of the free air, and the surface impedance of the layer, respectively. It is shown in Sec. II that $\cos \theta_p = -Z/Z_s(\theta_p)$. For thin layers, one pole can be present close to $\pi/2$ because $|Z_s|/Z \gg 1$ when l is sufficiently small. For thick layers the condition $|Z_s|/Z \gg 1$ is obtained when the flow resistivity is sufficiently large.

Thick porous layers can have an impedance very close to the one of a semi-infinite layer. The pole for a semi-infinite layer has properties very different from the ones of the pole related to a thin layer. It is recalled in this paper that the pole θ_p and the zero θ_z of the semi-infinite layer are related to opposite cosines, $\cos \theta_p = -\cos \theta_z$. There are strong analogies between the zero of the semi-infinite acoustic layer and the Brewster angle of total refraction of electromagnetic waves.[4] In Sec. II the theory for thin layers is reviewed as well as the approach used for pole localization and impedance measurement. In Sec. III the theory is modified to describe semi-infinite layers. Simulated measurements are performed in Sec. IV to illustrate the main limitations of the Chien and Soroka model and our method. Measurements of reduced impedance on Ottawa sand and glass beads are compared to theoretical predictions. Section V contains the conclusions.

II. THEORETICAL BACKGROUND

A. Generalization of the work by Chien and Soroka to the case of thin porous layers

A classical method of measuring porous ground surface impedance is based on the expression of the reflected monopole field pressure obtained by Chien and Soroka.[5] For locally reacting media the reflected pressure is given by[5]

$$p_r = \frac{\exp(ikR_1)}{R_1}[v(\cos \theta_0) + (1 - v(\cos \theta_0))F(p_e)], \quad (1)$$

where $F(p_e)$ is given by
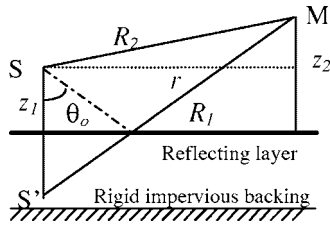
FIG. 1. Sketch of the experiment, the monopole source S, its image S', and the receiver at M above the porous layer.

$$F(p_e) = 1 + i\sqrt{\pi}p_e \exp(-p_e^2)\text{erfc}(-ip_e), \qquad (2)$$

and p_e is the numerical distance given by

$$p_e = \left(\frac{1}{2}ikR_1\right)^{1/2}\left(\frac{Z}{Z_s} + \cos\theta_0\right). \qquad (3)$$

A sketch of the measurement setup is presented in Fig. 1. R_1 is the distance from the image source S' to the receiver at M, R_2 is the distance from the source S to the receiver, r is the horizontal distance from the source to the receiver, z_1 and z_2 are the heights of the source and the receiver above the impedance surface, θ_0 is the specular angle of incidence, v is the reflection coefficient, Z and k are the characteristic impedance and the wave number in the free air, and Z_s is the surface impedance of the porous medium. The notation θ_0 is used for the specular angle of incidence related to a point source and θ for the angle of incidence of a plane wave. In the present work the time dependence is $\exp(-i\omega t)$. In general, the reflection coefficient for a plane wave at an angle θ, on a locally reacting medium, is given by

$$v(\cos\theta) = \frac{Z_s - Z/\cos\theta}{Z_s + Z/\cos\theta}. \qquad (4)$$

A pole of the reflection coefficient exists at a complex angle of incidence θ_p given by

$$\cos\theta_p = -\frac{Z}{Z_s}. \qquad (5)$$

Equations (3) and (4) can be rewritten using Eq. (5) as

$$p_e = \left(\tfrac{1}{2}ikR_1\right)^{1/2}(-\cos\theta_p + \cos\theta_0) \qquad (6)$$

and

$$v(\cos\theta) = \frac{\cos\theta + \cos\theta_p}{\cos\theta - \cos\theta_p}. \qquad (7)$$

Equation (1) is valid, as indicated in Ref. 5, for $kR_1 \gg 1$ and if both the pole θ_p and the specular angle θ_0 are close to $\pi/2$, which implies near grazing incidence and $|Z/Z_s| \ll 1$. When Eqs. (1), (2), (6), and (7) are used, it appears, like in the work by Brekhovskikh and Godin[6] for similar equations, and in previous works,[1,2] that the evaluation of Z_s from pressure measurements using Eq. (1) is primarily an evaluation of $\cos\theta_p$ followed by the evaluation of Z_s using Eq. (5).

The case of nonlocally reacting surfaces was not considered by Chien and Soroka.[5] For nonlocally reacting porous layers with a motionless frame, Z_s is dependent on the angle of incidence and Eq. (5) is replaced by

$$\cos\theta_p = -\frac{Z}{Z_s(\theta_p)}. \qquad (8)$$

The angle-dependent surface impedance for a layer of thickness l is given by

$$Z_s(\theta) = \frac{iZ_1}{\phi\cos\theta_1}\cot(k_1 l \cos\theta_1), \qquad (9)$$

where Z_1 and k_1 are the characteristic impedance and the wave number of the air saturating the porous medium, ϕ is the porosity, and θ_1 is the refraction angle. The refraction angle θ_1 is given by

$$\cos\theta_1 = \left(1 - \frac{1}{n^2} + \frac{\cos^2\theta}{n^2}\right)^{1/2}, \qquad (10)$$

where $n = k_1/k$ is the index of refraction. For thin nonlocally reacting layers, there are an infinite number of poles with a large imaginary part.[2] One pole, θ_p, will exist close to $\pi/2$ if, in Eq. (8), $Z/|Z_s(\theta_p)| \ll 1$. It has been shown in Ref. 1 that for nonlocally reacting thin porous layers with a motionless frame, Eqs. (1) and (2) can be used to evaluate the reflected field p_r, but Eq. (6) must be used for the numerical distance with $\cos\theta_p$ given by Eq. (8). The impedance Z_s that can be obtained from pressure measurements over thin layers from Eqs. (1)–(3) is $Z_s(\theta_p)$, close to $Z_s(\pi/2)$.

B. Pole localization and impedance measurement for thin porous layers

The method of Allard et al.[1] is based on Eq. (1). An equivalent version of Eq. (1) is used where v and p_e are replaced by their expressions given by Eqs. (6) and (7). The use of Eq. (7) for v is an approximation if the layer is not locally reacting, but this approximation is justified if $|\theta_0 - \theta_p|$ is sufficiently small such that the surface impedances $Z_s(\theta_p)$ and $Z_s(\theta_0)$ are not noticeably different. Equation (1) can be rewritten

$$p_r = \frac{\exp(ikR_1)}{R_1}\left[1 - \sqrt{2\pi kR_1}\exp\left(\frac{3\pi i}{4}\right)\right.$$

$$\left. \times\cos\theta_p \exp(-p_e^2)\text{erfc}(-ip_e)\right]. \qquad (11)$$

In this equation, $\cos\theta_p$ [or $Z_s(\theta_p)$] is the one parameter that characterizes the porous layer. A measurement of the reflected pressure field p_r is obtained by measuring the total field successively over the porous layer (p_t^l) and over a rigid impervious surface (p_t^r) at the same distances z_1, z_2, and r. Measurements are performed at small source-receiver distance, from 20 to 30 cm, close to grazing incidence to satisfy the condition θ_0 close to $\pi/2$. The total fields p_t^r and p_t^l with a point source of unknown amplitude K are given by

$$p_t^r = K\left[\frac{\exp(ikR_1)}{R_1} + \frac{\exp(ikR_2)}{R_2}\right], \qquad (12)$$

Hickey et al.: Impedance and Brewster angle measurement

TABLE I. Parameters for the porous materials

Parameters	Ottawa sand	Glass beads	Material 3
porosity ϕ	0.393	0.41	0.95
flow resistivity σ (N m^{-4} s)	94000.0	150000.0	40000.0
thermal permeability k_0' (m^2)	0.6×10^{-9}	3.8×10^{-10}	1.2×10^{-9}
viscous dimension Λ (μm)	32.4	26.5	50.0
thermal dimension Λ' (μm)	97.2	79.5	150.0
tortuosity α_∞	1.54	1.5	1.1
thickness l (cm)	8.0	3.0	8.0

$$p_t^l = K\left[p_r + \frac{\exp(ikR_2)}{R_2} \right]. \tag{13}$$

The reflected pressure field p_r is evaluated from the measured ratio p_t^l/p_t^r using

$$p_r = \frac{p_t^l}{p_t^r}\left[\frac{\exp(ikR_1)}{R_1} + \frac{\exp(ikR_2)}{R_2} \right] - \frac{\exp(ikR_2)}{R_2}. \tag{14}$$

From Eq. (11) $\cos \theta_p$ is given by

$$\cos \theta_p = [1 - R_1 p_r \exp(-ikR_1)]/[(2\pi kR_1)^{1/2}$$
$$\times \exp(3\pi i/4)\exp(-p_e^2)\mathrm{erfc}(-ip_e)], \tag{15}$$

At small source receiver distances, and for small $|\cos \theta_p - \cos \theta_0|$, the numerical distance p_e is generally small. The term $\mathrm{erfc}(-ip_e)$ can be approximated for small argument as

$$\mathrm{erfc}(-ip_e) = 1 + \frac{2i}{\sqrt{\pi}}p_e\left(1 + \frac{p_e^2}{3} + \frac{p_e^4}{2!5} + \frac{p_e^6}{3!7} + \cdots \right), \tag{16}$$

where p_e is given by Eq. (6). Simple iterative methods[1,2] can be used to evaluate $\cos \theta_p$ with Eqs. (6) and (15) when p_r is evaluated using Eq. (14) from measurements of p_t^r and p_t^l. The surface impedance $Z_s(\theta_p)$ with θ_p close to $\pi/2$ is obtained from Eq. (8).

Measured $\cos \theta_p$ are generally close to the predicted values obtained with the motionless frame model. The full Biot theory[7,8] must be used when frame resonances are excited. It has been verified experimentally[1,2] that for thin porous layers with $|k_1 l| \ll 1$, the pole θ_p close to $\pi/2$ has a Re $\sin \theta_p > 1$ and a surface wave contribution can exist at large numerical distances. The z_2 dependence of the surface wave related to the pole is $\exp(ikz_2 \cos \theta_p)$ with Im $\cos \theta_p > 0$, Re $\cos \theta_p < 0$, Im $\cos \theta_p \gg -$Re $\cos \theta_p$. The wave exponentially decreases with the distance z_2 from the porous surface. It is related to a trapped mode in the porous layer.[9] The reflection of electromagnetic TM waves by a thin grounded dielectric presents a strong analogy with this case.[4,9]

III. POLES AND ZEROS OF THE REFLECTION COEFFICIENT FOR SEMI-INFINITE LAYERS

Two granular media with a high flow resistivity, Ottawa sand and glass beads, are studied. The grain size distribution of the sand has a broad peak around 0.5 mm. The second medium is a mixture of glass beads of diameters ranging from 355 to 425 μm. The parameters that characterize both materials to be used in the predictions are given in Table I.
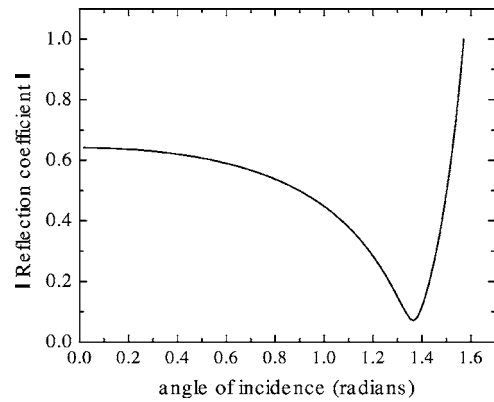
FIG. 2. Predicted modulus of the reflection coefficient at 4 kHz for a semi-infinite layer of Ottawa sand.

Tortuosity α_∞ and characteristic dimensions Λ and Λ' have been evaluated from time of flight ultrasonic measurements[10] under the simplifying hypothesis $\Lambda' = 3\Lambda$. Porosity and flow resistivity have been measured with the methods described in Refs. 11 and 12. Reasonable values of the thermal permeability are used. The thickness of the sand layer is $l = 8$ cm. The surface impedance is predicted by using the model of Johnson et al.[13] for the effective density ρ_1 and the model suggested by Lafarge[14] for the incompressibility χ_1 of the equivalent fluid. In Eq. (9), $k_1 = \omega(\rho_1/\chi_1)^{1/2}$ and $Z_1 = (\rho_1\chi_1)^{1/2}$. For frequencies higher than 1 kHz, the surface impedance of the layer given by Eq. (9) is very close to the surface impedance of the semi-infinite sand layer Z_s^∞ given by

$$Z_s^\infty = Z_1/(\phi \cos \theta_1). \tag{17}$$

The modulus of the reflection coefficient for the semi-infinite layer, calculated at 4 kHz from Eqs. (4) and (17) as a function of the angle of incidence θ, is shown in Fig. 2. A minimum in the reflection coefficient exists at $\theta_m = 1.365$. The unique pole for a semi-infinite layer is located at θ_p such that

$$\sin \theta_p = \left[\frac{(\rho_1/(\rho_0\phi))^2 - n^2}{(\rho_1/(\rho_0\phi))^2 - 1} \right]^{1/2} \tag{18}$$

and

$$\cos \theta_p = -\left[\frac{n^2 - 1}{(\rho_1/(\rho_0\phi))^2 - 1} \right]^{1/2}, \tag{19}$$

where ρ_0 is the air density and n is the index of refraction. Comparing the predicted $\cos \theta_p = -0.203 + i2.9 \times 10^{-2}$ from Eq. (19) to $\cos \theta_m = 0.204$ shows that it is opposite to Re $\cos \theta_p$ to a good approximation. This is due to the structure of the reflection coefficient v in Eq. (4). A given $\sin \theta_p$ is related to two opposite values for $\cos \theta$: $\cos \theta_p$, which is related to a zero of the denominator, and $\cos \theta_z = -\cos \theta_p$, which is related to a zero of the numerator, and a zero of v. The zero and the pole are related to opposite $\cos \theta$. The imaginary part of $\cos \theta_p$ is small and $|\cos \theta_z - \cos \theta_m|$ is small, thereby the localization of the minimum. The strong analogies between this zero and the zero of the reflection coefficient of a semi-infinite dielectric

layer (or a conducting layer of sufficient thickness) for electromagnetic transverse magnetic waves are discussed in Ref. 9. For the electromagnetic waves, the zero corresponds to the Brewster angle of the medium, and the hypothetical surface wave related to the pole is the Zenneck wave.[4] This wave does not exist at large numerical distances because $\mathrm{Re}\sin\theta_p < 1$. Raspet and Baird[15] have shown that $\mathrm{Re}\sin\theta_p < 1$ is not always true for semi-infinite porous layers described using the empirical laws of Delany and Bazley.[16] However, this condition has always been verified for semi-infinite porous layers when the causal equivalent densities and incompressibilities of the present work are used, and for the measurements described in the next section. However, we have not found any general proof that $\mathrm{Re}\sin\theta_p < 1$. A zero and a pole related to the same $\sin\theta_p$ is a general consequence of Eq. (4) and exist also for thin layers, but in this case $\mathrm{Re}\sin\theta_p > 1$, and there is no real angle θ_m close to θ_z. For semi-infinite layers, the complex Brewster angle can be partially localized with the Tamura method[17] of measuring the reflection coefficient at oblique incidence, from the minimum of the modulus of the reflection coefficient, but $\mathrm{Im}\cos\theta_p$ is not directly measurable with this method.

If the frame displacements are taken into account, another pole is present for porous frames saturated by a light fluid,[18,19] which is very close to the Rayleigh pole. The velocity of the Rayleigh waves in porous foams[20] is much smaller than the sound speed in air, and the pole is related to a complex angle of incidence far from $\pi/2$. This is probably the case for the unconsolidated media in the present study where no manifestation of this pole has been observed.

IV. VALIDITY OF THE METHOD FOR THICK LAYERS

A. Simulations

Equations (1)–(3) or the equivalent formulation with Eq. (11) gives an asymptotic estimation of the reflected field which is valid only for sufficiently large kr. A simulated measurement of $\cos\theta_p$ for a semi-infinite layer of sand using Eq. (15) is performed for $r=15$ cm, $z_1=z_2=0.5$ cm. The pressure p_r is calculated without approximations using the Sommerfeld integral[6]

$$p_r = ik \int_0^\infty \frac{q\,dq}{\mu} v(\mu) J_0(kqr) \exp(ik\mu(z_1+z_2)), \qquad (20)$$

where $\mu = \sqrt{1-q^2}$, $\mathrm{Im}\,\mu \geqslant 0$. The reflection coefficient v is given by Eq. (4) and the surface impedance Z_s by Eq. (9). The first test concerns the asymptotic behavior of the approximate expressions, Eq. (1) or (11). For the test, a layer of Ottawa sand of thickness $l=8$ cm, described by the parameters given in Table I, is used. The exact $\cos\theta_p$ is compared to the result of the simulated measurement for $r=15$ cm, $z_1=z_2=0.5$ cm, as a function of frequency in Fig. 3. The difference between the exact value and the simulated measurement slowly increases when frequency decreases. At 500 Hz, $kr=1.4$, the exact value and the simulated measurement remain surprisingly close to each other. The second test concerns the fact that as indicated in Ref. 5, Eq. (1) and then Eq. (11) are valid only for θ_0
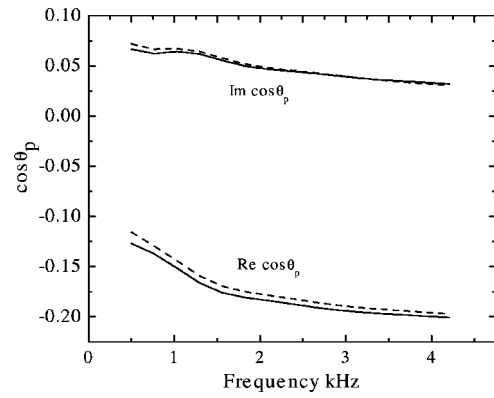
FIG. 3. $\cos\theta_p$ as a function of frequency for a layer of Ottawa sand of thickness $l=8$ cm. Prediction:———; simulated measurement $r=15$ cm, $z_1=z_2=0.5$ cm:---.

close to $\pi/2$. In Fig. 4 the exact value and the simulated measurement are compared at 4 kHz for $r=15$ cm, and $z_1=z_2$ varying from 0.25 to 2.5 cm. The real part of $\cos\theta_p$ is predicted with a good precision over the whole interval, but the error on $\mathrm{Im}\cos\theta_p$ becomes considerable for $z_1+z_2 > 2$ cm. A simulation has been performed to test the validity of the method for large $\cos\theta_p$. For semi-infinite layers, $|\cos\theta_p|$ is a decreasing function of the flow resistivity. In order to increase $|\cos\theta_p|$ a porous medium with a lower flow resistivity, $\sigma=40\,000$ N m^{-4} s, and a thickness $l=8$ cm, has been used in the simulation. The parameters which characterize the material, denoted as material 3, are given in Table I. The simulated measurement of $\cos\theta_p$ and the exact value are compared in Fig. 5 for $z_1=z_2=5$ mm, $r=15$ cm, and frequencies varying from 0.5 to 5 kHz. The error in $\mathrm{Im}\cos\theta_p$ can be important at high frequencies when $|\cos\theta_p|$ is too large, similar to the case for a large $\cos\theta_0$. A last point concerns the result of the measurements, when the measurement has been performed with a good precision on a material with a high flow resistivity. A large flow resistivity is related to a weak dependence of the surface impedance on the angle of incidence and the results of outdoors measurements around grazing incidence are generally considered as surface impedances at normal incidence. In Fig. 6 the simulated measurement of $Z_s(\theta_p)/Z$ for a layer of thickness $l=8$ cm of Ottawa sand
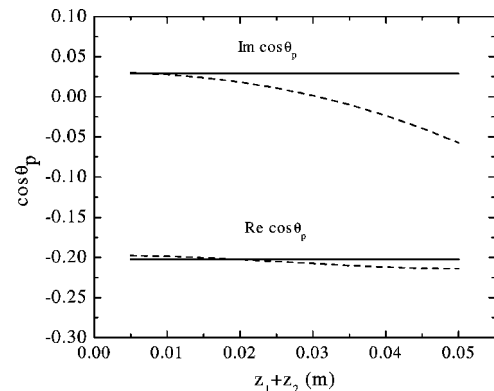
FIG. 4. $\cos\theta_p$ at 4 kHz for a layer of Ottawa sand of thickness $l=8$ cm. Prediction:———; simulated measurement $r=15$ cm, 0.5 cm$<z_1+z_2$ <5 cm:---.
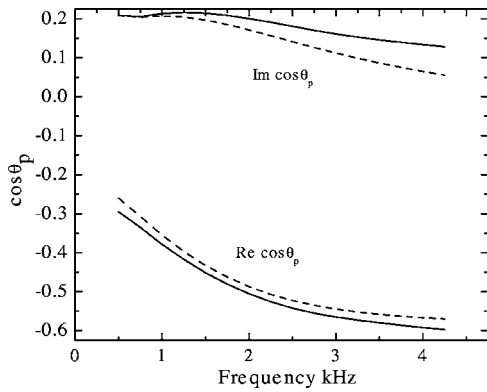
FIG. 5. $\cos \theta_p$ as a function of frequency, material 3. Prediction:———;
simulated measurement $r=15$ cm, $z_1=z_2=0.5$ cm:---.

with $z_1=z_2=5$ mm, $r=15$ cm is compared to the predicted
reduced impedances at normal and at grazing incidence.
The predicted measurement in the high-frequency range is
noticeably different from the impedance at normal inci-
dence and close to the impedance at grazing incidence.

B. Measurements

Measurements have been performed on two granular
media, a layer of Ottawa sand of thickness $l=8$ cm and a
layer of glass beads of thickness $l=3$ cm. The source is a
pipe fed by a compression driver. The internal diameter of
the pipe is 4 mm. The receiver is a $\frac{1}{4}$-in. Bruel and Kjaer
microphone. Different source-receiver distances have been
used, ranging from 15 to 30 cm. As shown in Fig. 4, the
systematic error on Im $\cos \theta_p$ can be important if z_1+z_2 is too
large. Some measurements have been performed with the
source very close to the surface (z_1 small). In this case there
was a systematic dependence of the result on the source re-
ceiver distance, probably due to the fact that the source is not
an ideal point source and is modified when it is too close to
the reflecting surface. A distance of 3 mm between the pipe
and the surface was sufficient to avoid this systematic effect.
The sound emission at the end of the pipe is not noticeably
modified for this distance when the rigid impervious surface
is replaced by the porous layer. Predictions of $\cos \theta_p$ for the
sand and measurements for $r=30$ cm, $z_1=4$ mm, and z_2
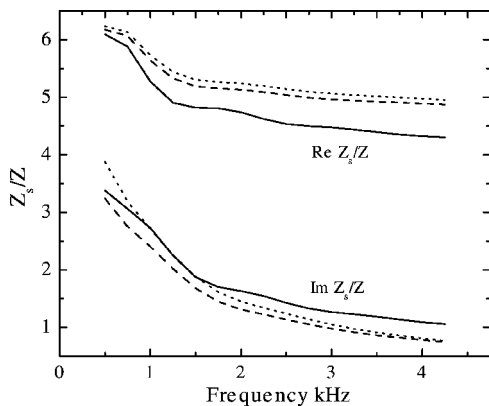
FIG. 6. Z_s/Z as a function of frequency for a layer of Ottawa sand of
thickness $l=8$ cm. Simulated measurement:· · ·; surface impedance at grazing
incidence:---, and surface impedance at normal incidence———.
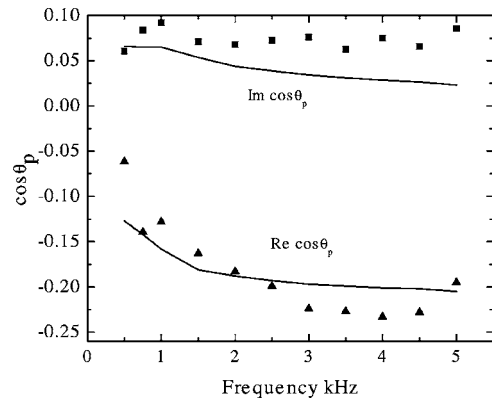
FIG. 7. $\cos \theta_p$ for a layer of Ottawa sand of thickness $l=8$ cm.
Prediction:———; measurement in an anechoic room $r=30$ cm, z_1
$=0.4$ cm, $z_2=0.7$ cm; real part ▲ ▲; imaginary part ■ ■.

$=8$ mm are presented in Fig. 7. At 0.5 kHz, the error on
Re $\cos \theta_p$ is large. This error is smaller at higher frequencies,
but the measured Im $\cos \theta_p$ is systematically larger than the
predicted one. Measurements with the same geometry and
predictions of the reduced surface impedance $Z_s(\theta_p)/Z$ of the
layer of glass beads are presented in Fig. 8. At 1 kHz, the
error on Im $Z_s(\theta_p)/Z$ is large. This error becomes smaller at
higher frequencies and a good agreement between prediction
and measurement is obtained. At frequencies higher than
1 kHz a good order-of-magnitude measurement of $\cos \theta_p$ and
$Z_s(\theta_p)$ can be obtained for the studied materials, which have
a fairly high surface impedance. Higher surface impedances
could probably be measured with a source that more closely
approximates a point source.

V. CONCLUSION

A method previously developed to evaluate the surface
impedance of thin porous layers close to grazing incidence
and with small source-receiver distances is used for thick
layers of granular porous materials with a high flow resistiv-
ity. The evaluated impedance is the only acoustical param-
eter needed to predict the reflected monopole pressure field
for specular angles of incidence close to $\pi/2$. For a locally
reacting surface the evaluated impedance is the surface im-
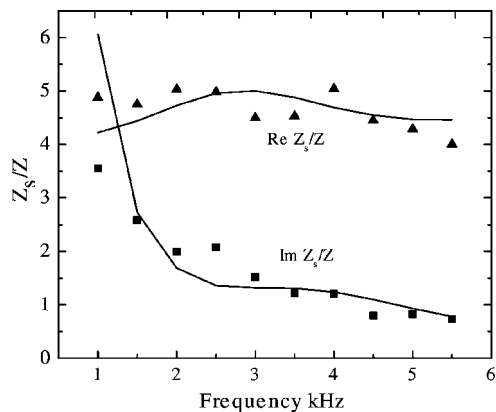
FIG. 8. Reduced impedance $Z_s(\theta_p)/Z$ for a layer of glass beads of thickness
$l=3$ cm. Prediction:———; measurement in an anechoic room $r=30$ cm,
$z_1=0.4$ cm, $z_2=0.7$ cm; real part ▲ ▲; imaginary part ■ ■.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Hickey *et al.*: Impedance and Brewster angle measurement   1507

pedance independent on the angle of incidence. For a thick porous layer with a high flow resistivity, simulations show that the evaluated impedance is close to the impedance at grazing incidence and can noticeably differ from the impedance at normal incidence. There is a pole of the reflection coefficient at the complex angle of incidence associated to the evaluated impedance. For sufficiently thick porous layers this pole is related in a straightforward way to the Brewster angle of the medium, and impedance measurement simultaneously leads to an evaluation of the pole and the Brewster angle. Measurements of the surface impedance close to grazing incidence can be performed on materials with a large flow resistivity for areas of about 1 m², at frequencies larger than 1 kHz, with a very simple experimental setup. The evaluation is performed with a classical expression of the reflected monopole field close to grazing incidence. This expression which is widely used for outdoors ground impedance measurements with large source receiver distances is shown to be applicable to smaller source receiver distances at higher frequencies for acoustically thick porous layers.

ACKNOWLEDGMENTS

APPENDIX: DERIVATION OF REFLECTED PRESSURE

The general formulation of the "reference integral method" by Brekhovskikh and Godin[6] is used in Ref. 1 to adapt the pole subtraction method for the case of nonlocally reacting porous layers with a finite thickness. The asymptotic expression Eq. (11) for the reflected pressure p_r has been obtained, and the different approximations which can lead to the simplified expressions Eqs. (1), (2), (6), and (7) for thin porous layers are indicated. The calculation of Ref. 1 can be simplified for the case of semi-infinite layers. The reflection coefficient [Eq. (A2) in Ref. 1] can be written

$$v = 1 - 2\sqrt{n^2 - q^2}\left[\sqrt{n^2 - q^2} + \frac{Z_1}{\phi Z}n\sqrt{1 - q^2}\right]^{-1}, \quad (A1)$$

where $q = \sin \theta$ and $F(q)$ is now given by

$$F(q) = -2\sqrt{\frac{q(n^2 - q^2)}{1 - q^2}}\left[\sqrt{n^2 - q^2} + \frac{Z_1}{\phi Z}n\sqrt{1 - q^2}\right]^{-1}. \quad (A2)$$

Let $G(q)$ be the factor in brackets in Eq. (A2). The new parameter a [Eq. (A11) in Ref. 1] is given by

$$a = -2\sqrt{\frac{q_p(n^2 - q_p^2)}{1 - q_p^2}}\frac{1}{G_q'(q_p)}, \quad (A3)$$

$$G_q'(q_p) = -\frac{q_p}{\sqrt{n^2 - q^2}} - \frac{Z_1}{\phi Z}n\frac{q_p}{\sqrt{1 - q_p^2}}. \quad (A4)$$

When $|n|$ is sufficiently large, θ_p given by Eqs. (18) and (19) is close to $\pi/2$ and the first term on the right-hand side of

Eq. (A4) can be neglected. Then a can be written, like in Ref. 1,

$$a = 2Z/(Z_s(\theta_p)\sqrt{q_p}), \quad (A5)$$

and p_r has the same expression as in Ref. 1,

$$p_r = \frac{\exp(ikR_1)}{R_1}$$
$$\times \left\{v(\cos \theta_0) + \frac{Z\sqrt{\sin \theta_0}}{Z_s(\theta_p)\sqrt{\sin \theta_p}\sin[(\theta_p - \theta_0)/2]}C\right\}, \quad (A6)$$

$$C = 1 + i\sqrt{\pi}s_p\sqrt{kR_1}\exp(-kR_1s_p^2)\mathrm{erfc}(-i\sqrt{kR_1}s_p), \quad (A7)$$

$$s_p = \exp\frac{i\pi}{4}\sqrt{2}\sin\frac{\theta_p - \theta_0}{2}. \quad (A8)$$

In Eq. (A6), $Z/Z_s(\theta_p)$ can be replaced by $-\cos \theta_p$. If θ_0 and θ_p are close to $\pi/2$, $(\sin \theta_0/\sin \theta_p)^{1/2}$ can be removed in Eq. (A6) and in Eq. (A8) $\sin((\theta_p - \theta_0)/2)$ can be replaced by $(\cos \theta_0 - \cos \theta_p)/2$. The term $-2\cos \theta_p/(\cos \theta_0 - \cos \theta_p)$ corresponds to $1 - v$ in Eq. (1). The conditions of validity of Eq. (1) are, like in Ref. 5, θ_0 and θ_p close to $\pi/2$.

[1] J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurement around grazing incidence for nonlocally reacting thin porous layers," J. Acoust. Soc. Am. **113**, 1210–1215 (2003).
[2] J. F. Allard, V. Gareton, M. Henry, G. Jansens, and W. Lauriks, "Impedance evaluation from pressure measurements near grazing incidence for nonlocally reacting porous layers," Acust. Acta Acust. **89**, 595–603 (2003).
[3] K. Attenborough, "Review of ground effects on outdoor sound propagation from continuous broadband point sources," Appl. Acoust. **24**, 289–319 (1988).
[4] R. E. Collin, *Field Theory of Guided Waves* (McGraw-Hill, New York, 1960).
[5] C. F. Chien and W. W. Soroka, "Sound propagation along an impedance plane," J. Sound Vib. **43**, 9–20 (1975).
[6] L. M. Brekhovskih and O. A. Godin, *Acoustics of Layered Media II, Point Source and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 1992).
[7] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid, I Low frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[8] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid, II Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[9] J. F. Allard, M. Henry, J. Tizianel, L. Kelders, and W. Lauriks, "Surface waves over bead layers," J. Acoust. Soc. Am. **105**, 3021–3025 (1999).
[10] J. F. Allard, B. Castagnede, M. Henry, and W. Lauriks, "Evaluation of tortuosity in acoustic porous materials saturated by air," Rev. Sci. Instrum. **65**, 754–755 (1994).
[11] M. Stinson and G. A. Daigles, "Electronic system for the measurement of flow resistance," J. Acoust. Soc. Am. **83**, 2422–2428 (1988).
[12] Y. Champoux, M. R. Stinson, and G. A. Daigles, "Air-based systems for the measurement of porosity," J. Acoust. Soc. Am. **89**, 910–926 (1991).
[13] 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).
[14] D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[15] R. Raspet and G. E. Baird, "The acoustic surface wave above a complex impedance ground surface," J. Acoust. Soc. Am. **85**, 638–640 (1989).
[16] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).

[17]M. Tamura, "Spatial Fourier transform method of measuring reflection coefficient at oblique incidence," J. Acoust. Soc. Am. **88**, 2259–2264 (1990).

[18]S. Feng and D. L. Johnson, "High-frequency acoustic properties of a fluid/porous solid interface. I. New surface mode," J. Acoust. Soc. Am. **74**, 906–914 (1983).

[19]S. Feng and D. L. Johnson, "High-frequency acoustic properties of a fluid/porous solid interface. II. The 2D reflection coefficient Green's function," J. Acoust. Soc. Am. **74**, 915–924 (1983).

[20]J. F. Allard, G. Jansens, G. Vermeir, and W. Lauriks, "Frame-borne surface waves in air-saturated porous media," J. Acoust. Soc. Am. **111**, 690–696 (2002).

A comparison of two different sound intensity measurement principles[a)]

Finn Jacobsen[b)]
Acoustic Technology, Ørsted•DTU, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

Hans-Elias de Bree[c)]
Microflown Technologies BV, P.O. Box 300, 6900 AH Zevenaar, The Netherlands

The dominating method of measuring sound intensity in air is based on the combination of two pressure microphones. However, a sound intensity probe that combines an acoustic particle velocity transducer with a pressure microphone has recently become available. This paper examines, discusses, and compares the two measurement principles with particular regard to the sources of error in sound power determination. It is shown that the phase calibration of intensity probes that combine different transducers is very critical below 500 Hz if the measurement surface is very close to the source under test. The problem is reduced if the measurement surface is moved further away from the source. The calibration can be carried out in an anechoic room. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1984860]

I. INTRODUCTION

The most successful method of measuring sound intensity in air is the "two-microphone" (or "*p-p*") method, which makes use of two closely spaced pressure microphones and relies on a finite-difference approximation to the sound pressure gradient.[1] Both the IEC standard on instruments for the measurement of sound intensity[2] and the corresponding North American ANSI standard[3] deal exclusively with *p-p* measurement systems. The alternative "*p-u*" method, which involves combining a pressure microphone with a particle velocity transducer, has been hampered by the absence of reliable transducers for the acoustic particle velocity; see, e.g., the historical overview in Ref. 1. The situation is different in underwater acoustics since "in water, direct measurement [of the acoustic particle velocity] is simple,"[4] and several *p-u* measurement systems have recently been developed; see, e.g., Refs. 4,5. This paper is concerned with air acoustics. Some years ago a micromachined transducer called the "Microflown" became available for measurement of the acoustic particle velocity in air,[6] and an intensity probe based on this device in combination with a small pressure microphone is now in commercial production.[7,8] Yet another method based on determining the sound pressure from an approximation to the divergence of the particle velocity (the "*u-u*" method, which involves six velocity transducers)[9] has, to the authors' knowledge, never been used in air.

The purpose of this paper is to compare, discuss, and examine the main limitations of the *p-u* and the *p-p* measurement principles in sound power determination under very difficult measurement conditions.

II. THE *p-p* MEASUREMENT PRINCIPLE

The *p-p* measurement principle employs two pressure microphones. The particle velocity component in the direction of the axis of the probe is obtained by a finite-difference approximation to the pressure gradient in Euler's equation of motion, and the sound pressure is simply the average of the two pressure signals. The most important limitations of this measurement technique are caused by the finite difference approximation, scattering and diffraction, and instrumentation phase mismatch.

The accuracy of the finite-difference approximation and the effect of scattering and diffraction obviously depend on the geometry of the microphone arrangement. Several configurations are possible, but in the early 1980s it was shown experimentally that the face-to-face configuration with a solid "spacer" between the two microphones is particularly favorable.[10] Much later it was discovered that the effect of scattering and diffraction not only tends to counterbalance the finite-difference error but in fact for a certain length of the spacer almost perfectly cancels it under virtually any sound field condition encountered in practice.[11] A practical consequence is that the upper frequency limit of a sound intensity probe based on two 1/2 in. microphones separated by a 12 mm spacer in the face-to-face arrangement is about 10 kHz, which is about an octave higher than the frequency limit determined by the finite-difference approximation.[11] The combination of 1/2 in. microphones and a 12 mm

spacer is now regarded as optimal, and longer spacers are only used when the focus is exclusively on low frequencies.[12]

Unless the measurement is compensated for phase mismatch the microphones for measurement of sound intensity with the *p-p* method have to be phase matched extremely well, and state-of-the-art sound intensity microphones are matched to a maximum phase response difference of 0.05° below 250 Hz and a phase difference proportional to the frequency above 250 Hz (say, 0.2° at 1 kHz).[13] The proportionality to the frequency is a consequence of the fact that phase mismatch in this frequency range is caused by differences between the resonance frequencies and the damping of the two microphones.[14]

It can be shown that a small phase mismatch error φ_{pe} gives rise to a bias error that can be approximated by[15,16]

$$\hat{I}_r \simeq I_r - \frac{\varphi_{pe}}{k\Delta r}\frac{p_{\mathrm{rms}}^2}{\rho c} = I_r\left(1 - \frac{\varphi_{pe}}{k\Delta r}\frac{p_{\mathrm{rms}}^2/\rho c}{I_r}\right), \quad (1)$$

where I_r is the "true" intensity (unaffected by phase mismatch), $\hat{I}_r$ is the biased estimate, p_{rms} is the rms value of the sound pressure, k is the wave number, Δr is the microphone separation distance, ρ is the density of air, and c is the speed of sound. This expression shows that the effect of a given phase error is inversely proportional to the frequency and the microphone separation distance and is proportional to the ratio of the mean square sound pressure to the sound intensity. If this ratio is large then even the small phase errors mentioned earlier will give rise to significant bias errors. Because of phase mismatch it will rarely be possible to make reliable measurements below, say, 80 Hz, except under very favorable semi free-field conditions unless a longer spacer than the usual 12 mm spacer is used.[17]

The ratio of the phase error to the product of the frequency and the microphone separation distance can be measured (usually in the form of the so-called "pressure-residual intensity index"[1]) by exposing the two pressure microphones to the same pressure in a small coupler.[1–3] Modern sound intensity analyzers automatically determine the ratio of the mean square pressure to the intensity during the intensity measurements.[18] Thus one has a clear indication of whether the bias error is serious or not.

The global version of Eq. (1) is found by integrating the normal component over a surface that encloses a source. The result is[16,19]

$$\hat{P}_a = \int_S \hat{\mathbf{I}} \cdot d\mathbf{S} \simeq P_a\left(1 - \frac{\varphi_{pe}}{k\Delta r}\frac{\int_S (p_{\mathrm{rms}}^2/\rho c)dS}{\int_S \mathbf{I} \cdot d\mathbf{S}}\right), \quad (2)$$

where P_a is the "true" sound power of the source within the surface and $\hat{P}_a$ is the biased estimate. The ratio of the surface integral of the mean square pressure to the surface integral of the intensity (in decibels) is known as the pressure-intensity index of the measurement.[1]

Calibration of *p-p* sound intensity measurement systems involves calibrating the two pressure microphones with a pistonphone in the usual manner and determining the pressure-residual intensity index in a small coupler driven by a wide-band signal as mentioned earlier.[1]

III. THE *p-u* MEASUREMENT PRINCIPLE

A. General considerations

A *p-u* sound intensity measurement system combines two fundamentally different transducers. The sound intensity is simply the time average of the instantaneous product of the pressure and particle velocity signal,[1]

$$I_r = \langle pu_r \rangle_t = \tfrac{1}{2}\mathrm{Re}\{pu_r^*\}, \quad (3)$$

where $\langle \rangle_t$ indicates time averaging, and the latter expression is based on the complex representation of harmonic variables. However, irrespective of the measurement principle used in measuring the particle velocity there is one fundamental problem: the pressure and the particle velocity transducer will invariably have different phase responses.[1] One must compensate for this "*p-u* phase mismatch," otherwise the result may well be meaningless. In fact even a small residual *p-u* mismatch error can have serious consequences under certain conditions. This can be seen by introducing such a small phase error, φ_{ue}, in Eq. (3). The result is[20]

$$\begin{aligned}\hat{I}_r &= \tfrac{1}{2}\mathrm{Re}\{p\hat{u}_r^*\} = \tfrac{1}{2}\mathrm{Re}\{pu_r^* e^{-j\varphi_{ue}}\}\\ &= \mathrm{Re}\{(I_r + jJ_r)(\cos\varphi_{ue} - j\sin\varphi_{ue})\} \simeq I_r\\ &\quad + \varphi_{ue}J_r,\end{aligned} \quad (4)$$

where

$$\hat{u}_r = u_r e^{j\varphi_{ue}} \quad (5)$$

is the particle velocity estimate, and

$$J_r = \tfrac{1}{2}\mathrm{Im}\{pu_r^*\} = \mathrm{Re}\{I_r + jJ_r\} \quad (6)$$

is the reactive intensity.[21,22] Whereas the (active) intensity describes the net flow of sound energy the reactive intensity describes the nonpropagating part of the energy, which is merely flowing back and forth, corresponding to the instantaneous particle velocity being in quadrature with the sound pressure. Many sources have strongly reactive near fields at low frequencies where they mainly generate evanescent waves. Near such a source the air is essentially moving back and forth as if it were incompressible.[1] Equation (4) demonstrates that even a small uncompensated *p-u* phase mismatch error will give rise to a significant bias error when $J_r \gg I_r$. On the other hand it also shows that substantial *p-u* phase errors can be tolerated if $J_r \ll I_r$. For example, even phase mismatch of 35° gives a bias error of less than 1 dB under such conditions. In other words, the phase calibration is critical when measurements are carried out under near field conditions, but not at all critical if the measurements are carried out in the far field. The "reactivity" (the ratio of the reactive to the active intensity) indicates whether this source of error is of concern or not.

The global version of Eq. (4) is found by integrating over a surface that encloses a source,[20]
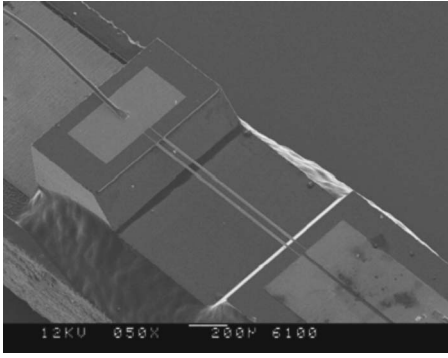
FIG. 1. The two heated wires of the Microflown particle velocity transducer.

FIG. 2. A Microflown 1/2 in. sound intensity probe. The particle velocity sensor can be seen on the solid cylinder to the right; the electret microphone is mounted inside the other cylinder.

$$\hat{P}_a = \int_S \text{Re}\{(\mathbf{I} + j\mathbf{J})(\cos \varphi_{ue} - j \sin \varphi_{ue})\} \cdot d\mathbf{S} \simeq P_a$$

$$+ \varphi_{ue} \int_S \mathbf{J} \cdot d\mathbf{S}$$

$$= P_a \left(1 + \varphi_{ue} \frac{\int_S \mathbf{J} \cdot d\mathbf{S}}{\int_S \mathbf{I} \cdot d\mathbf{S}} \right), \tag{7}$$

and this shows that uncompensated *p-u* phase mismatch is a potential source of error when the reactivity (which in the global case is the ratio of the surface integral of the reactive intensity to the surface integral of the active intensity) is large. This will typically occur at low frequencies when the measurement surface is close to the source. Thus the reactivity is an important field indicator for *p-u* probes. By contrast the pressure-intensity index is not relevant for *p-u* probes, as also pointed out by Fahy.[1]

Calibration of *p-u* sound intensity measurement systems involves exposing the probe to a sound field with a known relation between the pressure and the particle velocity, for example a plane propagating wave or a simple spherical wave.[1] Fahy is rather pessimistic about such calibration techniques, though; he says, "At best this is a check, and not an accurate means of calibration."[1]

B. The Microflown intensity probe

The MF particle velocity transducer consists of two short, thin, closely spaced wires of silicon nitride coated with platinum and heated by a dc current to about 300 °C;[6–8] see Fig. 1. The resistance of the wires depends on the temperature. An acoustic particle velocity signal in the perpendicular direction changes the temperature distribution instantaneously, because one of the wires will be cooled more than the other by the airflow, and this difference in resistance is measured with a bridge circuit that provides a signal proportional to the particle velocity. At low frequencies the sensitivity of this device increases 6 dB per octave. Between 100 and 1 kHz the frequency response is relatively flat. Between 1 and 10 kHz there is a rolloff of 6 dB per octave caused by diffusion effect related with the distance between the two wires, and above 10 kHz the sensitivity decreases an additional 6 dB per octave because of the thermal heat capacity of the wires. The particle velocity transducer is combined with a small electret condenser microphone in the 1/2 in. sound intensity probe shown in Fig. 2. The velocity trans-

ducer is mounted on a small, solid cylinder, and the condenser microphone is mounted inside another, hollow cylinder. The geometry of this arrangement increases the velocity and thus the sensitivity of the velocity transducer.

IV. DISCUSSION

Inspection of Eqs. (2) and (7) shows that *p-p* and *p-u* sound intensity measurement systems are affected differently by extraneous noise. Sources outside the measurement surface do not contribute to the surface integral of the "true" intensity [in the the denominator of the second term on the right-hand side of Eq. (2)], but they increase the surface integral of the mean square pressure (the numerator of the second term), from which it follows that even a very small phase error imposes restrictions on the amount of extraneous noise that can be tolerated in sound power measurements with a *p-p* sound intensity measurement system. (This conclusion was anticipated by Pascal as early as in 1981.[23]) By contrast, sources outside the measurement surface do not in general increase the reactivity [the second term on the right-hand side of Eq. (7)], and thus they do not in general increase the error due to *p-u* phase mismatch.

High values of the pressure-intensity index can occur in the entire frequency range;[11] therefore *p-p* phase mismatch can be of concern also at high frequencies. On the other hand very reactive sound fields are unlikely to occur except at low frequencies;[16] therefore *p-u* phase mismatch will generally be a problem only at low frequencies, and it will usually improve the situation to move away from the source.

The simple expedient of reversing a *p-p* probe makes it possible to eliminate the influence of *p-p* phase mismatch, because the intensity changes sign but the error does not.[24,25] Unfortunately, most *p-p* intensity probes are not symmetrical and therefore not suitable for real measurements with the probe reversed. By contrast the MF probe can easily be reversed. However, reversing a *p-u* probe simply changes the sign of the result, including the bias error, as easily seen from Eq. (4).[26] In other words, there is no simple way of getting rid of the phase error of a *p-u* probe; it must be calibrated with sufficient accuracy.

V. EXPERIMENTAL RESULTS

A. Free-field calibration

To validate the foregoing considerations and examine the performance of the MF sound intensity probe some experiments have been carried out. Initially the device was tested in the small anechoic room at DTU at a position 2 m from a small loudspeaker. A Brüel & Kjær (BK) "pulse" analyzer of type 3560 in one-third octave mode was used in all measurements. First the directional response of the particle velocity sensor was measured in the vertical and in the horizontal plane, and found to be in reasonable agreement with the expected cosine behavior in the entire frequency range. Next, the pressure response was measured; it was found to be in fair agreement with the response of a 1/2 in. free-field microphone of type BK 4191 up to 6.3 kHz. And finally the frequency response between the pressure and the velocity signal of the device was determined. This response, which took values over an interval of almost 30 dB and 140° in the frequency range from 50 Hz to 10 kHz, served as a correction factor to the cross spectrum in subsequent intensity measurements.

In the first place it was assumed that the probe had been exposed to a plane propagating wave in the velocity calibration measurement. However, as soon as the velocity calibration was applied to sound power measurements it became apparent that 2 m distance from the source is not enough to ensure plane wave conditions at low frequencies. Thus the measurement was repeated in the large anechoic room at DTU at positions 2, 3, and 4 m from the loudspeaker. This room is very good at frequencies down to 50 Hz.[27] The velocity calibration was corrected for the "near field effect," that is, the factor in parentheses in the expression for the relation between the particle velocity and the pressure generated by a monopole,

$$u_r(r) = \frac{p(r)}{\rho c}\left(1 + \frac{1}{jkr}\right). \tag{8}$$

Figure 3 shows the amplitude and phase corrections measured at 2 and 4 m distance, with and without the near field effect taken into account. (The corrections measured at 3 m distance are similar.) Below 200 Hz it is clearly necessary to correct for the near field effect even at a distance of 4 m. The fact that the corrected calibration curves differ from each other below 100 Hz simply demonstrates that a loudspeaker is not a monopole.

In what follows the active and the reactive intensity has been estimated in one-third octave bands using

$$\hat{I}_r = \text{Re}\left\{\frac{S_{\hat{p}\hat{u}}H_{pu}}{H_{\hat{p}\hat{u}}|H_{p\hat{p}}|^2}\right\}, \tag{9}$$

$$\hat{J}_r = -\text{Im}\left\{\frac{S_{\hat{p}\hat{u}}H_{pu}}{H_{\hat{p}\hat{u}}|H_{p\hat{p}}|^2}\right\}, \tag{10}$$

where $S_{\hat{p}\hat{u}}$ is the cross spectrum between the two signals from the MF probe, $H_{\hat{p}\hat{u}}$ is the ratio of the velocity to the pressure from the calibration measurement, H_{pu} is the corresponding theoretical value of the ratio of the velocity to the pressure
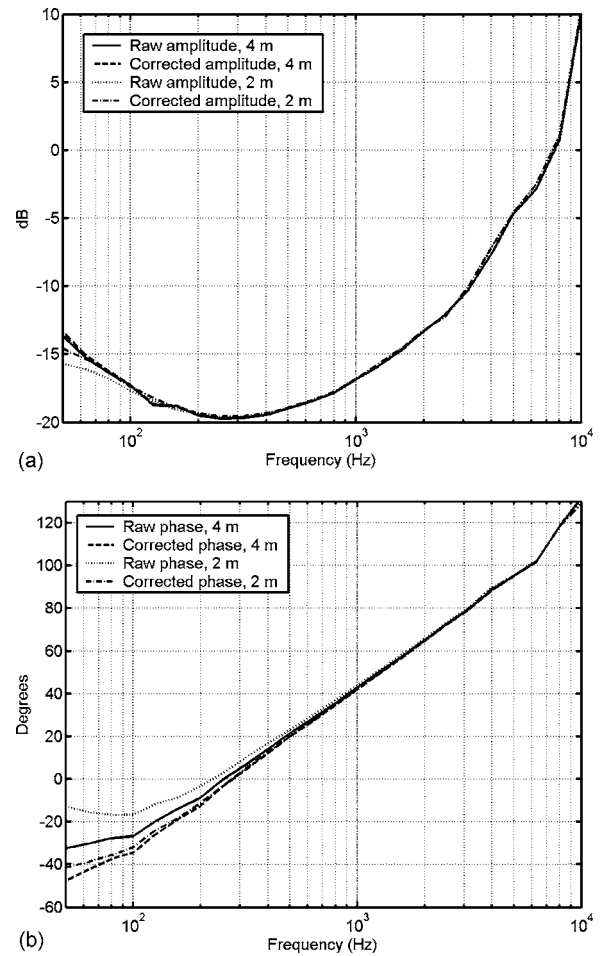
FIG. 3. Amplitude (a) and phase (b) calibration of the velocity signal relative to the pressure signal, measured at 2 and 4 m distance in an anechoic room, with and without correction for the finite distance to the loudspeaker.

[from Eq. (8)], and the pressure correction,$H_{p\hat{p}}$, takes account of fact that the pressure sensitivity of the device is not completely flat. The pressure correction is shown in Fig. 4.

B. Sound power measurements

The next experiments took place in a large hall where the sound power of a "sound source" of type BK 4205 was
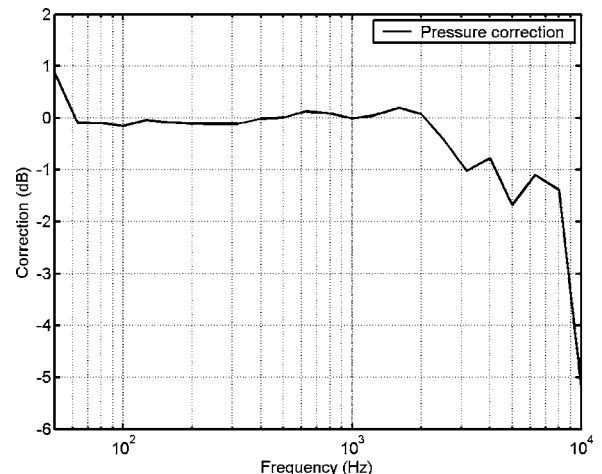
FIG. 4. Pressure correction of the MF probe.
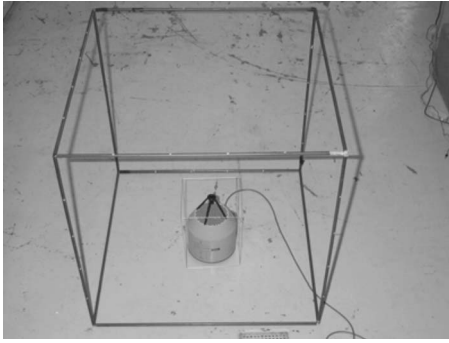
FIG. 5. Large and small measurement surface and the BK 4205 sound source.

FIG. 7. Ratio of reactive to active intensity on two surfaces enclosing the BK 4205 sound source.

measured by scanning over a surface enclosing the source with two different intensity probes, the MF probe and a *p-p* probe of type BK 3599 with microphones of type BK 4181. The two channels of the BK *p-p* sound intensity measurement system were much better matched than required of "class 1" systems by the IEC standard;[2] see Ref. 28. The reactive intensity was also measured. Two different measurement surfaces were used, one with an area of 5 m[2], and a very small one with an area of about 0.4 m[2] (see Fig. 5), and each measurement was repeated using a different scanning pattern. The repeatability was found to be very good in all cases.

Figure 6 shows the results of the four sound power measurements. The results determined with the MF probe have been processed using the presumably most accurate phase and amplitude calibration function, the one determined at 4 m distance and corrected for the near field effect. As can be seen all four measurements are in good agreement up to 6.3 kHz. At 8 and 10 kHz the MF probe overestimates slightly, perhaps because the pressure response of the device depends on the direction of incidence whereas the pressure calibration has been determined for axial incidence, or perhaps because the geometry of the arrangement affects the particle velocity. No influence of the measurement surface can be seen, in spite of the fact that the sound field on the small surface very

close to the source is strongly reactive at low frequencies, as demonstrated by Fig. 7. The fluctuations in the measured reactivity at high frequencies on the large measurement surface are of no concern. In this frequency range the reactive intensity decays rapidly with the distance to the source, and thus it takes very small values on the large measurement surface.[16,26] Besides, measuring reactive intensity in a predominantly active sound field with a *p-p* measurement system requires extremely accurate amplitude calibration.[29]

That the phase calibration of the *p-u* probe can be critically important is demonstrated by Fig. 8, which shows the error of the sound power measurements with the MF probe on the two measurement surfaces with the velocity calibration measured at 2 and 4 m distance, without and with correction for the finite distance to the loudspeaker in the calibration measurement. The reference is the measurement with the BK probe on the large surface. It is apparent that the reactive sound field on the small measurement surface amplifies the influence of *p-u* phase mismatch, as predicted by Eq. (7).

The BK 4205 sound source is an enclosed loudspeaker,

FIG. 6. Sound power of the BK 4205 sound source, measured with a BK intensity probe and with a MF intensity probe on two different measurement surfaces.
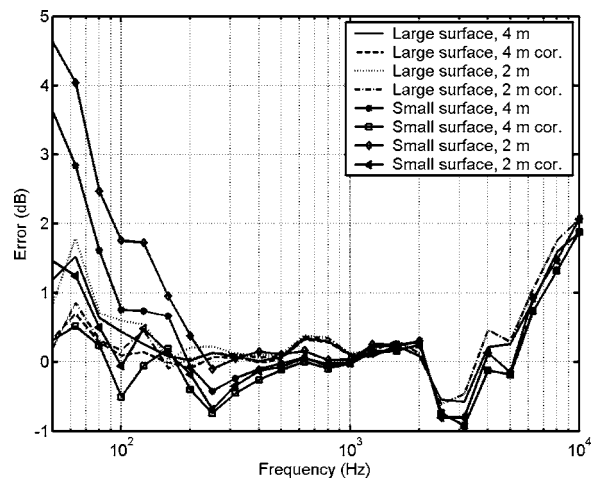
FIG. 8. Estimation error of the MF probe using different phase calibrations. A measurement with a BK intensity probe on the large measurement surface is used as the reference.
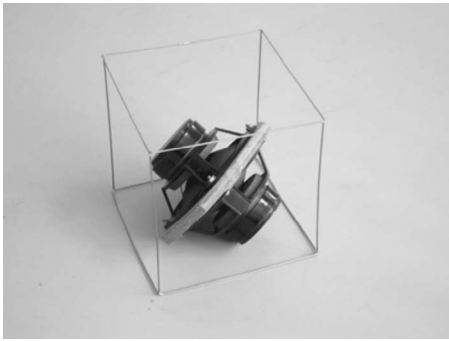
FIG. 9. Dipole source inside a very small measurement surface.

FIG. 11. Sound power of the dipole, measured with a BK intensity probe and with a MF intensity probe on two different measurement surfaces.

which is a well-behaved source. To examine the performance of the MF probe under more severe conditions the sound power of a dipole has been determined using an extremely small measurement surface with an area of 0.2 m². The dipole was constructed by mounting two loudspeaker units against each; see Fig. 9. Placed on concrete floor this source generated an exceptionally reactive sound field on the small measurement surface below 500 Hz, as shown in Fig. 10. In this case the BK probe produced negative intensity values at 50 and 63 Hz; hence the missing part of the corresponding curve. However, the "true" average intensity on the small surface can be calculated from the average intensity on the large surface. The MF data on the small surface were affected by residual p-u phase mismatch, as will become apparent in what follows.

Figure 11, which corresponds to Fig. 6, shows the results of sound power measurements with the two devices on the two measurement surfaces. Again, missing parts of the curves indicate meaningless negative sound power estimates. The two probes are in good agreement on the large measurement surface except for the overestimation of the MF probe at 8 and 10 kHz, as before. It can also be seen that the BK probe performs reasonably well except at 50 and 63 Hz on the small surface, considering the difficulties in measuring on such a small surface with the somewhat bulky probe. However, on the small surface the MF probe obviously has problems at low frequencies.

Figure 12, which corresponds to Fig. 8, shows the error of the MF results determined using the various calibrations. As can be seen, enormous errors occur when the small measurement surface is used, confirming once again that the phase calibration of the device is critically important when it is used in strongly reactive sound fields.

Finally it should be mentioned that the influence of strong background noise from sources outside the measurement surface has been examined. However, no significant difference between the performance of the p-p and the p-u measurement system under such conditions was observed.[28] As shown in Sec. II such background noise amplifies the error due to p-p phase mismatch, but with the very well-matched p-p intensity probe used in these measurements this effect could not be detected. The only effect of the noise was increased random errors, in all probability caused by the fact that the approximation to the surface integral in sound power measurement becomes more critical the higher the level of extraneous noise.

FIG. 10. Ratio of reactive to active intensity on two surfaces enclosing the dipole.
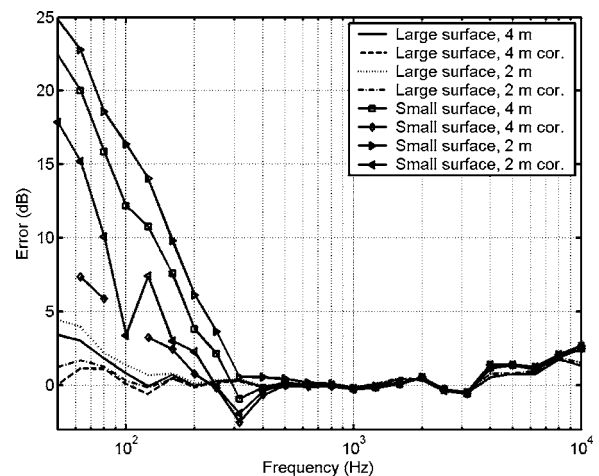
FIG. 12. Estimation error of the MF probe using different phase calibrations. A measurement with a BK intensity probe on the large measurement surface is used as the reference.
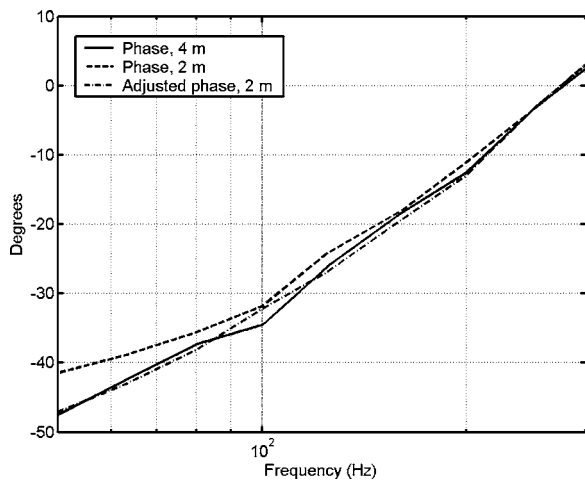
FIG. 13. Original and adjusted phase calibration.

C. Improved phase calibration

Not everybody has easy access to a large anechoic room of high quality; therefore it has been examined whether one can improve a free-field calibration made too close to the loudspeaker simply by manual adjustment of the phase calibration until the error that occurs in a very reactive sound field disappears. Figure 13 shows the modified phase correction, and Fig. 14 shows the result of such a procedure in the frequency range from 50 to 200 Hz. The starting point was the phase calibration measured at 2 m distance and corrected for the phase angle between pressure and particle velocity due to the finite distance on the assumption that the loudspeaker was a perfect monopole. The adjustment must be made in small steps. Below 100 Hz a phase increment of 0.1° has drastic consequences for the measurement on the small surface (the intensity estimate can change its sign, for example), confirming again that the phase calibration is critically important under such conditions. Modifying the phase calibration so as to remove the overestimation in the results on the small measurement surface at low frequencies has a very limited effect on the results determined on the large measurement surface, and also, not shown, a very modest influence on the measurement of the sound power of the BK
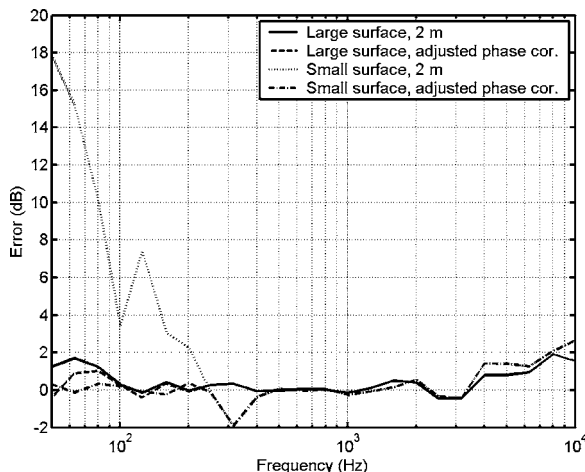
4205 using the large surface, but a positive influence on the (fairly small) error in the measurement of the sound power of the BK 4205 made on the small surface (not shown).

Obviously this method cannot take account of errors in the *amplitude* calibration. However, these errors are quite small, cf. Fig. 3(a), and their effect do not depend on the sound field conditions.

VI. CONCLUSIONS

The *p-u* sound intensity measurement principle has been examined theoretically and experimentally, and compared with the established *p-p* method with particular regard to the influence of phase mismatch on sound power estimation. It is more difficult to calibrate *p-u* intensity probes than *p-p* intensity probes, and whereas *p-p* phase mismatch in conventional *p-p* measurement systems can be almost eliminated by reversing the intensity probe, probe reversal has no effect on the influence of *p-u* phase mismatch in measurements with *p-u* measurement systems. Strongly reactive sound fields exacerbate the influence of *p-u* phase mismatch, but have no influence on the effect of *p-p* phase mismatch. Such sound field conditions are reflected in a high value of the reactivity, that is, the ratio of the surface average of the reactive intensity to the surface average of the active intensity. This will rarely be a problem above 500 Hz, and can be avoided simply by moving the measurement surface further away from the source under investigation. By contrast, background noise from sources outside the measurement surface increases the influence of *p-p* phase mismatch, but has no influence on the effect of *p-u* phase mismatch. Such background noise is reflected in a high value of the pressure-intensity index. This problem can occur in the entire frequency range.

The experimental part of the investigation was carried out with a *p-u* sound intensity probe produced by Microflown. The results show that it is possible to measure sound power reliably with the Microflown intensity probe from 50 Hz to 6.3 kHz if strongly reactive near fields, which tend to make the phase calibration very critical, are avoided. The measurements have also shown that it is possible to expand the range of measurement to near field conditions if the phase can be calibrated with sufficient accuracy below 500 Hz; at higher frequencies the phase calibration is less critical. The calibration can be carried out in a large anechoic room, and the extreme sensitivity to *p-u* phase mismatch in very reactive sound fields can even be turned to advantage by adjustment of the phase correction until the error disappears.

[1] F. J. Fahy, *Sound Intensity* 2nd ed. (E & FN Spon, London, 1995).
[2] IEC Publication 61043, "Instruments for the measurement of sound intensity–Measurements with pairs of pressure sensing microphones," 1993.
[3] American National Standard ANSI S1.9-1996, "Instruments for the mea-

FIG. 14. Estimation error with improved calibration.

surement of sound intensity," 1996.

[4] T. B. Gabrielson, D. L. Gardner, and S. T. Garret, "A simple neutrally buoyant sensor for direct measurement of particle velocity and intensity in water," J. Acoust. Soc. Am. **97**, 2227–2237 (1995).

[5] K. Kim, T. Gabrielson, and G. C. Lauchle, "Development of an accelerometer-based underwater acoustic intensity sensor," J. Acoust. Soc. Am. **116**, 3384–3392 (2004).

[6] H.-E. de Bree, P. Leussink, T. Korthorst, H. Jansen, T. Lammerink, and M. Elwensoek, "The Microflown: A novel device measuring acoustical flows," Sens. Actuators, A **SNA054/1-3**, 552–557 (1996).

[7] H.-E. de Bree, "The Microflown: An acoustic particle velocity sensor," Acoust. Aust. **31**, 91–94 (2003).

[8] R. Raangs, W. F. Druyvesteyn, and H.-E. de Bree, "A low-cost intensity probe," J. Audio Eng. Soc. **51**, 344–357 (2003).

[9] K. J. Bastyr, G. C. Lauchle, and J. A. McConnell, "Development of a velocity gradient underwater acoustic intensity sensor," J. Acoust. Soc. Am. **106**, 3178–3188 (1999).

[10] G. Rasmussen and M. Brock, "Acoustic intensity measurement probe," Proceedings of Recent Developments in Acoustic Intensity Measurement, Senlis, France, 1981, pp. 81–88.

[11] F. Jacobsen, V. Cutanda, and P. M. Juhl, "A numerical and experimental investigation of the performance of sound intensity probes at high frequencies," J. Acoust. Soc. Am. **103**, 953–961 (1998).

[12] F. Jacobsen, V. Cutanda, and P. M. Juhl, "A sound intensity probe for measuring from 50 Hz to 10 kHz," Brüel & Kjær Tech. Rev. **1**, 1–8 (1996).

[13] Anonymous, "Product data, sound intensity pair–type 4197," Brüel & Kjær, Nærum, Denmark.

[14] E. Frederiksen and O. Schultz, "Pressure microphones for intensity measurements with significantly improved phase properties," Brüel & Kjær Tech. Rev. **4**, 11–12 (1986).

[15] T. G. Nielsen, "A field calibrated real time system for intensity measurements," Proceedings of Inter-Noise 88, Avignon, France, 1988, pp. 123–126.

[16] F. Jacobsen, "Sound field indicators: Useful tools," Noise Control Eng. J. **35**, 37–46 (1990).

[17] S. Gade, "Validity of intensity measurements in partially diffuse sound field," Brüel & Kjær Tech. Rev. **4**, 3–31 (1985).

[18] Anonymous, "Technical documentation, sound intensity software BZ 7205 for 2260 Investigator," Brüel & Kjær, Nærum, Denmark.

[19] F. Jacobsen, "A simple and effective correction for phase mismatch in intensity probes," Appl. Acoust. **33**, 165–180 (1991).

[20] F. Jacobsen, "Spatial sampling errors in sound power estimation based upon intensity," J. Sound Vib. **145**, 129–149 (1991).

[21] J. A. Mann, III, J. Tichy, and A. J. Romano, "Instantaneous and time-averaged energy transfer in acoustic fields," J. Acoust. Soc. Am. **82**, 17–30 (1987).

[22] F. Jacobsen, "Active and reactive, coherent and incoherent sound fields," J. Sound Vib. **130**, 493–507 (1989).

[23] J.-C. Pascal, "Estimation de la puissance acoustique à l'aide d'intensimètres," Proceedings of Recent Developments in Acoustic Intensity Measurement, Senlis, France, 1981, pp. 179–185.

[24] J. Y. Chung, "Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch," J. Acoust. Soc. Am. **64**, 1613–1616 (1978).

[25] S. J. Elliott, "Errors in acoustic intensity measurements," J. Sound Vib. **78**, 439–445 (1981).

[26] F. Jacobsen and E. S. Olsen, "The influence of microphone vents on the performance of sound intensity probes," Appl. Acoust. **41**, 25–45 (1994).

[27] F. Ingerslev, O. J. Pedersen, P. K. Møller, and J. Kristensen, "New rooms for acoustic measurements at the Danish Technical University," Acustica **19**, 185–199 (1967/68).

[28] F. Jacobsen and H.-E. de Bree, "Intensity-based sound power determination under adverse sound field conditions: p-p versus p-u probes," Proceedings of Twelfth International Congress on Sound and Vibration, Lisbon, Portugal, 2005.

[29] F. Jacobsen, "Active and reactive sound intensity in a reverberant sound field," J. Sound Vib. **143**, 231–240 (1990).

Data-based mode extraction with a partial water column spanning array

S. C. Walker,[a] Philippe Roux, and W. A. Kuperman
Marine Physical Laboratory of the Scripps Institution of Oceanography,
University of California—San Diego, La Jolla, California 92093-0238

In a shallow ocean waveguide the acoustic field can be characterized by depth-dependent modes propagating in range with an associated propagating wavenumber. Though recently developed methods for determining the modes from recorded acoustic data alone without ocean or bottom modeling have shown promise, they are only applicable when the acoustic field is sampled over the entire water column. This paper presents a method for determining the acoustic modes from measured data alone when the field is sampled over only a portion of the water column. The method requires broadband sources at many ranges, e.g., a moving source, in order to construct the frequency-wavenumber (f-k) structure of the waveguide. Because modal propagation is dispersive, the modes are characterized by a discrete set of wavenumbers that vary continuously with frequency. Due to the discreteness of the modal wavenumbers, it is possible to isolate the modes in the f-k domain and extract them individually with a singular value decomposition (SVD). Because the modes are extracted individually the full-spanning and degeneracy limitations of the SVD are removed. Theory, simulation, and laboratory data confirm the process. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1993149]

I. INTRODUCTION

The normal mode structure of an ocean waveguide is not only fundamental to describing propagation, but also provides opportunities for various types of array and signal processing.[1] In range-independent environments, matched mode or mode filtering methods[2–4] offer an effective alternative to more general matched field methods[5] that tend to be computationally costly and highly sensitive to environmental mismatch. Matched mode methods apply a knowledge of the modal components of the field to measured data over a vertical array of sensors. Typically, the modal components are constructed from theory using parameters derived from a detailed knowledge of the environment. The difficulty and high cost of obtaining the necessary environmental information has motivated research into the development of model-free methods for determining the modes of acoustic propagation. Of particular interest are methods that can be directly applied to acoustic data measured on a vertical line array (VLA) without any *a priori* information. Current research strategies range from active methods involving feedback between a pair of VLAs[6] to passive methods in which a VLA is used to accumulate information from acoustic sources and/or noise in the environment.[7–10] This paper demonstrates that, under certain conditions, normal modes can be extracted passively from the received acoustic field in a waveguide without a full water column spanning array, without environmental information and without any modeling.

We introduce a method for determining the acoustic modes of propagation from sound received on a partial-spanning VLA from a horizontally moving source. The method involves a combination of masking individual modal dispersion curves contained in a frequency-wavenumber (f-k) analysis of the data, and a subsequent singular value decomposition (SVD). The method is related to earlier work by Frisk and Lynch[11] in which they Hankel transformed (over range) acoustic data (at a single frequency) recorded on a single receiver from sources spanning a range aperture at a single depth in order to estimate the Green's function characterizing the shallow ocean waveguide. Extending this idea to broadband signals measured at many depths using the receivers of the VLA leads to a comprehensive knowledge of the depth-dependent f-k structure of the waveguide. By combining this knowledge with the physics of modal dispersion, it possible to overcome the inherent limitations to the applicability of other SVD based methods; namely singular value degeneracy and the full-spanning requirement.

SVD methods are based on the premise that the equation describing the decoupled full-spanning cross spectral density matrix (CSDM) of the frequency domain acoustic field is isomorphic to its SVD. Existing SVD methods rely on range, frequency, and ambient noise ensemble averaging to decouple the modes in the CSDM.[9,10] In this paper we introduce a frequency-wavenumber (f-k) modal dispersion curve isolation, or mode isolation (MI), scheme for decoupling the modes in the CSDM. By decomposing the acoustic field into its wavenumber components, it is possible to isolate the modes in the f-k domain prior to formation of the CSDM. Our scheme involves the formation of several CSDMs, each of which contains significant contributions from only one modal component of the acoustic field. Then a SVD is applied to each f-k averaged CSDM in order to extract each modal depth dependence one at a time. Not only does this eliminate the ambiguities arising from singular value degen-

---

[a]Electronic mail: shane@physics.ucsd.edu

eracies, it frees the SVD from its fully spanning constraint so that modes can be extracted over any portion of the water column for any sound speed profile.

Section II reviews the theory of SVD-based mode extraction. The conditions under which the SVD of the CSDM is expected to be isomorphic to the depth-dependent modes are discussed in Sec. II C. The range-averaged form of the CSDM introduced in Ref. 9 is used as an example. As well, the limitations of the SVD are addressed, namely those involving singular value degeneracies and normalization over a partial half-space. (Throughout this paper, the term "half-space" is meant to refer to all media below the surface of the water, i.e., the water column and all bottom layers.) The MI method is introduced and its advantages over other SVD methods are discussed in Sec. II D. Mode extraction results for both simulated and laboratory data are presented in Sec. III.

II. THEORY: SVD MODE EXTRACTION FROM VLA DATA

In this section we explicitly develop the relationship between acoustic data on an array, normal mode expansions, and the SVD process. After presenting this analysis, we then formally develop our extraction method.

A. Normal mode theory applied to the VLA pressure field

Normal mode theory states that the single frequency component, at angular frequency ω_j, of pressure field $p(z', z_q, \omega_j, r_h)$, at receiver point $(z = z_q, r = 0)$ in a range-independent waveguide with depth-varying density, $\rho(z)$, due to a time-varying pressure signal originating from point source, $S(t - t_0)$, located at the source point $(z = z', r = r_h)$ is given by

$$P_{qhj}(z') \equiv p(z', z_q, \omega_j, r_h)$$
$$= \frac{\sqrt{2\pi}e^{i\pi/4}}{\rho(z')} \sum_{m=1}^{M} \frac{\phi_m(z', \omega_j)\phi_m(z_q, \omega_j)}{\sqrt{k_m(\omega_j)r_h}} e^{-ik_m(\omega_j)r_h} \widetilde{S}(\omega_j).$$
$$(1)$$

Here M is the number of propagating modes supported by the environment, $\widetilde{S}(\omega_j)$ represents the Fourier transform of $S(t - t_0)$

$$\widetilde{S}(\omega_j) \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} S(t - t_0)e^{-i\omega_j t}\, dt, \quad (2)$$

and $k_m(\omega_j)$ is the modal wavenumber. The $\phi_m(z_q, \omega_j)$ are the normalized, orthogonal depth-dependent mode functions satisfying

$$\int_0^{D'} \frac{\phi_m(z_q, \omega_j)\phi_m(z_q, \omega_j)}{\rho(z_q)}\, dz_q \equiv \delta_{mn}. \quad (3)$$

Rigorously, δ_{mn} represents a Kronecker-delta function when D' spans the entire half-space. The variables z_q, ω_j, and r_h are subscripted in order to facilitate the connection between the theoretical discussions in the continuous limit to follow and their discretized experimental implementations. Explic-
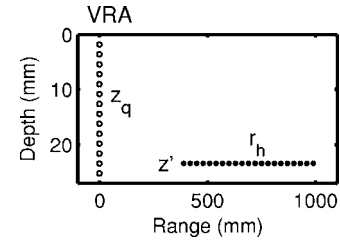
FIG. 1. Experimental setup. VLA composed of Q transducers at depths z_q (open circles) located at range $r = 0$ records acoustic field from sources at ranges $r = r_h$ and depth z'. The acoustic fields from sources at the same depth for many ranges (closed circles) are recorded individually.

itly, discretization is introduced to describe the experimental situation shown in Fig. 1 such that z_q is the depth of the qth VLA receiver element, r_h is the range of the hth source, and ω_j is the jth signal frequency component. Under an integral, a subscript serves as a label on a continuous variable. Under a summation, a subscript serves as a proper index. Functions of subscripted independent variables may also adopt the subscripted notation. Two notable examples are

$$k_m(\omega_j) \to k_{mj}, \quad \phi_m(z_q, \omega_j) \to \phi_{qmj}.$$

B. The SVD

Any $Q \times Q$ complex matrix $\mathbf{A}$ can always be written as a product of matrices such that

$$\mathbf{A} = \mathbf{USV}^\dagger, \quad A_{qq'} = \sum_{\alpha=1}^{N} u_{q\alpha}s_{\alpha\alpha}v_{\alpha q'}^*, \quad (4)$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, $\mathbf{V}^\dagger$ is the complex conjugate transpose of $\mathbf{V}$, and $\mathbf{S}$ is a diagonal matrix whose elements are the singular values of $\mathbf{A}$ and are positive. N is the number of rows of $\mathbf{Q}$ (and $\mathbf{V}$). $\mathbf{U}$ and $\mathbf{V}$ are unique when the singular values are nondegenerate. Performing a SVD on $\mathbf{A}$ yields $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$. By design, the SVD implementation we use orders the column vectors of $\mathbf{U}$ and $\mathbf{V}$ according to their respective singular values, with the vector corresponding to the greatest singular value appearing in the first column, the vector corresponding to the next highest singular value in column 2, and so forth.

C. The SVD of the decoupled CSDM and the normal modes

The SVD-based mode extraction algorithms defined in Refs. 8 and 9 involve the formation of the spatial cross-spectral density matrix (CSDM), $\mathbf{C}$, as the outer product of the complex frequency domain pressure field, $\mathbf{P} \equiv P_{qhj}(z')$ data array, with its complex conjugate transpose (for either constant h or constant j), $\mathbf{P}^\dagger$, averaged over range and/or frequency. In general, $\mathbf{C}$ is a complex $Q \times Q$ matrix where Q is the number of elements on the VLA. In accordance with Sec. II B, $\mathbf{C}$ may be written as a product of matrices $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{S}$. Under the proper conditions, an isomorphism exists between some of the columns of $\mathbf{U}$ and some of the propagating depth-dependent modes. The remainder of this section briefly summarizes the connection between the SVD of the

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Walker *et al.*: Shallow water mode extraction   1519

CSDM and the depth-dependent modes using the range averaging or multiple range (MR) method of Ref. 9 as an example.

In the MR method, the CSDM, $\mathbf{C}^{MR}$, is defined as an average over ranges,

$$\mathbf{C}^{MR} \equiv \langle \mathbf{P}\mathbf{P}^{\dagger} \rangle^{MR},$$
(5)

$$C_{qq'}^{MR} = \frac{1}{H}\sum_{h=1}^{H} P_{qhj}(z')P_{hq'j}^{*}(z') = \sum_{m,n=1}^{M,M} \phi_{qmj}\phi_{nq'j}^{*}\delta_{mn}^{MR},$$

where

$$\delta_{mn}^{MR} \equiv \frac{2\pi|\widetilde{S}_{j}|^{2}\phi_{mj}(z')\phi_{nj}^{*}(z')}{H\rho^{2}(z')\sqrt{k_{mj}k_{nj}}}\left(\sum_{h=1}^{H}\frac{e^{-ik_{mnj}r_{h}}}{r_{h}}\right),$$
(6)

$$k_{mnj} \equiv k_{mj} - k_{nj}.$$
(7)

The MR method utilizes the information from a single frequency component, ω_{j}, of many acoustic pressure fields recorded at points z_{q} on the receiver VLA. The pressure fields are accumulated sequentially from sources at many ranges, r_{h}, all at the same depth z'. For a single frequency (constant j) then, $P_{qhj}(z')$ are the elements of a $Q \times H$ matrix for Q VLA elements and H range samples. For the case of adequate range sampling over a sufficiently large aperture, the summation over range in Eq. (6) results in the term in parentheses approaching an unnormalized Kronecker-delta function in the indices m and n. In this case, δ_{mn}^{MR} approaches a diagonal matrix so that Eq. (5) reduces to the approximation

$$C_{qq'}^{MR} \approx \sum_{m=1}^{M} \phi_{qmj}\phi_{mq'j}^{*}\delta_{mm}^{MR}.$$
(8)

Comparison to Eq. (4) makes explicit the isomorphism between the matrices $\mathbf{U}$ and $\mathbf{V}$ and the depth-dependent modes, ϕ_{qmj}, as well as the isomorphism between the singular values and the diagonal elements δ_{mm}^{MR},

$$U_{q\alpha} \equiv \vec{U}_{\alpha} \leftrightarrow \vec{\phi}_{mj} \equiv \phi_{qmj}, \quad s_{\alpha\alpha} \leftrightarrow \delta_{mm}^{MR}.$$
(9)

As the quantity δ_{mm}^{MR} is proportional to the intensity of mode m, the modes will be ordered according to their intensities measured at the VLA. Therefore, column 1 of the matrix $\mathbf{U}$ (and $\mathbf{V}$ as well), $\vec{U}_{1}$, returned by applying a SVD to the CSDM, is isomorphic to the depth-dependent mode of greatest intensity normalized to unity, and so forth. This isomorphism is valid only for modes which are completely decoupled from all other modes, i.e., for modes m for which δ_{mn}^{MR} approaches a Kronecker-delta function for all $n \neq m$. CSDM mode decoupling depends on the total number of sources, H, the span in range, or range aperture, of the sources, Δr, and the sampling interval of the sources, dr.[8,9]

Applying similar reasoning to a frequency-averaged form of the CSDM, a method we refer to as the frequency averaging or multiple frequency (MF) method, an isomorphism can be established between the frequency-averaged acoustic modes and the SVD column vectors in the case of limited bandwidth where the frequency-averaged acoustic modes are an adequate approximation to the true frequency-dependent modes. Proceeding further, it is possible to improve mode decoupling, and hence the mode extraction results, by combining the MR and MF methods. This hybrid method, MRMF, applies the SVD to a time- and range-averaged form of the CSDM to extract the modes.

CSDM mode decoupling is a necessary but not sufficient condition for the extraction of a given mode. For the case of degenerate singular values, $\mathbf{U}$ and $\mathbf{V}$ are not unique. Specifically, all column vectors of $\mathbf{U}$ whose corresponding singular values are degenerate are not unique. Hence, an isomorphism cannot be established between these vectors and any of the depth-dependent modes. As well, because $\mathbf{U}$ is unitary, the isomorphism is valid only when the matrix formed by the set of modes, $\mathbf{\Phi}_{j} \equiv \{\phi_{qmj}\}$ (the subscript j is included in reference to the MR case where a single frequency component is considered), is unitary. This is only true when $\mathbf{\Phi}_{j}$ spans the entire half-space, otherwise, $\Sigma_{q=1}^{Q}\phi_{mqj}\phi_{qnj}$ is not diagonal. This requires that the VLA cover the entire water column and that none of the propagating modes penetrate significantly into the bottom. Not only are the range, frequency, and ambient noise averaging techniques so far proposed in the literature constrained by the full-spanning requirement, but they are also hampered by degeneracy limitations (see Sec. II B).

As a final note, because $\mathbf{U}$ is comprised of Q columns, Q is the maximum number of modes that can be extracted. All the modes can be extracted when $Q \geq M$.

D. Using the f-k structure of the waveguide to extract the normal modes

In this section we augment the utility of range and frequency information in SVD-based methods by considering the underlying physics of modal dispersion in the f-k domain. Not only will such considerations allow us to overcome the degeneracy limitations of the earlier methods, but they also make it possible to extract all the depth-dependent modes projected over a partially spanning VLA.

The mode isolation (MI) method capitalizes on an extra bit of information not considered by the previously discussed mode extraction algorithms: namely, the model wavenumber pattern of an f-k dispersion representation. By accumulating the frequency domain VLA pressure fields from sources at many ranges all at the same depth, it is possible to apply a Fourier transformation over range as well as time. We note that the source ranges need not be evenly spaced.[12] In the case of varying sampling interval, the sampling intervals must be known. The range-wavenumber transformation results in the frequency-wavenumber structure of the waveguide projected over the VLA. This information is useful because in a horizontally stratified waveguide the modal wavenumbers are discrete and localized in the f-k domain, making it possible to isolate the modes (depending on the wavenumber resolution) from one another so that their depth dependence may be extracted individually. The effect of the isolation process is to reweight each isolated mode such that its effective excitation is much greater than the effective excitations of the other modes. This ensures that, upon apply-

ing the SVD to the isolated mode CSDM, the column vector of $\mathbf{U}$ corresponding to the greatest singular value, denoted here as $\vec{U}_1$, is associated with the isolated mode. By design, the SVD calculates the orthogonal column vectors of $\mathbf{U}$ sequentially according to their singular value. As each vector is calculated, it is constrained to be orthogonal with respect to each and every previously calculated vector. Because $\vec{U}_1$ is calculated first, it is not constrained by the orthogonalization process. Rather, it is merely normalized to unity over the spanning space. Furthermore, in the infinite wavenumber resolution limit ($\Delta r \rightarrow \infty$), the mode isolation process guarantees $\vec{U}_1$ to be nondegenerate. As a result, we conclude that $\vec{U}_1$ is isomorphic to the isolated depth-dependent mode projected over the VLA, regardless of the amount of the water column spanned by the VLA.

The remainder of this section is dedicated to the characterization of the MI process in a quantitative framework. Taking the discrete Fourier transform over the sampled ranges of the pressure field, we define the f-k domain VLA pressure field

$$\widetilde{P}_{qh'j}(z') \equiv \sum_{h=1}^{H} P_{qhj}(z') e^{-i\widetilde{k}_{h'} r_h}. \tag{10}$$

$\widetilde{k}_{h'}$ represents the wavenumber conjugate to the range, r_h. Because it is assumed that all the recorded pressure fields originate from the same source depth, the explicit depth dependence of the f-k domain pressure field shall be suppressed for the remainder of the paper. Substituting Eq. (1) and rearranging yields

$$\widetilde{P}_{qh'j} = \frac{\sqrt{2\pi} e^{i\pi/4}}{\rho(z')} \sum_{m=1}^{M} \frac{\phi_{mj}(z')\phi_{qmj}}{\sqrt{k_{mj}}} \widetilde{S}_j \gamma_{mh'j}, \tag{11}$$

where all the wavenumber-dependent phase factors have been included in $\gamma_{mh'j}$,

$$\gamma_{mh'j} \equiv \sum_{h=1}^{H} e^{-i(\widetilde{k}_{h'} + k_{mj})r_h}. \tag{12}$$

For constant range sampling interval, dr, $\gamma_{mh'j}$ reduces to the classical array diffraction result,

$$\gamma_{mh'j} = e^{-i(\widetilde{k}_{h'} + k_{mj})R_0} \frac{\sin((\widetilde{k}_{h'} + k_{mj})Hdr/2)}{\sin((\widetilde{k}_{h'} + k_{mj})dr/2)}, \tag{13}$$

where R_0 is the range of the nearest source (at $r=r_1$). The relevance of this result is that $\gamma_{mh'j}$ is peaked around $\widetilde{k}_{h'} = -k_{mj}$. Equation (13) demonstrates explicitly the dependence of the wavenumber resolution of the f-k pressure field on the range aperture, $\Delta r = Hdr$. $\gamma_{mh'j}$ approaches a delta function, $\gamma_{mh'j} \rightarrow \delta(\widetilde{k}_{h'} + k_{mj})$, in the infinite wavenumber resolution limit ($\Delta r \equiv |r_H - r_1| \rightarrow \infty$). For each q then, we recognize $\widetilde{P}_{qh'j}$ to be the f-k diagram of modal propagation at depth z_q defining the frequency-dependent wavenumber characteristics of the propagating depth-dependent modes. In the f-k domain, the modes appear as distinct, localized curvilinear regions centered on lines defined by $\widetilde{k}_{h'} = -k_{mj}$. The
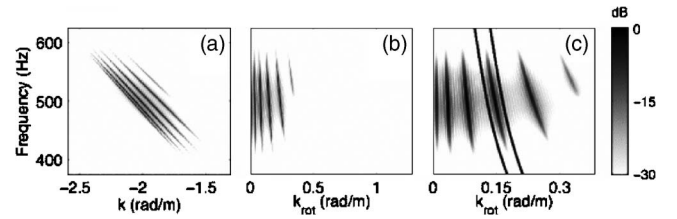
FIG. 2. Depth-averaged modulus of the f-k structure of a simulated waveguide. Propagating modes appear as localized dark curves. (a) Unrotated f-k structure. (b) Rotated f-k structure. (c) Magnified rotated f-k structure showing the mode 4 mask region (black lines). Simulated waveguide parameters: Pekeris waveguide, $D=15$ m, $c_w=1500$ m/s, $c_b=1800$ m/s, $\rho_b =1800$ kg/m³, $dr=2$ m, $\Delta r=1$ km, $r_c=3$ km (r_c is the range to the center of the range aperture), and $z'=15$ m.

localization is a consequence of the discreteness of the modal wavenumber spectrum. Figure 2(a) shows an example of the discrete wavenumber spectrum of a simulated ocean Pekeris waveguide. All f-k domain dispersion figures show the depth-averaged modulus of the spectrum. Because the modal wavenumbers are localized in the f-k domain, it is possible to isolate them. The isolation is accomplished by means of a 2D mask applied in the vicinity of a given mode projected over all VLA elements.

At this point, it is convenient to introduce a modification to our transformation procedure that facilitates the mode isolation process. We emphasize that this modification in no way affects the analysis of the underlying waveguide dispersion physics. Returning to Eq. (10), we introduce a frequency-dependent weighting factor (in parentheses) and define the rotated pressure field

$$\hat{\widetilde{P}}_{qh'j} \equiv \sum_{h=1}^{H} P_{qhj}(e^{i\widetilde{k}_j r_h}) e^{-i\widetilde{k}_{h'} r_h}. \tag{14}$$

Analogously to Eq. (12), the wavenumber-dependent complex phase dependence of the rotated pressure field is

$$\gamma_{mh'j} \equiv \sum_{h=1}^{H} e^{-i(\widetilde{k}_{h'} + k_{mj} - \hat{k}_j)r_h}. \tag{15}$$

As before, for the case of constant range sampling, Eq. (15) reduces to the array diffraction result similar to Eq. (13), only it is peaked about $\widetilde{k}_{h'} = \hat{k}_j - k_{mj}$. The weighting factor transforms the f-k dispersion structure according to the choice for $\hat{k}_j$. The transformation is arbitrary. For example, choosing $\hat{k}_j$ equal to a constant results in a simple translation in wavenumber of the dispersion structure. For our purposes, we find it convenient to define $\hat{k}_j \equiv k_{1j}$ where k_{1j} is the frequency-dependent mode 1 wavenumber. This choice of transformation effectively rotates the f-k structure such that mode 1 curve appears at $\widetilde{k}_{h'} \approx 0$ for all frequencies. We approximate the mode 1 wavenumber $k_{1j} \approx \omega_j/v_{g1}$, where v_{g1} is a frequency independent approximation to the mode 1 group velocity, $v_{g1}=1500$ m/s over the bandwidth.[13] Because the modal wavenumber spectrum density increases as mode number decreases, modal wavenumbers of modes 1 and 2 will be the most difficult to resolve, and, therefore, the most difficult to isolate. This choice of rotation facilitates

the isolation of the low-order modes by rendering them nearly vertical. Figure 2(b) shows the rotation applied to the f-k structure of Fig. 2(a). In this example using simulated shallow ocean waveguide data, all five propagating modes are well resolved in wavenumber.

Referring to the f-k dispersion structure, the mask of mode ν is defined as the set of $\tilde{k}_{h'}$ for which

$$\hat{k}_j - k_{\nu j} - \epsilon_\nu \leq \tilde{k}_{h'} \leq \hat{k}_j - k_{\nu j} + \epsilon_\nu \qquad (16)$$

for each frequency component of interest. The width (in wavenumber) of the mode ν mask is determined by the quantity ϵ_ν. The mask width, and hence ϵ_ν, should be large enough that the mask contains all the important features of mode ν, but still small enough that it does not incorporate significant features of adjacent modes. We adopt the notation $\{X\}_\nu$ to represent the mode ν masking operation applied to the quantity X where X is a synthesis of modal components, $X = \sum_{m=1}^{M} x_m$. Because it is an isolation operation, the masking operation conceptually behaves similarly to the Kronecker delta function, $\delta_{m\nu}$. It isolates the mode ν component of quantity X, $\{X\}_\nu \approx x_\nu$. Applying the masking operation to the rotated f-k pressure field yields

$$\{\hat{\tilde{P}}_{qh'j}\}_\nu \approx \frac{\sqrt{2\pi} e^{i\pi/4}}{\rho(z')} \frac{\phi_{\nu j}(z') \phi_{q\nu j}}{\sqrt{k_{\nu j}}} \tilde{S}_j \gamma_{\nu h'j}. \qquad (17)$$

This form for the masking operation applied to the pressure field is valid in the limit that $\gamma_{mh'j} \rightarrow \gamma_{\nu h'j}$ over the mask region [the region defined by condition (16)]. This is indeed the case when mode ν is well isolated and a mask containing only mode ν can be defined unambiguously. That is, as long as the wavenumber resolution is adequate to render mode ν distinct from its neighboring modes, $\nu-1$ and $\nu+1$, in the f-k domain, the dominant contribution to the pressure field, and hence the CSDM, within the mask comes from mode ν.

Because the depth dependence of the modes is unaffected by the transformation process $(r \rightarrow k)$, it is not necessary to transform back from wavenumber to range in order to form the CSDM. Defining the mask such that there are the same number of wavenumber elements, H', for each frequency, $\{\hat{\tilde{P}}_{qh'j}\}_\nu$ is a $Q \times H' \times J$ array which can be used to form the CSDM directly. J here is the number of frequency components included in the mask. The CSDM is then defined as the double sum over the mode ν mask region

$$C_{qq'} = \frac{1}{H'J} \sum_{h',j=1}^{H'J} \{\hat{\tilde{P}}_{qh'j}\}_\nu \{\hat{\tilde{P}}_{qh'j}\}_\nu^*. \qquad (18)$$

Notice the sum over range is replaced by a sum over the wavenumbers defined by the mask.

Because the CSDM involves frequency averaging, care must be taken in establishing an isomorphism between the modes and the column vectors of the SVD matrix $\mathbf{U}$. Rigorously, the isomorphism is only valid when the depth dependence of the modes can be considered frequency independent. In practice this means that the frequency must be limited to a bandwidth over which the modal depth dependence is nearly constant. In this case the extracted modes can

be thought of as an average over the bandwidth, $\phi_{q\nu j} \rightarrow \langle \phi_{q\nu} \rangle$, so that upon substitution by Eq. (17), Eq. (18) takes the form

$$C_{qq'} = \langle \phi_{q\nu} \rangle \langle \phi_{\nu q'}^* \rangle \tilde{\delta}_\nu, \qquad (19)$$

where

$$\tilde{\delta}_\nu \equiv \frac{2\pi \langle \phi_\nu(z') \rangle \langle \phi_\nu^*(z') \rangle}{H'J\rho^2(z')} \left(\sum_{h',j=1}^{H'J} \left(\frac{|\tilde{S}_j|^2}{\sqrt{k_{\nu j} k_{\nu j}}} \right) \gamma_{\nu h'j} \gamma_{\nu h'j}^* \right). \qquad (20)$$

Clearly then, when mode ν is well resolved in wavenumber [as is the case with mode 4 in Fig. 2(c)], performing the singular value decomposition of $\mathbf{C}$ yields the matrix $\mathbf{U}$ such that the column vector $\vec{U}_1$ is isomorphic to the frequency-averaged depth-dependent mode ν with corresponding singular value s_{11} proportional to $\tilde{\delta}_\nu$,

$$U_{q1} \leftrightarrow \langle \phi_{q\nu} \rangle, \quad s_{11} \leftrightarrow \tilde{\delta}_\nu. \qquad (21)$$

In the case where the masked mode is not well resolved in wavenumber, it is not possible to define a mask that contains only mode ν and the SVD will not work for this mode.

To determine the resolvability of the modal wavenumbers, we consider the range/wavenumber relations of numerical Fourier transform implementations such as the fast Fourier transform (FFT),

$$\Delta r = \frac{2\pi}{d\tilde{k}}, \quad dr = \frac{2\pi}{\Delta\tilde{k}}. \qquad (22)$$

The mode ν wavenumber will be well resolved when the rotated wavenumber sampling interval is much less than the minimum interval separating adjacent modes within the bandwidth, $d\tilde{k} \ll |k_{\nu j} - k_{(\nu-1)j}|_{min}$. This condition determines the minimum range aperture required to resolve mode ν, Δr_ν, in wavenumber,

$$\Delta r_\nu \gg \frac{2\pi}{|k_{\nu j} - k_{(\nu-1)j}|_{min}}. \qquad (23)$$

Because $|k_{\nu j} - k_{(\nu-1)j}|_{min}$ decreases as ν decreases, this effectively sets a lower limit on the lowest mode that is extractable for a given range aperture. At the other end of the spectrum, the width of the rotated wavenumber domain, $\Delta\tilde{k}$, of the f-k diagram determines the highest-order mode appearing in the f-k structure. Then the upper limit on ν that is extractable for a given range sampling interval is set by the condition

$$dr = \frac{2\pi}{|k_{\nu j} - k_{1j}|_{max}}. \qquad (24)$$

In the context of ocean applications, the range aperture needed to resolve mode ν (in the low mode order limit, $\nu \ll 2D/\lambda_w$) is governed by the inequality

$$\Delta r_\nu \geqslant \frac{8D^2}{\lambda_w(2\nu-1)}. \tag{25}$$

Here D is the water depth, and $\lambda_w \equiv f_0/c_w$ is the acoustic wavelength in the water (characterized by sound speed c_w) at frequency f_0. For example, in order to extract all the propagating modes excited by a moving broadband source radiating at a central frequency of 150 Hz in a 100-m-deep waveguide, the source would have to cover greater than 8 km in range aperture. However, extracting modes 3 and greater would require only $\frac{1}{5}$ the range aperture. The sampling interval requirement can be written in terms of physical parameters as $dr \approx \lambda_w(c_b/|c_b-c_w|)$. For water and bottom sound speeds $c_w = 1500$ m/s and $c_b = 1800$ m/s, the sampling interval can be as large as 18 m ($=6\lambda_w$).

III. RESULTS: APPLICATION OF THE MI METHOD TO DATA

In this section we include comparisons of the depth-dependent mode extraction results of the MRMF and MI methods. Mode extractions using both simulated and experimental data are shown. The simulated mode extractions are useful because they allow us to compare the extracted modes with the true modes used to generate the simulated pressure fields. The experimental mode extractions provide a demonstration that the MI mode extraction method is experimentally applicable, a crucial first step proof of concept for shallow ocean applications. In the following sections, the vectors returned by the SVD operation are called SVD vectors. This is done in order to differentiate them from the theoretical modes used to generate the simulation. This differentiation also serves to emphasize that these objects are not always faithful representations of (isomorphic to) the true modes of acoustic propagation.

The effectiveness of range averaging (MR) and ambient noise decoupling SVD methods over a full-spanning array have been demonstrated.[9] Except for degeneracy limitations, our work shows that the MRMF technique works as well as the MI method when applied to data over a full spanning array. Therefore, we shall not dwell on full-spanning applications, and instead discuss the more interesting case of partial-spanning applications.

A. Comparison of MI and MRMF methods applied to a partial-spanning VLA simulation

Range and frequency-averaging CSDM decoupling scheme SVD methods, such as the MR, MF, and MRMF methods, have been applied to experimental data. The success achieved by such methods has been demonstrated to be limited by the constraints mentioned in Sec. II C, namely the degree of mode decoupling in the CSDM, the degeneracy of the singular values, and the full-spanning requirement. Of these, the latter is by far the most stringent constraint. Whereas some of the modes may be successfully extracted under conditions in which only some of the modes are decoupled and some of the modes are degenerate, these methods fail to extract any of the modes when the VLA does not span the entire water column. The MI technique, on the other
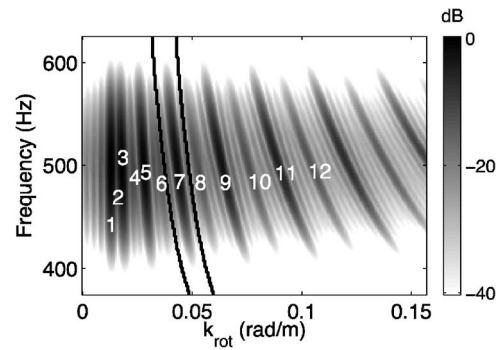
FIG. 3. Depth-averaged modulus of the rotated f-k structure of a simulated waveguide. Propagating modes appear as dark gray curved regions. Modes are enumerated for clarity. Modes 2, 4, and 6 were weakly excited by the source. Hence, they appear to be missing. The mode 7 mask is shown (solid black lines) as an example. Simulation parameters: Pekeris waveguide, $D = 60$ m, $c_w = 1500$ m/s, $c_b = 1800$ m/s, $\rho_b = 1800$ kg/m^3, $dr = 20$ m, $\Delta r = 2$ km, $r_c = 3$ km, and $z' = 30$ m.

hand, can extract all modes for which the wavenumber resolution is adequate over a partial-spanning array. In this section we use simulation to demonstrate the MRMF and MI methods' respective abilities to extract the depth-dependent modes of acoustic propagation over a partial-spanning VLA. Because the data is simulated it is possible to compare the extracted modes to theory.

We simulated the VLA pressure fields resulting from distant broadband (400–600 Hz) sources in a shallow ocean Pekeris waveguide characteristic of near shore environments of interest. The water depth was 60 m. The VLA consisted of 33 evenly spaced elements with element 1 at a depth of 12 m and element 33 at depth 44 m. The resulting f-k dispersion is shown in Fig. 3. Clearly, there are many propagating modes. The first 12 modes are labeled for clarity. Modes 2, 4, and 6 are weakly excited and do not appear on the dispersion plot. This results from the source depth being chosen to be at the midpoint of the water column.

Figure 4(a) shows the SVD vectors produced by the MRMF method applied to the simulated pressure field. The MRMF SVD vectors were matched to all the theoretical modes. Each SVD vector is plotted with its closest matching theoretical mode. On the other hand, as seen in Fig. 4(b), when it comes to extracting depth-dependent modes projected over only a portion of the water column, the MI method performs remarkably well. Whereas in the MRMF case we do not automatically know to which mode a given SVD vector corresponds, in the MI case we know which mode we are attempting to extract. As the wavenumber separation between the modes decreases, each mask contains increasing residuals from adjacent modes. If the mode for which the mask is defined is weakly excited compared to an adjacent mode, the extraction process will result in a mix of adjacent modes. As a result, we see that the weakly excited modes (2, 4, and 6) are not well extracted. The remaining modes are well extracted, with one caveat. It is possible to properly determine mode 1 in this example only because mode 2 is absent. Otherwise, due to the limited wavenumber resolution, it would not have been possible to isolate mode 1 from mode 2.
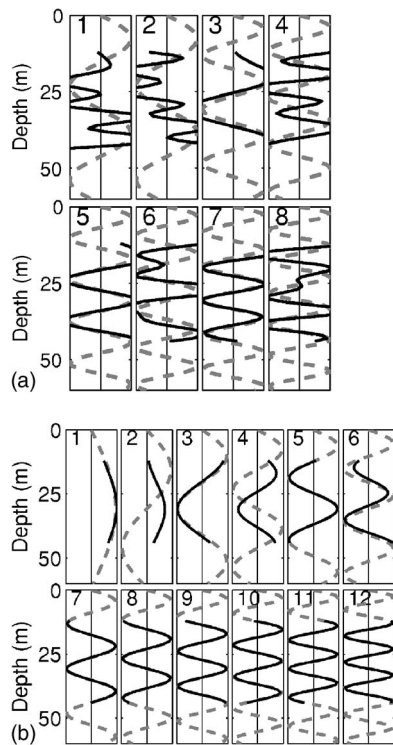
FIG. 4. Comparison of MRMF and MI method mode extractions using simulated pressure field data over part of water column (12–44 m). (a) Depth-dependent modes extracted using the MRMF method. (b) Depth-dependent modes extracted using MI method. The SVD vectors (solid lines) are plotted with the theoretical modes (dashed lines) used to generate the data. Each numbered panel shows one SVD vector along with corresponding theoretical mode. In the MRMF case only the real components of the eight best matching MRMF SVD vectors are shown along with their best matching theoretical modes. Simulation parameters: same as Fig. 3.

B. MI method applied to experimental data

Figure 5 shows mode extraction results of the MI method applied to experimental data. The experiment was designed to mimic a typical shallow ocean range-independent environment at approximately a 1:10 000 scale. Whereas ocean environments of interest typically involve ranges on the order of kilometers, depths on the order of tens of meters, and wavelengths on the order of meters, the experimental ranges were on the order of a meter (ranges between 390 and 900 mm), the experimental depth was on the order of tens of millimeters ($D \approx 27$ mm), and the acoustic wavelength was on the order of millimeters ($\lambda \approx 2.5$ mm). The waveguide consisted of approximately 27 mm ($\approx 10.5\lambda$) of fresh water over a homogenous sand bottom. A vertical line array (VLA) measured the pressure fields individually from $H=301$ identical broadband ultrasonic sources (400 kHz bandwidth at 600 kHz carrier frequency) located at evenly spaced range intervals between 390 and 990 mm ($\approx 160 - 400\lambda$). The VLA consisted of $Q=31$ evenly spaced hydrophones spanning most of the water column (see Fig. 1). The absolute depths of the array elements were unknown. The minimum and maximum source ranges were constrained by the physical dimensions of the waveguide apparatus. The water depth and source frequency were chosen to produce ten propagating modes in a regime where the wavenumber resolution produced by the range aperture would be adequate
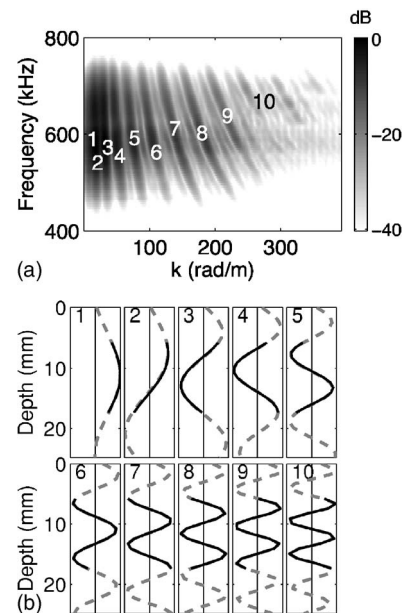
FIG. 5. MI method mode extraction results using ultrasonic experimental pressure field data over full-spanning (VLA elements 1–31) and partial-spanning (elements 8–22) VLAs. (a) Depth-averaged modulus of the rotated f-k structure of the experimental waveguide. Propagating modes appear as dark gray curved regions. Modes are enumerated for clarity. (b) MI SVD vectors from partial-spanning VLA data are indicated with solid lines. Full-spanning VLA SVD vectors appear as dashed lines. Only modes 1–10 are shown. Experimental parameters: fresh water over sand bottom, $D =27$ mm, $c_w \approx 1490$ m/s, c_b unknown, ρ_b unknown, $z' \approx 26$ mm, $dr =1$ mm, $\Delta r=600$ mm, and $r_c=690$ mm.

for resolving most of the modal wavenumbers. The source depth was chosen to be near the bottom to ensure that all the modes would be excited by the source. As seen in the resulting f-k diagram of Fig. 5(a), there were indeed ten propagating modes, seven of which were well isolated. Figure 5(b) shows the MI SVD vectors resulting from both full-spanning (dashed lines) and partial-spanning (dark lines) VLA geometries. The full spanning case utilized data from all 31 VLA elements, while the partial-spanning extraction used data from elements 8–22 ($\approx 6-17$ mm). Excluding mode 1, the full-spanning SVD vectors appear remarkably sinusoidal. Furthermore, the partial-spanning SVD vectors coincide perfectly with the corresponding portions of the full-spanning vectors. As mentioned, this would not be the case for the MRMF method. The wavenumber resolution was not adequate to isolate modes 1 and 2 (and perhaps mode 3).

IV. SUMMARY AND DISCUSSION

We have demonstrated the conditions under which the singular value decomposition (SVD) of a cross-spectral density matrix (CSDM) is expected to yield column vectors that are isomorphic to the depth-dependent modes of acoustic propagation. Typically, in order for the isomorphism to be valid (*i*) the CSDM must sample the entire half-space [the vertical line array (VLA) over which acoustic data is recorded must at least span the entire water column]; (*ii*) the modes must be decoupled in the CSDM; and (*iii*) the singular values returned by the SVD must be nondegenerate. The constraints posed by these conditions limit the applicability

of the current collection of SVD and eigenvalue decomposition mode extraction methods in the literature. The full spanning requirement alone is a formidable constraint to overcome in ocean applications. Add to this the degeneracy limitation, which arises whenever two or more modes arrive at the VLA with nearly equal intensities, and it is unlikely that such techniques will have significant practical applications in the ocean.

In order to address these shortcomings, we introduced a variation on the current SVD methods which incorporates a knowledge (by measurement) of the dispersion characteristics of the waveguide environment. Because the modes are dispersed and localized in the frequency-wavenumber (f-k) domain, it is possible to isolate and deal with them individually. Our mode isolation (MI) technique proved effective at extracting the modes of acoustic propagation over a VLA covering only a portion of the water column. In addition, because the modes are extracted individually, the MI technique does not suffer from degeneracy limitations.

The MI method is subject to two constraints: the soft constraint that the bandwidth be limited to a regime in which the depth dependence of the modes is weakly frequency dependent, and the hard constraint imposed by the available range aperture. The latter constraint, related to the CSDM mode decoupling requirement mentioned in (*ii*) above, determines the wavenumber resolution. It is the wavenumber resolution that dictates whether or not a given modal wavenumber will be well localized in the f-k domain. The MI technique is expected to perform well when the modes are well localized in the f-k domain regardless of the sound speed profile.

As a practical matter, it is worth mentioning that it is not necessary to determine the absolute ranges of the sources under the condition that they are evenly spaced. In this case transforming to wavenumber leads to the same fundamental f-k modal dispersion. Any error in estimating the range interval between the sources results only in a simple linear wavenumber scaling of the modal dispersion. In ocean applications, time gating the acoustic signal from a passing ship traveling at a constant velocity far from its closest point of approach might fulfill our requirement of evenly spaced acoustic sources. A misestimation of the ship's velocity would result in a linear scaling of the ranges and would not adversely affect the application of the MI mode extraction technique.

[1] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).

[2] C. S. Clay, "Waveguides, arrays, and filters," Geophysics **31**, 501–506 (1966).

[3] T. C. Yang, "A method of range and depth estimation by modal decomposition," J. Acoust. Soc. Am. **82**, 1736–1745 (1987).

[4] G. R. Wilson, R. A. Koch, and P. J. Vidmar, "Matched mode localization," J. Acoust. Soc. Am. **84**, 310–320 (1988).

[5] A. B. Baggeroer, W. A. Kuperman, and P. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**(4), 401–424 (1993).

[6] J. R. Buck, J. C. Preisig, M. Johnson, and J. Catipovic, "Single-Mode Excitation in the Shallow-Water Acoustic Channel Using Feedback Control," IEEE J. Ocean. Eng. **22**(2), 281–291 (1997).

[7] S. N. Wolf, D. K. Cooper, and B. J. Orchard, "Environmentally adaptive signal processing in shallow water," Oceans '93, Engineering in Harmony with Ocean Proceedings, **1**, 99–104 (1993).

[8] P. Hursky, W. S. Hodgkiss, and W. A. Kuperman, "Matched field processing with data-derived modes," J. Acoust. Soc. Am. **109**, 1355–1366 (2001).

[9] T. B. Neilsen and E. K. Westwood, "Extraction of acoustic normal mode depth functions using vertical line data," J. Acoust. Soc. Am. **111**, 748–756 (2002).

[10] W. A. Kuperman and F. Ingenito, "Spatial correlation of surface generated noise in a stratified ocean," J. Acoust. Soc. Am. **67**, 1988–1996 (1980).

[11] G. V. Frisk and J. F. Lynch, "Shallow water waveguide characterization using the Hankel transform," J. Acoust. Soc. Am. **76**, 205–216 (1984).

[12] L. Greengard and J.-Y. Lee, "Accelerating the Nonuniform Fast Fourier Transform," SIAM Rev. **46**(3), 443–454 (2004).

[13] B. Nicolas, J. Mars, and J.-L. Lacoume, "Geoacoustical parameters estimation with impulsive boat-noise sources," IEEE J. Ocean. Eng. **28**(3), 494–501 (2003).

Matched field processing with contrast maximization

K. C. Shin

Underwater R&D Institute, NEX1 FUTURE Co., Ltd. Gyeonggi-do, 449–910, Korea

J. S. Kim[a]

Department of Ocean Engineering, Korea Maritime University, Busan, 606-791, Korea

The high sidelobe level of the ambiguity surface has been a major drawback of matched field processing in ocean waveguides compared to the free space beamforming. In this paper, a contrast-maximized optimization scheme that enhances the contrast between the acoustic power output of the source region and the power output of the total region of interest is presented. The method utilizes the signal vectors determined from the contrast-maximizing formulation. Through numerical simulation and oceanic data analysis, the presented algorithm is shown to be effective in improving the contrast between the target and the averaged background sidelobe level. It was also shown that, even when the resolution of the source region in conventional matched field processing is degraded due to severe bottom attenuation, the resolution of the source is not affected in the contrast-maximized method. The same principle and method can be applied to the time-reversal processing to maintain the focal size in an ocean environment with high bottom attenuation. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1993131]

I. INTRODUCTION

Matched field processing (MFP) is an approach for localizing acoustic sources by comparing acoustic data with solutions of the wave equation. MFP can also be viewed as a generalized beamforming technique that takes advantage of the spatial structure of the sound field of a signal propagating in the ocean to provide improved localization performance compared to the conventional plane wave beamformer.[1] One of the advantages associated with the use of the MFP is that the ambiguity surface provides the localization result in range and depth terms, while the plane wave beamformer gives only the direction.

However, since the matched field processor exploits the spatial complexities of the acoustic field in the waveguide, the ambiguity surface produced by MFP typically contains, in addition to the signal peak, false peaks or sidelobes, which make the localization of an acoustic source ambiguous. The conventional MFP (CMFP) in an ocean waveguide tends to suffer from higher sidelobe levels than the conventional beamforming in free space, where the sidelobe level is around 13 dB less than the mainlobe level.

In order to avoid the ambiguity or to reduce the sidelobe effects, incoherent and coherent processors are widely used.[2–6] Sidelobe effects are additionally mitigated by using adaptive MFP (AMFP)[7–10] which minimizes the beamformer energy interfering with the signal by imposing a constraint that a signal in look-direction is passed without distortion.

The purpose of this paper is to present a method to suppress the sidelobe level of ambiguity surface produced by MFP in an ocean waveguide. In order to achieve lower side-lobes level, an adaptivity was introduced to find the signal vector to be correlated, or to be beamformed in the case of plane wave beamforming. The signal vector was determined from an optimization scheme to maximize the contrast between the source region and the search region of interest. The same principle has been applied to the generation of an acoustically bright spot in active noise control based on the contrast-maximizing optimization scheme using multiple sources in a free space.[11] In this paper, an optimization method for contrast maximization was extended to MFP in an ocean waveguide. The same principle and method can be applied to the time-reversal processing in order to reduce the focal size in an ocean waveguide with high bottom attenuation.

The remainder of this paper is organized as follows. In Sec. II, the two optimization schemes, maximization of power and contrast, are reviewed and the issues in waveguides are discussed. The simulated results of MFP in an ocean waveguide are presented and discussed in Sec. III. In Sec. IV, ocean acoustic data are analyzed to show the effectiveness of the contrast-maximized method in MFP. Section V summarizes the work.

II. THEORY

The conventional MFP is reviewed in Sec. II A, and then reformulated from the view point of power-optimization problem in Sec. II B. Finally, the theory on the contrast-maximized optimization method in MFP is presented.

In order to demonstrate the applicability of the contrast-maximized optimization method in MFP, two regions were defined as shown in Fig. 1. The search region was denoted as V_t, and the source region as V_s, which was assumed included inside V_t.

---

[a]Author to whom correspondence should be addressed. Electronic mail: jskim@hhu.ac.kr
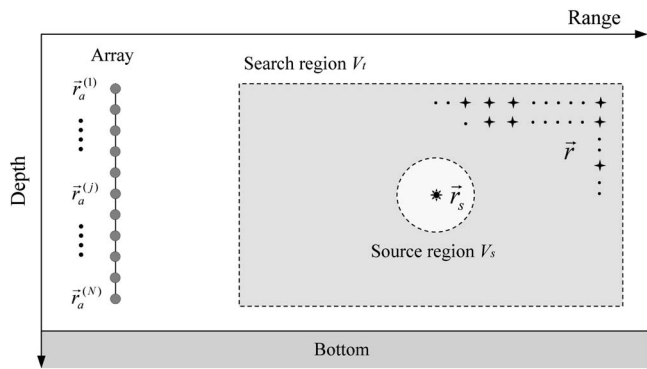
FIG. 1. (Color online) Schematics of acoustic regions in an ocean wave-guide.

A. Review of the conventional MFP

Defining the received acoustic pressure at the jth array element as $d(\vec{r}_a^{(j)}|\vec{r}_s)$ in frequency domain, MFP can be formulated as a correlation process between the replica generated from the propagation model and the data measured at an array:

$$P(\vec{r}) = \sum_{j=1}^{N} d^*(\vec{r}_a^{(j)}|\vec{r}_s) g(\vec{r}|\vec{r}_a^{(j)}) = \mathbf{d}^+ \mathbf{g}(\vec{r}|\mathbf{r}_a) = \mathbf{d}^+ \mathbf{w}, \qquad (1)$$

where the subscripts a and s denote the array and source, respectively, and $g(\vec{r}|\vec{r}_a^{(j)})$ represents the normalized Green's function at the arbitrary location $\vec{r}$ propagated from the jth array element location $\vec{r}_a^{(j)}$ as shown in Fig. 1. The number of the array elements is denoted as N. The superscripts $(\ )^*$ and $(\ )^+$ denote complex conjugate and Hermitian transpose, respectively. In a vector notation, the data vector, $\mathbf{d}$, and the replica vector, $\mathbf{w}$, which is mathematically equivalent to the normalized Green's function $\mathbf{g}$, are $(N \times 1)$ column vectors. Note that the position vectors are written in *italic* letters with arrows and column vectors and matrices are written in boldface letters.

B. Conventional MFP as a power optimization problem

Now, in order to introduce the "acoustical contrast" concept to MFP, the space-averaged power output is defined as the acoustical brightness of a given region. Replacing the signal vector $\mathbf{d}$ by $\mathbf{x}$, which is a more generalized form of the data vector to be correlated with the replica vector, the space-averaged power output σ_s and σ_t in the regions V_s and V_t can be formulated as

$$\sigma_s = \frac{1}{V_s} \int_{V_s} P(\vec{r}) P^*(\vec{r}) \, dV = \mathbf{x}^+ \mathbf{R}_s \mathbf{x}, \qquad (2)$$

$$\sigma_t = \frac{1}{V_t} \int_{V_t} P(\vec{r}) P^*(\vec{r}) \, dV = \mathbf{x}^+ \mathbf{R}_t \mathbf{x}, \qquad (3)$$

where

$$\mathbf{R}_s = \frac{1}{V_s} \int_{V_s} \mathbf{g}(\vec{r}|\mathbf{r}_a) \mathbf{g}^+(\vec{r}|\mathbf{r}_a) \, dV = \frac{1}{V_s} \int_{V_s} \mathbf{w}\mathbf{w}^+ \, dV, \qquad (4)$$

$$\mathbf{R}_t = \frac{1}{V_t} \int_{V_t} \mathbf{g}(\vec{r}|\mathbf{r}_a) \mathbf{g}^+(\vec{r}|\mathbf{r}_a) \, dV = \frac{1}{V_t} \int_{V_t} \mathbf{w}\mathbf{w}^+ \, dV. \qquad (5)$$

Each element of $\mathbf{R}_s$ and $\mathbf{R}_t$ represents the spatial correlation of the Green's functions, $\mathbf{g}$, in the two regions produced by each array element. It is worth noting that the correlation value in Eq. (2) has a maximum at the point source, where V_s is given by a point located at $\vec{r}_s$. For this reason, $\mathbf{R}_s$ is the covariance matrix of the replica vector for the acoustic field at the array propagated from the sources. In a similar way, the covariance matrix $\mathbf{R}_t$ can be constructed from replica vectors derived from the propagation model as shown in Eq. (5).

The conventional matched field processor or the Bartlett processor can be considered as a power optimization scheme, which gives a maximum power in a source region for given input power. In order to have maximum power at the source region, the problem can be stated as

$$\max_{\mathbf{x}} \sigma_s = \mathbf{x}^+ \mathbf{R}_s \mathbf{x} \ \text{subjected to} \ P_0 = \mathbf{x}^+ \mathbf{x}. \qquad (6)$$

The above equation may also be formulated as a pure (unconstrained) optimization problem using a form known as the Rayleigh quotient.[12,13] Therefore, maximizing the σ_s under given input power constraint is equivalent to maximizing the following ratio:

$$\alpha = \frac{\sigma_s}{P_0} = \frac{\mathbf{x}^+ \mathbf{R}_s \mathbf{x}}{\mathbf{x}^+ \mathbf{x}}. \qquad (7)$$

The ratio α represents the quantity of space-averaged output power generated in the source region by unit input power. The optimum data vector $\mathbf{x}_\alpha$, which maximizes the Rayleigh quotient in Eq. (7), corresponds to the eigenvector having the largest eigenvalue of $\mathbf{R}_s$. In conventional MFP, the optimum data vector $\mathbf{x}_\alpha$ reduces to the unprocessed data vector at the array $\mathbf{d}$ with a rank equal to one.

C. MFP as an acoustical contrast optimization problem

While the optimization problem in Sec. II B is to maximize the correlation in the source region, another optimization scheme to maximize the contrast between the source region and the search region can be formulated. The derivation of the contrast-maximized matched field processor is very similar to that of the maximum signal-to-interference ratio (SIR) for adaptive beamforming.[13]

The signal vector can be chosen to directly maximize the contrast between the source region and the search region. Assuming that $\mathbf{R}_s$ and $\mathbf{R}_t$ are known [which are the cross-spectral density matrix calculated from the replica vector in Eqs. (4) and (5)], we may choose to maximize the ratio of the space-averaged power output in the source space (σ_s) and the space-averaged power output in the search space (σ_t). That corresponds to the maximization of the following cost function:
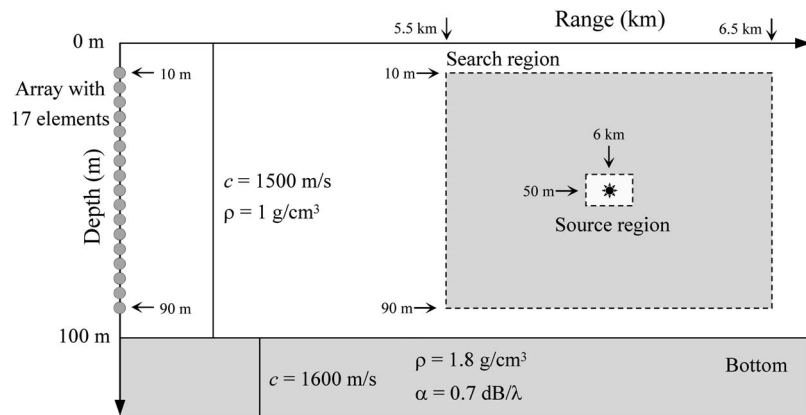
FIG. 2. (Color online) Setup for numerical simulation in the Pekeris waveguide.

$$\beta = \frac{\sigma_s}{\sigma_t} = \frac{\mathbf{x}^+\mathbf{R}_s\mathbf{x}}{\mathbf{x}^+\mathbf{R}_t\mathbf{x}}. \tag{8}$$

This function, expressed as the ratio of two quadratic forms, essentially determines the ratio of the spaced-averaged output power in V_s to that in V_t (Fig. 1) and is well known as Rayleigh quotient. Now the problem is reduced to the determination of the optimum signal vector, which maximizes the Rayleigh quotient. Taking the derivative of Eq. (8) with respect to $\mathbf{x}$ and setting it to zero, we obtain

$$\mathbf{R}_s\mathbf{x} = \beta\mathbf{R}_t\mathbf{x}, \tag{9}$$

which appears to be a joint eigenproblem. The value of β is bounded to the minimum and maximum eigenvalues of the symmetric matrix $\mathbf{R}_t^{-1}\mathbf{R}_s$. The signal vector that maximizes β is given by the following form:

(a)

(b)
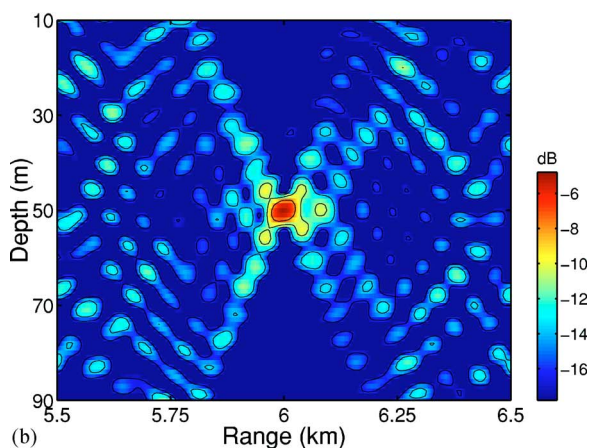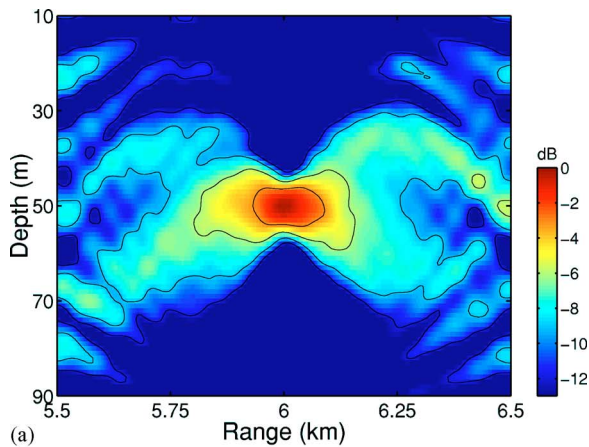
FIG. 3. (Color online) Simulation results. (a) Ambiguity surfaces by conventional MFP. (b) Ambiguity surfaces by contrast maximized MFP.

(a)

(b)
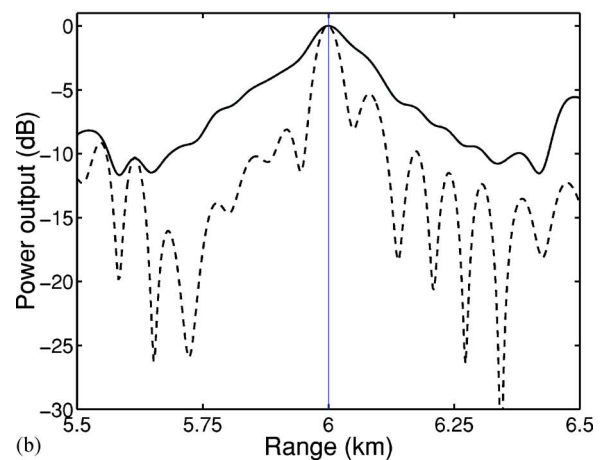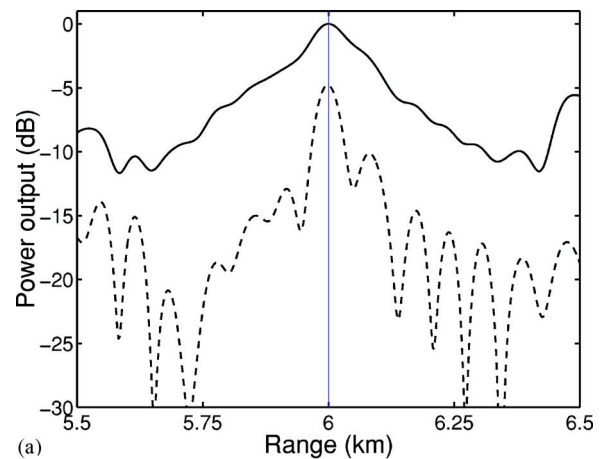
FIG. 4. (Color online) Ambiguity surfaces as function of range for point source case: (a) unnormalized plot and (b) normalized plot. The solid and dashed lines represent results by conventional MFP and contrast maximized MFP, respectively.

Shin and Kim: Matched field processing with contrast maximization
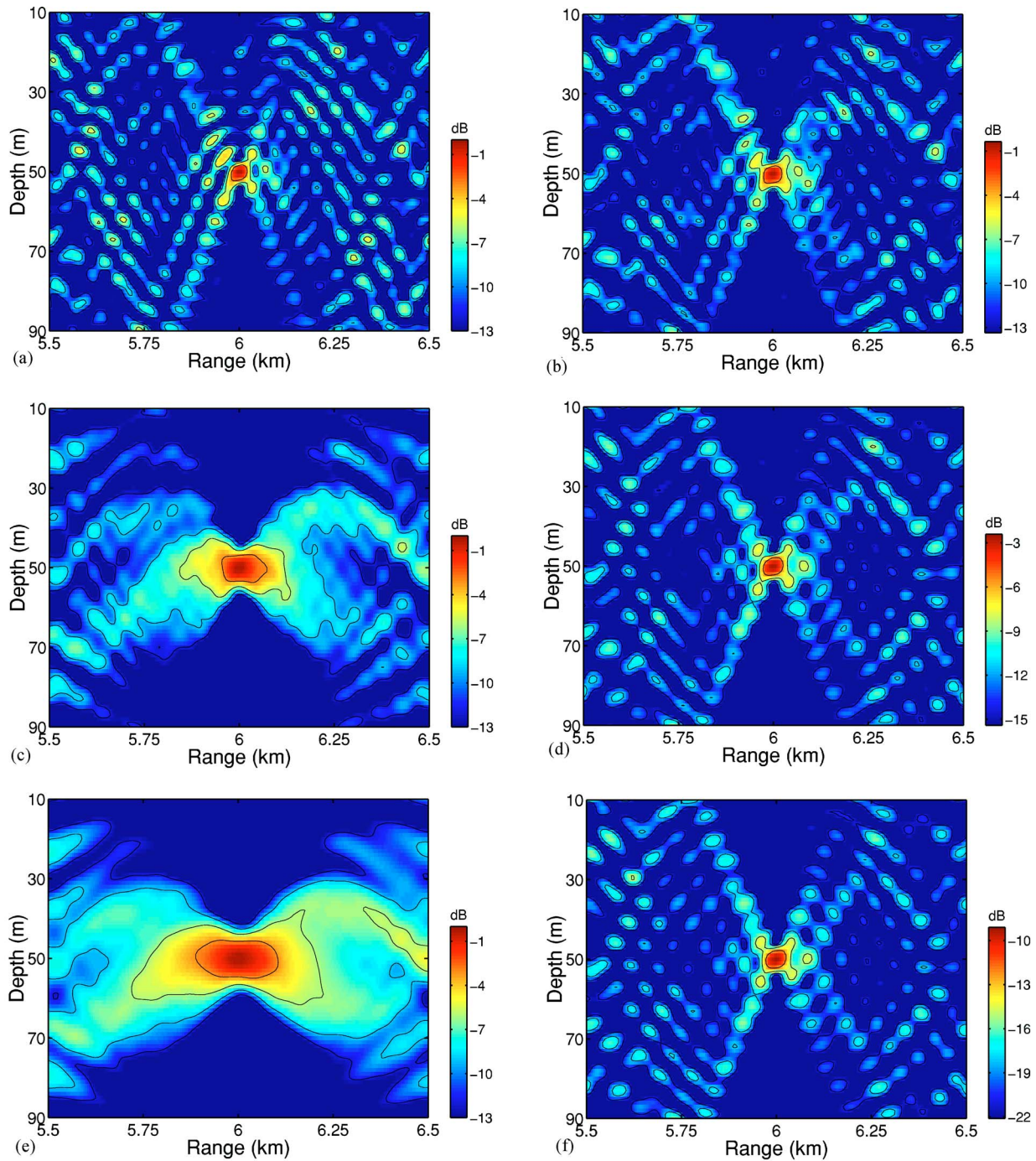
FIG. 5. (Color online) Simulation results for various bottom attenuation. (a), (c), and (e) were with the conventional signal vector. (b), (d), and (f) were with a contrast signal vector. The attenuation for (a) and (b) is $0\text{dB}/\lambda$, and $0.5\text{dB}/\lambda$, for (c) and (d), $1.0\text{dB}/\lambda$, for (e) and (f).

$$(\mathbf{R}_t^{-1}\mathbf{R}_s)\mathbf{x}_\beta = \beta_{\max}\mathbf{x}_\beta. \qquad (10)$$

The solution to the above Eq. (10) is well known and is equal to the eigenvector corresponding to the maximum eigenvalue of $\mathbf{R}_t^{-1}\mathbf{R}_s$. Thus, if the pressure fields produced by each control source can be predicted within V_t, which is inherently assumed in MFP, it is possible to construct the covariance matrix $\mathbf{R}_s$ and $\mathbf{R}_t$ based on Eqs. (4) and (5). Now, the optimal signal vector $\mathbf{x}_\beta$ obtained from the data vector $\mathbf{d}$ can be calculated and used as the signal vector to be correlated in the contrast-maximized optimization scheme.

In summary, we define the variables and state how they

are found. The data vector measured at the array is denoted as $\mathbf{d}$, and the processed data vector to be used for a different optimization scheme, such as contrast maximization, is expressed as $\mathbf{x}$. The processed data vectors, $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$, in Secs. II B and II C, represent the generalized data vectors for conventional MFP and contrast-maximized MFP and can be calculated from the eigenanalysis. Finally, the covariance matrices $\mathbf{R}_s$ and $\mathbf{R}_t$ are found from Eqs. (4) and (5), respectively.

III. SIMULATION IN AN OCEAN WAVEGUIDE

In Secs. II B and II C, two types of signal vector were considered. One is the conventional data vector $\mathbf{x}_\alpha$, which

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005   Shin and Kim: Matched field processing with contrast maximization   1529
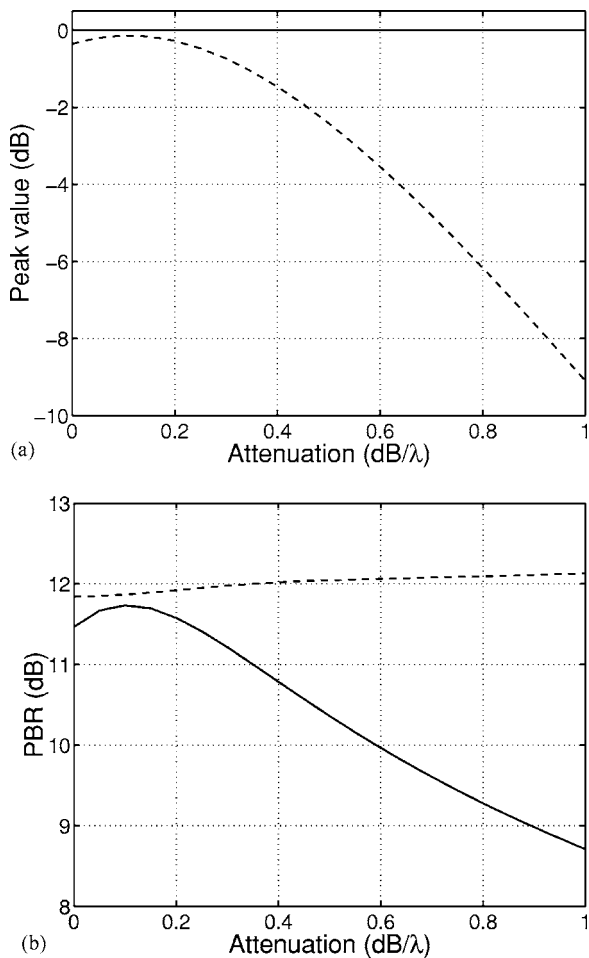
FIG. 6. Variation of (a) peak value and (b) PBR as functions of the bottom attenuation. The solid and dashed lines represent results by conventional and contrast signal vector, respectively.

gives a maximum power in a source location for given input power of $\mathbf{x}^+\mathbf{x}$. The other is the contrast signal vector $\mathbf{x}_\beta$, which is to maximize the contrast between the finite region containing source and the entire search region. In this section, we show that the peak-to-background resolution (PBR) of a contrast-maximized matched field processor with the contrast signal vector is improved by about 3 dB compared to the Bartlett processor.

The experimental setup for a numerical simulation is shown in Fig. 2. The center frequency of the sound source was 500 Hz with unit strength. The field was sampled by a vertical line array containing 17 hydrophones spaced 5 m apart ranging from 10 to 90 m in depth. The ambiguity surfaces were computed for source ranges between 5.5 and 6.5 km away from the array and at source depths of 10 to 90 m. As propagation model, KRAKEN[14] was used throughout the simulation.

A. Statistical quantification of the ambiguity surface

To quantify the performance of overall sidelobe reduction or background suppression for two types of signal vectors, a measure of peak value (P) relative to mean background level (μ) was used. It was called peak-to-background resolution (PBR) and was defined as[15]

$$ \text{PBR} = 10 \log_{10}\left(\frac{P-\mu}{\mu}\right) \text{ dB}, \tag{11} $$

where the mean background level was calculated by excluding source region. We did not infer statistical behavior from them, but merely used them as a measure of "goodness."

B. Interpretation and analysis

Figure 3(a) shows the ambiguity surface produced by the conventional signal (data) vector in the case of the point source. The main peak was seen at the true source location: 50-m depth and 6000-m range. A value of PBR=9.60 dB was obtained for this surface. Figure 3(a) was compared with Fig. 3(b), which is the ambiguity surface obtained by contrast-maximized optimization scheme. Figure 3(b) clearly illustrates that the main peak was still accurately located at the true source location, and additionally the size of the source resolution was greatly improved. Also, it can be seen that PBR is about 12.03 dB, and therefore has improved by about 3 dB.

Figure 4 shows the ambiguity surface as a function of range for the Bartlett processor (solid line) and the contrast method (dashed line) in the case of point source. This figure describes the important difference between two optimization techniques. In Fig. 4(a), the power output of the contrast-maximized processor is approximately 5 dB lower than the power output of the Bartlett processor at the source location. This was due to the fact that the contrast-maximized technique did not maximize the power output at the source location, but maximized the contrast between the two regions. Figure 4(b) shows that the contrast-maximized processor had an excellent sidelobe suppression in an average sense compared to the Bartlett processor.

C. Effect of bottom attenuation

The physical principle on the size of resolution for time-reversal mirror (TRM)[16] is the same as the MFP. While a propagation model was used to match the source-generated sound fields in MFP, the time-reversed signal was physically backpropagated to the original source position using a source-receiver array in a TRM. Hence, the focal size realized with a TRM can be considered as the maximum achievable resolution of a conventional MFP for a given waveguide.

The excited acoustic field can be decomposed into modes, and each mode propagates with its own group speed. During the propagation between the source and the array, some high-order modes that had more interactions with the ocean bottom were removed by waveguide attenuation. The

Shin and Kim: Matched field processing with contrast maximization
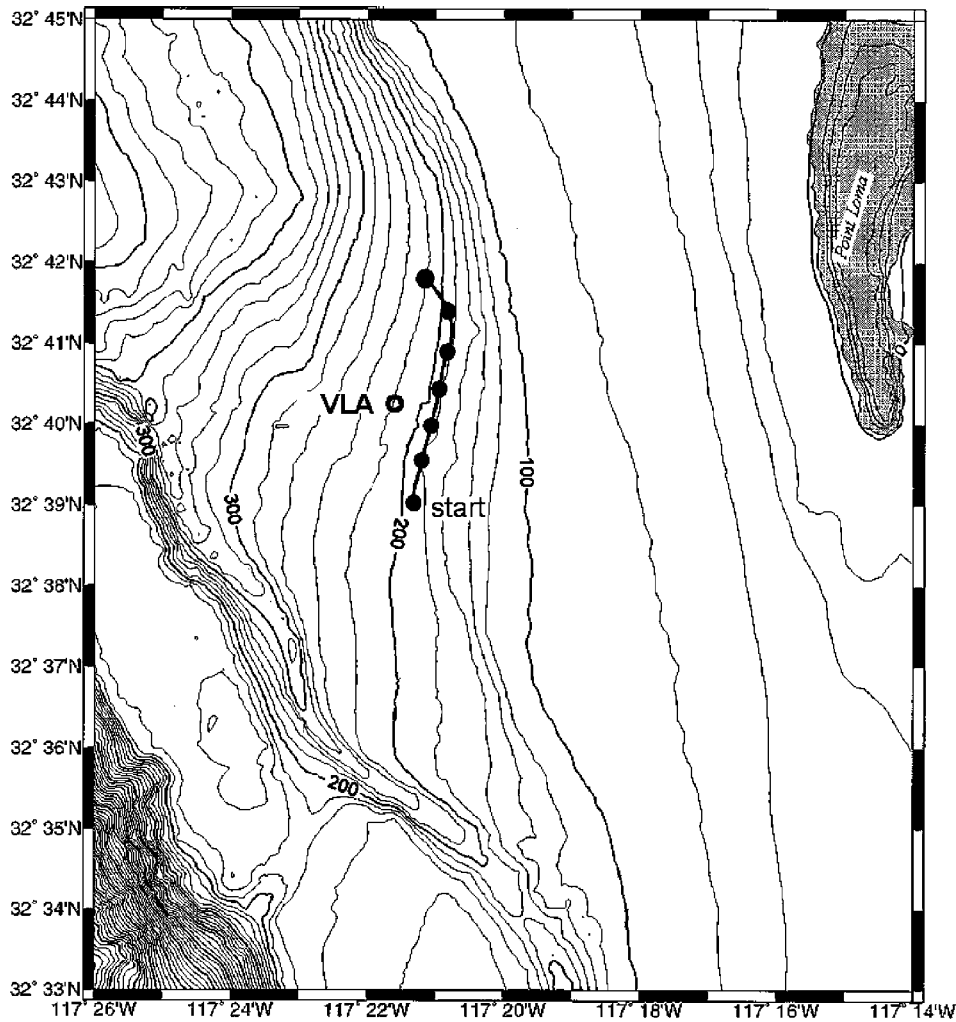
FIG. 7. The location and bathymetry for SwellEx-96 Event 59.

focal size or the resolution[11] of a MFP was determined by the remaining effective modes. The focal size in depth and range can be estimated roughly as

$$\Delta_z \approx \frac{D}{M}, \tag{12}$$

$$\Delta_R \approx \frac{2\pi}{k_1 - k_2}, \tag{13}$$

where D is the water depth, M is the number of effective modes, and k_1 and k_M are the horizontal wave numbers of the first and last effective modes, respectively. The vertical resolution is determined by the mode functions while the horizontal resolution is determined by the difference of wave numbers that is the smallest *interference pattern* of the acoustic field. The waveguide attenuation increases the focal size by reducing the number of effective modes. As a consequence, the ambiguity surface is expected to become more blurry and to have higher sidelobes than in the ideal lossless situation.

In this section, the result of numerical experiments on the effects of bottom attenuation on the size of resolution cell and the sideobe level is presented. For simulation purposes, the bottom attenuation was varied between 0 and 1 dB/λ under the point source assumption, while keeping all other parameters unchanged as shown in Fig. 2.

Figure 5 shows the ambiguity surface produced by the Bartlett processor for three bottom attenuation cases. The results by the conventional MFP are shown in Figs. 5(a), 5(c), and 5(e): the depth and range focal size increased with increasing bottom attenuation. The results by the contrast-maximized MFP, illustrated in Figs. 5(b), 5(d), and 5(f), show that there was no changes of focal size although the bottom attenuation increased. Figure 6 shows the peak value and PBR where the bottom attenuation varied from 0 to 1 dB/λ. The solid and dashed lines represent the results by the conventional MFP and the contrast-maximized MFP, respectively. In Fig. 6(a), the peak values by the conventional processor do not change with the attenuation, but the peak values by the contrast-maximized processor rapidly fall down approximately −9 dB at 1-dB/λ attenuation. However, in Fig. 6(b), the PBR by contrast-maximized processor is consistently greater than PBR by the conventional processor. Therefore, it was concluded that the contrast-maximized MFP suppressed the sidelobe levels by about 3 dB when compared to the conventional MFP.
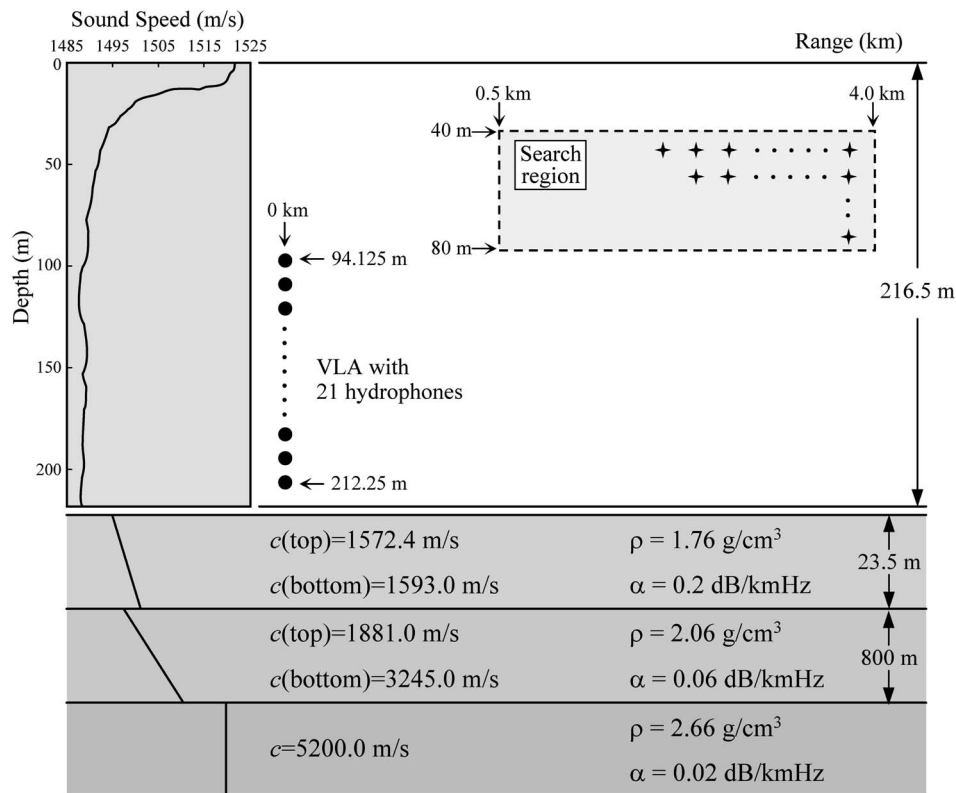
FIG. 8. (Color online) The sound speed profile and geoacoustic parameters for replica vector calculation.

IV. APPLICATION TO OCEANIC DATA

This section describes the application of the contrast-maximization technique to the SwellEx-96 data[17] to verify the applicability of the technique to the matched field processing in a shallow water waveguide. In Sec. IV A, the description of the experiment is given. In Sec. IV B, the results of the contrast-maximized technique in MFP are presented and the quantitative analysis utilizing PBR is described.

A. Description of SWellEx-96 Event 59

The detailed description of the experiment can be found in Ref. 17. Figure 7 shows a plan view of the SwellEx-96 experiment with towed source ship. The solid circle marks on Fig. 7 shows the track of the towed source every 5 min. In order to apply the contrast-maximization technique, the following eight frequencies were used: 49, 64, 79, 112, 130, 148, and 166 Hz. The field was sampled by a vertical line array containing 21 hydrophones spaced 5 m apart and ranging from 94.125 to 212.25 m in depth. The towed source was located between 0.5 and 4 km from the array and at a depth of 60 m.

The KRAKEN[14] normal mode program was used to generate the matched field replica vectors. The depth covered in generating the replica vector was 40 to 80 m with 1-m increment and the range covered was 0.5 to 4 km with 20-m increment. The sound speed profile and relevant parameters used in creating the replica vectors are shown in Fig. 8.

B. Data analysis

In order to track the underwater acoustic source, the ambiguity surface at estimated source depth was displayed based on the procedure described in Ref. 12. First, the ambiguity surfaces for range versus depth were constructed for each frequency and the surfaces were incoherently summed over the eight frequency components. Then, the frequency average of the ambiguity surface was generated for every time step and was displayed at the estimated source depth.

Figure 9(a) shows the ambiguity surface from the Bartlett processor at estimated source depth and Fig. 9(b) illustrates the results of the Bartlett processor with contrast signal vector. The source trajectory is clearly visible in both figures. However, the ambiguity surface from the Bartlett processor had higher sidelobes than the contrast-maximized Bartlett processor.

In order to perform the quantitative analysis of each ambiguity surface, PBR was calculated over the search region for every time step as shown in Fig. 10. The mean background level was calculated by excluding the 100-m range ×10-m depth surrounding the peak location. Figure 10 displays the variation in PBR from 0 to 30 min, for ambiguity surfaces. The solid and dashed lines refer to the PBRs by the conventional and the contrast signal vector, respectively. Figure 10 shows that the PBR significantly improved by about 2 dB.

(a)

(b)
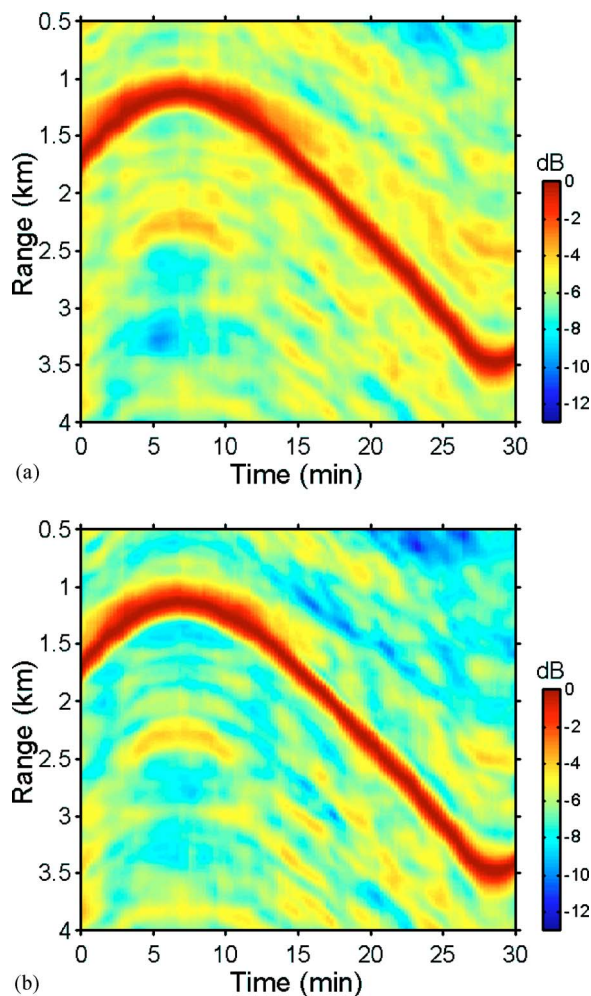
FIG. 9. (Color online) Plot of MFP-derived range versus time obtained from the eight pilot tones along the source track at that source depth with the peak MFP correlation: (a) Bartlett MFP and (b) Bartlett MFP with contrast signal vector.

V. CONCLUSION

A concept of contrast-maximized optimization was introduced to the matched field processing (MFP) in order to
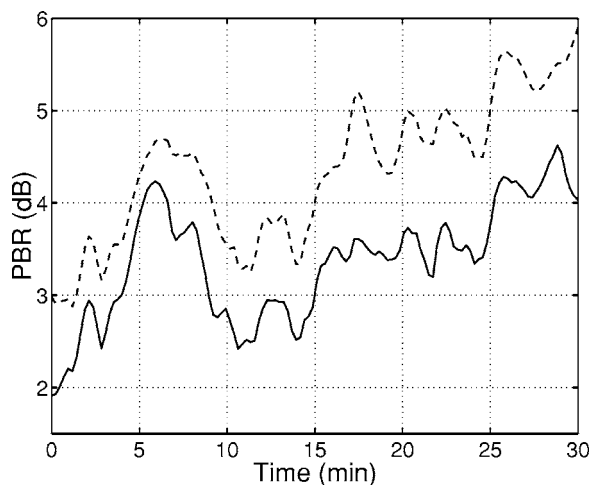
FIG. 10. Variation of PBRs as functions of the time using Bartlett incoherent broadband MFP. The solid and dashed lines represent results of PBRs by conventional and contrast signal vector, respectively.

reduce the sidelobe level of ambiguity surfaces. The signal vectors were determined from an optimization scheme to maximize the relative space-averaged power in the source region to that in the search region of interest. The application of the developed algorithm to numerical simulation and ocean experimental data showed that the peak-to-background resolution (PBR) is improved by about 3 dB. It was also observed that, even when the resolution of the source region in conventional matched field processing was degraded due to severe bottom attenuation, the resolution of the source was not affected in the contrast-maximized method. The same principle and method can be applied to the time-reversal processing to maintain the focal size in an ocean environment with high bottom attenuation.

[1] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).

[2] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**, 571–587 (1998).

[3] A. Tolstoy, "Computational aspects of matched field processing in underwater acoustics," in *Computational Acoustics*, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (North-Holland, Amsterdam, 1990), Vol. **3**, pp. 303–310.

[4] Z.-H. Michalopoulou and M. B. Porter, "Matched-field processing for broadband source localization," IEEE J. Ocean. Eng. **21**, 384–392 (1996).

[5] S. P. Czenszak and J. L. Krolik, "Robust wideband matched-field processing with a short vertical array," J. Acoust. Soc. Am. **101**, 749–759 (1997).

[6] G. J. Orris, M. Nicholas, and J. S. Perkins, "The matched-phase coherent multi-frequency matched field processor," J. Acoust. Soc. Am. **107**, 2563–2575 (2000).

[7] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1365–1376 (1987).

[8] H. Schmidt and A. B. Baggeroer, "Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch," J. Acoust. Soc. Am. **88**, 1851–1862 (1990).

[9] L. M. Zurk, N. Lee, and J. Ward, "Source motion mitigation for adaptive matched field processing," J. Acoust. Soc. Am. **113**, 2719–2731 (2003).

[10] B. F. Harrison, "The eigencomponent association method for adaptive interference suppression," J. Acoust. Soc. Am. **115**, 2122–2128 (2004).

[11] J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am. **111**, 1695–1700 (2002).

[12] C. E. Davila, "An Algorithm for Efficient, Unbiased, Equation-Error Infinite Impulse Response Adaptive Filtering," IEEE Trans. Signal Process. **42**, 1221–1226 (1994).

[13] J. Litva and T. K. Y. Lo, *Digital Beamforming in Wireless Communications* (Artech House, Boston 1996), pp. 35–56.

[14] M. B. Porter, "The KRAKEN Normal Mode Program," SACLANT Undersea Research Center SM-245, La Spezia, Italy (1991).

[15] D. R. Del Balzo, C. Feuillade, and M. M. Rowe, "Effects of water depth mismatch on matched-field processing in shallow water," J. Acoust. Soc. Am. **83**, 2180–2185 (1988).

[16] S. Kim, G. F. Edelmann, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Spatial resolution of time-reversal arrays in shallow water," J. Acoust. Soc. Am. **110**, 820–829 (2001).

[17] N. O. Booth, T. Abawi, P. W. Schey, and W. S. Hodgkiss, "Detectability of low-level broad-band signals using adaptive matched-field processing with vertical aperture arrays," IEEE J. Ocean. Eng. **25**, 296–313 (2000).

The inverse-scattering problem and global convergence

Stephen J. Norton

Geophex, Ltd., 605 Mercury Street, Raleigh, North Carolina 27603

In solving nonlinear inverse-scattering problems an iterative approach is usually required, particularly if the nonlinearity is strong. A typical approach is to use a descent algorithm to minimize a global error norm, such as the mean-square error, to arrive at a set of parameters that best predicts the data based on an assumed model of the unknown scattering object. In minimizing the error norm, the possibility of converging to a local minimum can lead to false solutions. A different strategy is proposed that is not based on gradient descent which, at least in principle, should avoid local minima and converge to the true solution (the global minimum). A simple inverse-scattering problem is used as an illustration. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1979407]

Pages: 1534–1539

I. INTRODUCTION

Nonlinear inverse-scattering problems can generally be solved only by iteration when the nonlinearity is strong. The usual strategy is to estimate the unknown scattering distribution (or a set of parameters that model the distribution) by iteratively minimizing an error functional. A common choice is the mean-squared difference between the actual observations and observations predicted on the basis of solutions to the forward-scattering problem. For highly nonlinear problems (e.g., when multiple scattering is significant), local minima in the error functional can lead to erroneous solutions. Here we propose a strategy that, in principle, avoids local minima. To illustrate the problem, suppose $\mathbf{p}$ denotes a vector of parameters that models an unknown scattering potential and let $u_m(\mathbf{p})$, $m=1,2,\ldots,M$, represent a set of M observations predicted on the basis of an assumed $\mathbf{p}$, where m could denote measurements, say, at different positions or frequencies. The field quantity $u_m(\mathbf{p})$ can be computed by solving the forward-scattering problem given $\mathbf{p}$. Now denote by $u_m(\mathbf{p}^{(true)})$ the set of actual observations, where $\mathbf{p}^{(true)}$ is the true parameter vector that we wish to determine. A common approach is to estimate $\mathbf{p}^{(true)}$ by searching for a $\mathbf{p}$ that minimizes the mean-square error

$$\mathcal{E}(\mathbf{p}) = \frac{1}{2} \sum_{m=1}^{M} [u_m(\mathbf{p}) - u_m(\mathbf{p}^{(true)})]^2. \qquad (1)$$

When minimizing $\mathcal{E}(\mathbf{p})$ by gradient descent, this approach will yield a solution whenever the gradient of $\mathcal{E}(\mathbf{p})$ vanishes, that is, whenever

$$\frac{\partial \mathcal{E}(\mathbf{p})}{\partial p_n} = \sum_{m=1}^{M} [u_m(\mathbf{p}) - u_m(\mathbf{p}^{(true)})]\frac{\partial u_m(\mathbf{p})}{\partial p_n} = 0, \qquad (2)$$

for all p_n, $n=1,2,\ldots,N$, where p_n are the components of $\mathbf{p}$. As discussed below, this is not always guaranteed to be the point $\mathbf{p}$ for which

$$u_m(\mathbf{p}) - u_m(\mathbf{p}^{(true)}) = 0 \qquad (3)$$

holds for all m. When (3) holds, we have presumably found the true solution (the global minimum), but when (2) holds, a

false solution (or local minimum) may occur, for reasons to be discussed. The approach described in this paper is based on a numerical solution to the system of nonlinear equations defined by (3) using a multi-dimensional Newton's method. When Newton's method converges, it typically finds the global minimum or else fails to converge. For theoretical reasons, when convergence occurs, Newton's method generally will not converge to a local minimum. The Appendix describes a modification of Newton's method that can improve stability, that is, widen the domain of starting values for which Newton's method converges.

In developing a nonlinear inverse-scattering algorithm, a key task in devising an iterative approach is the calculation of the derivatives of the observed field with respect to the parameters, p_n, that model the scattering object. Here we employ the adjoint method, as described in Sec. II (see Ref. 1 for a more detailed discussion of the application of the adjoint method to inverse-scattering problems). Section III describes the multi-dimensional Newton's method, and Secs. IV and V illustrate the technique with a simple example.

By way of illustration, we consider here a simple one-dimensional inverse Helmholtz problem, although the basic approach can be generalized to higher dimensions and to the time domain. Let $\gamma(x,\mathbf{p})$ denote a one-dimensional scattering potential, where $\mathbf{p}$ is a vector of parameters that models the potential. Our objective is to determine $\mathbf{p}$. The 1-D Helmholtz equation is

$$\frac{d^2u(x)}{dx^2} + k^2u(x) + \gamma(x,\mathbf{p})u(x) = S(x), \qquad (4)$$

where $u(x)$ is an appropriate field quantity, k is the wave number in the absence of the scatterer ($\gamma=0$), and $S(x)$ is the source. The components of $\mathbf{p}$ might, for example, represent the coefficients in some convenient basis expansion of $\gamma(x,\mathbf{p})$, i.e.,

$$\gamma(x,\mathbf{p}) = \sum_{n=1}^{N} p_n\phi_n(x), \qquad (5)$$

where $\phi_n(x)$ are a set of suitable basis functions, although more general parametrizations are also possible.[1]

Note that the field depends on both the wave number k and the parameter vector $\mathbf{p}$, and we will sometimes indicate this by writing $u(x,k,\mathbf{p})$. As before, denote by $u(x,k,\mathbf{p}^{(true)})$ the observed field, where $\mathbf{p}^{(true)}$ signifies the true set of parameters. Our task is to find $\mathbf{p}^{(true)}$ from the scattered field observed at some point, x_0, outside the scattering domain. In our example, we shall assume that M measurements are made at different frequencies ω_m or, equivalently, at M wave numbers $k_m = \omega_m/c$, where c is the sound speed in the homogeneous background medium (i.e., when $\gamma = 0$). Then, in the above notation used in (1), the predicted observations are $u_m(\mathbf{p}) \equiv u(x_0, k_m, \mathbf{p})$ and the actual observations are $u_m(\mathbf{p}^{(true)}) \equiv u(x_0, k_m, \mathbf{p}^{(true)})$.

II. PROPAGATION OF DERIVATIVES

In any iterative scheme, we begin by guessing the parameter vector, $\mathbf{p}$, and then, on the basis of our guess, we predict the observation, $u(x_0, k, \mathbf{p})$, and compare it to the actual observation, $u(x_0, k, \mathbf{p}^{(true)})$. A common approach, as described below, is to use a least-squares norm. We describe later an alternative to least-squares estimation. In either case, however, to update our estimate of $\mathbf{p}$, we require the partial derivatives, $\partial u(x_0, k, \mathbf{p})/\partial p$, where p is one component of $\mathbf{p}$. For brevity, we abbreviate $u(x,k,\mathbf{p})$ by $u(x)$ below. To compute $\partial u(x_0)/\partial p$ at the observation point x_0, we employ the adjoint field, $u^\dagger(x)$, defined as the solution to

$$\frac{d^2 u^\dagger(x)}{dx^2} + k^2 u^\dagger(x) + \gamma(x,\mathbf{p})u^\dagger(x) = \delta(x - x_0). \quad (6)$$

Now differentiate (4) with respect to p, where for brevity we write $u_p(x) \equiv \partial u(x)/\partial p$. This gives

$$\frac{d^2 u_p(x)}{dx^2} + k^2 u_p(x) + \gamma(x,\mathbf{p})u_p(x) = -\frac{\partial \gamma(x,\mathbf{p})}{\partial p}u(x). \quad (7)$$

Thus, $u_p(x)$ obeys a Helmholtz equation with the "source" given by the term on the right of (7). We wish to compute the derivative of the observed field, $u(x_0)$, with respect to p. To do this, note that

$$\frac{\partial u(x_0)}{\partial p} \equiv u_p(x_0) = \int_{-\infty}^{\infty} u_p(x)\delta(x - x_0)dx. \quad (8)$$

Substituting the left-hand side of (6) into (8) gives

$$\frac{\partial u(x_0)}{\partial p} = \int_{-\infty}^{\infty} u_p(x)\left[\frac{d^2 u^\dagger(x)}{dx^2} + k^2 u^\dagger(x) + \gamma(x,\mathbf{p})u^\dagger(x)\right]dx. \quad (9)$$

Integrating by parts twice results in

$$\frac{\partial u(x_0)}{\partial p} = \int_{-\infty}^{\infty}\left[\frac{d^2 u_p(x)}{dx^2} + k^2 u_p(x) + \gamma(x,\mathbf{p})u_p(x)\right]u^\dagger(x)dx. \quad (10)$$

Substituting (7) into (10) finally gives

$$\frac{\partial u(x_0)}{\partial p} = -\int_{-\infty}^{\infty}\frac{\partial \gamma(x,\mathbf{p})}{\partial p}u(x)u^\dagger(x)dx. \quad (11)$$

If, for example, we substitute the scattering model (5) into (11), we obtain on letting $p = p_n$, the derivative of the observed field with respect to p_n:

$$\frac{\partial u(x_0)}{\partial p_n} = -\int_{-\infty}^{\infty}\phi_n(x)u(x)u^\dagger(x)dx. \quad (12)$$

The key result (11) can also be derived by writing the solution to (7) with the aid of the Green's function, $g(x|x')$, which solves

$$\frac{d^2 g(x|x')}{dx^2} + k^2 g(x|x') + \gamma(x,\mathbf{p})g(x|x') = \delta(x - x'), \quad (13)$$

to obtain

$$u_p(x) = -\int\frac{\partial \gamma(x',\mathbf{p})}{\partial p}u(x')g(x|x')dx'. \quad (14)$$

Comparing (6) and (13), we see that $u^\dagger(x) = g(x_0|x)$, which shows that (11) and (14) are the same after setting $x = x_0$ in (14). However, we prefer to define the adjoint field, $u^\dagger(x)$, as the solution to (6), where the observation point, x_0, is defined as the location of the source of this field.

III. ITERATIVE ALGORITHMS AND THE LOCAL MINIMUM PROBLEM

In the above problem, we denoted the observations by $u(x_0, k_m, \mathbf{p}^{(true)})$, where $\mathbf{p}^{(true)}$ is the true set of parameters. For brevity, we write this as $u^{(obs)}(k_m)$. For a given parameter estimate, $\mathbf{p}$, we can simulate the observations, denoted $u(x_0, k_m, \mathbf{p})$, by using a forward algorithm to solve (4). Below we abbreviate $u(x_0, k_m, \mathbf{p})$ by $u(k_m, \mathbf{p})$. In the usual least-squares approach, we seek the solution, $\mathbf{p}$, that minimizes the mean-square error

$$\mathcal{E}(\mathbf{p}) = \frac{1}{2}\sum_{m=1}^{M}[u(k_m, \mathbf{p}) - u^{(obs)}(k_m)]^2. \quad (15)$$

Solutions occur when the gradient of $\mathcal{E}(\mathbf{p})$ vanishes, that is, when the partial derivatives of $\mathcal{E}(\mathbf{p})$ all vanish:[2]

$$\frac{\partial \mathcal{E}(\mathbf{p})}{\partial p_n} = \sum_{m=1}^{M}[u(k_m, \mathbf{p}) - u^{(obs)}(k_m)]\frac{\partial u(k_m, \mathbf{p})}{\partial p_n} = 0, \quad (16)$$

for $n = 1, 2, \ldots, N$. Convergence to a global minimum will occur from any point within some domain of $\mathbf{p}$ provided that the matrix of second derivatives of $\mathcal{E}(\mathbf{p})$ is positive definite everywhere within that domain. This matrix is

$$\frac{\partial^2 \mathcal{E}(\mathbf{p})}{\partial p_i \partial p_j} = \sum_{m=1}^{M}\left\{\frac{\partial u(k_m, \mathbf{p})}{\partial p_i}\frac{\partial u(k_m, \mathbf{p})}{\partial p_j} + [u(k_m, \mathbf{p})\right.$$
$$\left. - u^{(obs)}(k_m)]\frac{\partial^2 u(k_m, \mathbf{p})}{\partial p_i \partial p_j}\right\}. \quad (17)$$

In general, the second term on the right is responsible for local minima, if they occur. This term vanishes if $u(k_m, \mathbf{p})$ is linear in $\mathbf{p}$, leaving only the first term, resulting in a positive

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Stephen J. Norton: Nonlinear inverse scattering   1535

definite $\partial^2 \mathcal{E}(\mathbf{p})/\partial p_i \partial p_j$. In the latter case, $\mathcal{E}(\mathbf{p})$ is quadratic in $\mathbf{p}$, implying the existence of a unique global minimum.[3] An additional point can be made by writing (16) as the matrix equation

$$\sum_{m=1}^{M} e_m(\mathbf{p}) J_{mn}(\mathbf{p}) = 0, \tag{18}$$

where $e_m(\mathbf{p}) \equiv u(k_m, \mathbf{p}) - u^{(obs)}(k_m)$ and $J_{mn}(\mathbf{p})$ is the Jacobian matrix

$$J_{mn}(\mathbf{p}) = \frac{\partial u(k_m, \mathbf{p})}{\partial p_n}. \tag{19}$$

Generally speaking a global minimum will occur at a point $\mathbf{p}$ when (18) holds and the Jacobian, $J_{mn}(\mathbf{p})$, is full rank at $\mathbf{p}$. This then guarantees that $e_m(\mathbf{p}) = 0$ for all m at the point $\mathbf{p}$. A local minimum can occur at a point $\mathbf{p}$ whenever the gradient vanishes [i.e., when (18) holds at $\mathbf{p}$] and when $J_{mn}(\mathbf{p})$ is less than full rank at $\mathbf{p}$. Then (18) can hold while $e_m \neq 0$ for one or more m, signifying a local minimum at $\mathbf{p}$.

We now consider an approach that, in principle, avoids local minima, which is based on solving the system of M nonlinear equations

$$u(k_m, \mathbf{p}) = u^{(obs)}(k_m), \quad m = 1, 2, \ldots, M, \tag{20}$$

for the unknown vector $\mathbf{p}$. If a solution to (20) exists, the following approach should, in principle, avoid converging to a local minimum in the error surface $\mathcal{E}(\mathbf{p})$. We solve (20) using Newton's method, as follows. Newton's method also has the added theoretical benefit of second-order convergence, whereas gradient descent using the mean-square error gradient (16) provides only first-order convergence.

Denote the jth estimate of $\mathbf{p}$ by $\mathbf{p}^{(j)}$. We then Taylor expand $u(k_m, \mathbf{p}^{(j+1)})$ about the point $\mathbf{p}^{(j)}$. We then have, to first order in $\mathbf{p}^{(j+1)} - \mathbf{p}^{(j)}$,

$$u(k_m, \mathbf{p}^{(j+1)}) = u(k_m, \mathbf{p}^{(j)}) + \sum_{n=1}^{N} J_{mn}(p_n^{(j+1)} - p_n^{(j)}), \tag{21}$$

where J_{mn} is the Jacobian matrix which, in view of (11), then can be computed from

$$J_{mn} \equiv \frac{\partial u(k_m, \mathbf{p}^{(j)})}{\partial p_n} = -\int_{-\infty}^{\infty} \frac{\partial \gamma(x, \mathbf{p}^{(j)})}{\partial p_n} u(x, k_m) u^{\dagger}(x, k_m) dx. \tag{22}$$

As noted, J_{mn} is evaluated at $\mathbf{p}^{(j)}$. Now let $\mathbf{J}$ denote the matrix whose elements are J_{mn} and $\mathbf{u}(\mathbf{p})$ an M component vector whose elements are $u(k_m, \mathbf{p})$, $m = 1, 2, \ldots, M$. Newton's method is obtained by setting $\mathbf{u}(\mathbf{p}^{(j+1)}) = \mathbf{u}^{(obs)}$ in (21), which gives an equation for $\mathbf{p}^{(j+1)}$ in terms of $\mathbf{p}^{(j)}$:

$$\mathbf{u}^{(obs)} - \mathbf{u}(\mathbf{p}^{(j)}) = \mathbf{J}(\mathbf{p}^{(j+1)} - \mathbf{p}^{(j)}). \tag{23}$$

This linear system can be solved for $\mathbf{p}^{(j+1)}$, or, symbolically,

$$\mathbf{p}^{(j+1)} = \mathbf{p}^{(j)} - \mathbf{J}^{-1} [\mathbf{u}(\mathbf{p}^{(j)}) - \mathbf{u}^{(obs)}], \tag{24}$$

which is Newton's method in N variables. If there are more measurements, M, than unknowns, N, the system (23) is overdetermined, in which case, for example, the generalized inverse can be used to solve (23). To this end, the system

(23) is premultiplied by the transpose of $\mathbf{J}$, giving

$$\mathbf{J}^T [\mathbf{u}^{(obs)} - \mathbf{u}(\mathbf{p}^{(j)})] = \mathbf{J}^T \mathbf{J}(\mathbf{p}^{(j+1)} - \mathbf{p}^{(j)}), \tag{25}$$

and the update formula (24) now becomes

$$\mathbf{p}^{(j+1)} = \mathbf{p}^{(j)} - (\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T [\mathbf{u}(\mathbf{p}^{(j)}) - \mathbf{u}^{(obs)}], \tag{26}$$

where $(\mathbf{J}^T \mathbf{J})^{-1} \mathbf{J}^T$ is the generalized inverse.

Press *et al.*[4] point out that the Newton step defined by (24) is always a descent direction on the error surface $\mathcal{E}(\mathbf{p}) = [\mathbf{u}(\mathbf{p}) - \mathbf{u}^{(obs)}]^2/2$. To show this, abbreviate the step as $\delta \mathbf{p} = \mathbf{p}^{(j+1)} - \mathbf{p}^{(j)}$; then

$$\nabla \mathcal{E}^T \delta \mathbf{p} < 0, \tag{27}$$

which follows by noting that $\nabla \mathcal{E}^T = (\mathbf{u}(\mathbf{p}^{(j)}) - \mathbf{u}^{(obs)})^T \mathbf{J}$ and then from (24)

$$\nabla \mathcal{E}^T \delta \mathbf{p} = -|\mathbf{u}(\mathbf{p}^{(j)}) - \mathbf{u}^{(obs)}|^2 < 0. \tag{28}$$

The same argument can be used to show that the step implied by (26) is also a descent direction. Press *et al.*[4] use the fact that each Newton step is a descent direction to devise a strategy for adaptively modifying the step size if $\mathcal{E}(\mathbf{p})$ is not reduced at each iteration. In their approach, one backtracks along the descent direction if the current step fails to reduce the error $\mathcal{E}(\mathbf{p})$. Note, however, that the descent direction defined by (24) [or (26)] is not the same descent direction defined by the gradient of $\mathcal{E}(\mathbf{p})$ given by (16) in solving the least-squares problem. In the Appendix, we outline a modified Newton's method based on the idea of continuation.[5] In that approach, the Newton iteration is reformulated as a differential equation in $\mathbf{p}$ whose solution progresses continuously from a known solution, $\mathbf{p}_0$, to the desired solution. The idea here is to enlarge the domain of starting values from which global convergence succeeds.

IV. EXAMPLE: TWO-PARAMETER INVERSE PROBLEM

As an illustration, we consider a simple inverse Helmholtz problem with two unknown parameters. Although the problem is simple and artificial, it does succeed in showing how the Newton iteration converges to the global minimum, whereas mean-square minimization converges, from the same starting value, to a local minimum.

We consider a one-dimensional problem with a potential step of unknown magnitude and unknown location. The step occurs at the point $x = a$ and beyond the step the wave number is β. We wish to determine the parameters a and β from two measurements. The field $u(x)$ then obeys the following Helmholtz equation:

$$\frac{d^2 u(x)}{dx^2} + k^2 u(x) + (\beta^2 - k^2) H(x - a) u(x) = 0, \tag{29}$$

where k is the wave number to the left of the step and $H(x)$ is the unit step function. We assume a field $\exp(ikx)$ is incident on the step from the left. The total field is then of the form

$$u(x) = e^{ikx} + R e^{-ikx}, \quad x < a, \tag{30}$$

$$u(x) = Te^{i\beta x}, \quad x > a. \tag{31}$$

Imposing the boundary conditions of continuity of the field and its derivative at $x = a$ yields the coefficients R and T:

$$R = \frac{k - \beta}{k + \beta} e^{2ika}, \tag{32}$$

$$T = \frac{2k}{k + \beta} e^{i(k-\beta)a}. \tag{33}$$

The potential function $\gamma(x, \mathbf{p})$ for this problem is

$$\gamma(x, \mathbf{p}) = (\beta^2 - k^2) H(x - a) \tag{34}$$

and its derivatives with respect to a and β are

$$\frac{\partial \gamma(x, \mathbf{p})}{\partial a} = -(\beta^2 - k^2) \delta(x - a), \tag{35}$$

$$\frac{\partial \gamma(x, \mathbf{p})}{\partial \beta} = 2\beta H(x - a). \tag{36}$$

To compute the derivatives of the observed field at an observation point x_0 (where we assume $x_0 < a$) using (11), we need the adjoint field which, in this case, solves

$$\frac{d^2 u^\dagger(x)}{dx^2} + k^2 u^\dagger(x) + (\beta^2 - k^2) H(x - a) u^\dagger(x) = \delta(x - x_0). \tag{37}$$

Solving this equation, we find that the adjoint solution is given by

$$u^\dagger(x) = -\frac{i}{2k} u(x) e^{-ikx_0}, \tag{38}$$

for $x \geq x_0$, where $u(x)$ is given above in (30) and (31). Substituting $u(x)$ and $u^\dagger(x)$ into (11), we obtain for the derivatives of the observed field

$$\frac{\partial u(x_0)}{\partial a} = (\beta^2 - k^2) u(a) u^\dagger(a), \tag{39}$$

$$\frac{\partial u(x_0)}{\partial \beta} = -2\beta \int u(x) u^\dagger(x) dx. \tag{40}$$

Integrating (40), we finally obtain

$$\frac{\partial u(x_0)}{\partial a} = 2ik \left[\frac{k - \beta}{k + \beta} \right] e^{i2ka} e^{-ikx_0}, \tag{41}$$

$$\frac{\partial u(x_0)}{\partial \beta} = -\frac{2k}{(k + \beta)^2} e^{i2ka} e^{-ikx_0}. \tag{42}$$

As a check, these can be obtained by directly differentiating (30) with respect to β and a after setting $x = x_0$.

V. SIMULATION

Using the example of the last section, we devise a problem with both local and global minima. Starting from the same initial guess, $\mathbf{p}$, we show that the Newton iteration converges to the global minimum, whereas a steepest-
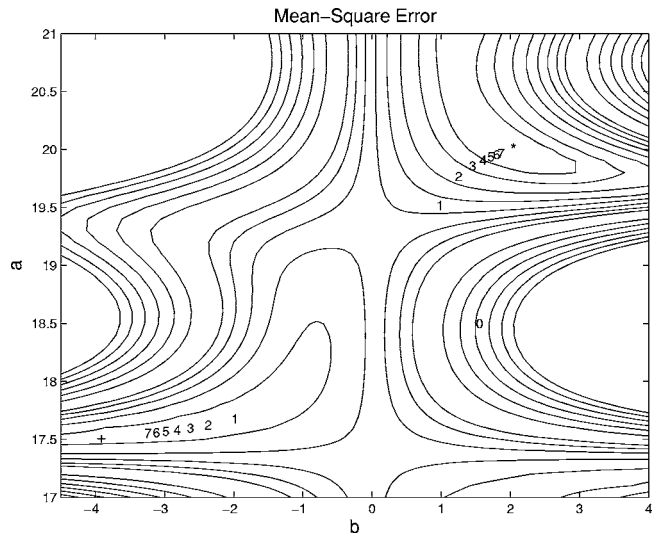
FIG. 1. Contour plot of mean-square error surface. The $*$ and $+$ signify, respectively, global and local minima, with the starting value for both algorithms identified by the 0. Newton's method converges to the global minimum and the method of steepest descent converges to one of many local minima. For Newton's method, the numbers label single iterations, whereas for steepest descent, each number stands for 1000 iterations.

descent algorithm that minimizes the mean-square error converges to a local minimum. The data consist of observations made at two frequencies (i.e., two values of k). In particular, the backscattered field is

$$u(x, k) = e^{ikx} + R e^{-ikx}, \tag{43}$$

where from (32) $R = [(k - \beta)/(k + \beta)]\exp(2ika)$. Recall that a is the location of a step discontinuity and β is the wave number beyond this step. Let us define as our unknowns a and $b \equiv (k - \beta)/(k + \beta)$ and let our data consist of the real part of the backscattered field observed at x_0 with wave numbers k_1 and k_2; that is, we measure

$$u_R(x_0, k_1) = \cos(k_1 x_0) + b \cos(2k_1 a - k_1 x_0),$$

$$u_R(x_0, k_2) = \cos(k_2 x_0) + b \cos(2k_2 a - k_2 x_0),$$

with a and b the unknowns.

This particular problem has multiple minima. We set the true parameter values to $a = 20$ and $b = 1$, and let $x_0 = 0$, $k_1 = 0.7$, and $k_2 = 0.1$. In an attempt to recover the parameters a and b, we applied both iterative algorithms, steepest descent and Newton's method, using the same initial starting point: $a = 18.5$ and $b = 1.5$. Figure 1 shows the paths of convergence of both algorithms on a contour plot of the mean-square error surface $\mathcal{E}(a, b)$, where the global minimum is marked by $*$ and a local minimum is marked by $+$. The starting value is identified by 0. For this particular example, Newton's method converges to the global minimum and the steepest-descent method converges to the local minimum. One can check that the Jacobian is full rank at the global minimum, but singular at the local minimum. The numbers in the figure denote single iterations of Newton's method (converging to

∗), where each number in the steepest-descent path (converging to +) stands for 1000 iterations. Thus, we see that steepest descent converges exceedingly slowly in this example. A more efficient descent algorithm, such as the conjugate-gradient method, would converge more rapidly, although once again to the local minimum.

VI. DISCUSSION AND CONCLUSION

We propose a method for solving nonlinear inverse-scattering problems that, in principle, should avoid the problem of convergence to local minima. In the simple example described in the last section, the author found that the basic Newton iteration, without the modification suggested by Press *et al.*,[4] either converged to the global minimum or failed to converge, depending on the starting values of the parameters. In the limited number of tests conducted, the algorithm never converged to a local minimum. On the other hand, the steepest-descent algorithm either failed to converge, converged to a local minimum, or occasionally converged to the global minimum, again depending on the starting parameters. To develop a more robust Newton algorithm, the modification suggested by Press *et al.* or a similar strategy might be used, particularly for higher dimensional problems. If no exact solution to (20) exists due to noise in the data or inaccuracies in the model, an algorithm needs to be devised to yield an approximate solution under these conditions. For example, a stabilized Newton iteration along the lines of the generalized update formula (26) should converge to a "reasonable" solution given uncertainties in the data or model. The latter formula is less likely to fail due to an ill-conditioned Jacobian. Another potentially promising approach is the continuation method outlined in the Appendix.

The purpose of this paper is to suggest a strategy for avoiding false solutions associated with local minima. However, the numerical issues of nonconvergence and stability associated with a multi-dimensional Newton's method are beyond our scope here. These issues will require further study for the development of a truly robust inverse-scattering algorithm.

APPENDIX: "MORPHING" A KNOWN SOLUTION INTO AN UNKNOWN SOLUTION

Ortega and Rheinboldt[5] describe a technique known as the method of continuation, in which the Newton iteration starts from the solution of a known problem and is allowed to continuously evolve toward the solution of the desired problem. The idea is both to increase the domain of starting values that leads to convergence to the global solution and to provide a more robust algorithm than the basic Newton iteration. Here we outline this approach and refer the reader to Ref. 5 for details and a discussion of convergence criteria. In view of (20), we can redefine our original problem as the task of finding the zeros, $\mathbf{p}$, of a system of equations $F(k_m, \mathbf{p}) = 0$, given by

$$F(k_m, \mathbf{p}) \equiv u(k_m, \mathbf{p}) - u^{(obs)}(k_m) = 0, \quad m = 1, 2, \ldots, M, \tag{A1}$$

in which a standard multi-variable Newton's method would be used. Letting $\mathbf{F}(\mathbf{p})$, $\mathbf{u}(\mathbf{p})$, and $\mathbf{u}^{(obs)}$ denote vectors of M components, this can be written

$$\mathbf{F}(\mathbf{p}) \equiv \mathbf{u}(\mathbf{p}) - \mathbf{u}^{(obs)} = 0. \tag{A2}$$

Suppose we consider instead the problem of finding the zeros of

$$\mathbf{F}(\mathbf{p}, \tau) \equiv \mathbf{u}(\mathbf{p}) - \mathbf{u}^{(obs)} + (\tau - 1)[\mathbf{u}(\mathbf{p}_0) - \mathbf{u}^{(obs)}] = 0, \tag{A3}$$

where τ varies between 0 and 1 and $\mathbf{p}_0$ is a set of starting parameters. The idea is to start the iteration at a known solution defined when $\tau = 0$, that is, $\mathbf{F}(\mathbf{p}_0, 0) = 0$ [which follows from (A3)] and end at the desired solution when $\tau = 1$, that is, $\mathbf{F}(\mathbf{p}, 1) = 0$ [which is (A2)]. When τ varies from 0 to 1, we see that the solution to (A3) varies continuously from $\mathbf{p}_0$ to $\mathbf{p}_1$, where $\mathbf{p}_1$ is the solution to (A2), that is, the desired solution. Note that $\mathbf{p}$ is a function of τ, which we signify by writing $\mathbf{p}_\tau$. An iterative algorithm can be devised (as τ varies from 0 to 1) by differentiating (A3) with respect to τ. This gives

$$\mathbf{J}(\mathbf{p}_\tau) \frac{d\mathbf{p}_\tau}{d\tau} + \mathbf{u}(\mathbf{p}_0) - \mathbf{u}^{(obs)} = 0, \tag{A4}$$

where $\mathbf{J}$ is the Jacobian defined by (22), which leads to the differential equation

$$\frac{d\mathbf{p}_\tau}{d\tau} = -\mathbf{J}(\mathbf{p}_\tau)^{-1}[\mathbf{u}(\mathbf{p}_0) - \mathbf{u}^{(obs)}], \tag{A5}$$

with the initial condition $\mathbf{p}_0$. If we discretize τ by writing τ_k, where $k = 0, 1, \ldots, N$, (A5) can be approximated using, for example, the simple integration scheme of Euler. Letting $\mathbf{p}_k$ denote $\mathbf{p}_{\tau_k}$, this results in

$$\mathbf{p}_{k+1} = \mathbf{p}_k - (\tau_{k+1} - \tau_k)\mathbf{J}(\mathbf{p}_k)^{-1}[\mathbf{u}(\mathbf{p}_0) - \mathbf{u}^{(obs)}]. \tag{A6}$$

Note the resemblance to Newton's method as defined by (24). The difference, however, is that the quantity in square brackets in (A6) is not updated at each iteration, as is done in (24), as well as the presence of the weighting $(\tau_{k+1} - \tau_k)$. As before, each iteration requires a solution to the forward and adjoint scattering problems to compute the Jacobian. Since the above integration scheme may require more iterations than Newton's method, we may sacrifice computational efficiency for the benefit of widening the domain of convergence. Note also from (24) that Newton's method will fail at any point where the determinant of the Jacobian vanishes. However, a more sophisticated integration scheme than Euler's method applied to (A5) should be able to pass through these singular points, provided the singularities are integrable (alternatively the integration path $\mathbf{p}_\tau$, $0 \leq \tau \leq 1$, through parameter space can be deformed around these points). The above arguments suggest that a method based on integrating (A5) using an adaptive integration algorithm, although computationally more costly, may well be more stable than a simple Newton's iteration, but a study of this approach is beyond our scope.

[1]S. J. Norton, "Iterative inverse-scattering algorithms: Methods for comput-ing the Frechet derivative," J. Acoust. Soc. Am. **106**, 2653–2660 (1999).
[2]If we define $F_n(\mathbf{p}) = \partial \mathcal{E}(\mathbf{p})/\partial p_n$, local minima will occur for zeros of the system of nonlinear equations

$$F_n(\mathbf{p}) = 0, \quad n = 1, 2, \dots, N.$$

However, as pointed out in Ref. 4, these equations are the components of the gradient of a scalar and are not independent. In fact, they obey the $N(N-1)/2$ integrability conditions

$$\frac{\partial F_n(\mathbf{p})}{\partial p_m} = \frac{\partial F_m(\mathbf{p})}{\partial p_n},$$

which follows from the definition of $F_n(\mathbf{p})$. For N equations in N unknowns, these conditions imply, generally speaking, the possibility of multiple solu-tions to the first equation in this footnote and thus multiple local minima.
[3]P. H. Gill, W. Murray, and M. H. Wright, *Practical Optimization* (Aca-demic, New York, 1981).
[4]W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery, *Nu-merical Recipes in Fortran*, 2nd ed. (Cambridge U.P., New York, 1992), p. 375.
[5]J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equa-tions in Several Variables* (Academic, New York, 1970), pp. 230–235.

Analysis of models for the synapse between the inner hair cell and the auditory nerve

Xuedong Zhang
Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Laurel H. Carney[a]
Departments of Biomedical and Chemical Engineering and Electrical Engineering and Computer Science, Institute for Sensory Research, Syracuse University, Syracuse, New York 13244

A general mathematical approach was proposed to study phenomenological models of the inner-hair-cell and auditory-nerve (AN) synapse complex. Two models (Meddis, 1986; Westerman and Smith, 1988) were studied using this unified approach. The responses of both models to a constant-intensity stimulus were described mathematically, and the relationship between model parameters and response characteristics was investigated. The mathematical descriptions of the two models were essentially equivalent despite their structural differences. This analytical approach was used to study the effects of adaptation characteristics on model parameters and of model parameters on adaptation characteristics. The results provided insights into these models and the underlying biophysical processing. This analytical method was also used to study offset adaptation, and it was found that the offset adaptation of both models was limited by the models' structures. A modified version of the synapse model, which has the same onset adaptation but improved offset adaptation, is proposed here. This modified synapse model produces more physiologically realistic offset adaptation and also enhances the modulation gain of model AN fiber responses, consistent with AN physiology. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1993148]

I. INTRODUCTION

The inner-hair-cell (IHC) and auditory-nerve (AN) complex is a critical element in the peripheral auditory system that converts a mechanical signal (the response of the cochlea to sound) into a neural signal. Physiological studies (Westerman and Smith, 1984; Rhode and Smith, 1985; Westerman, 1985) have provided insight into the temporal dynamics of IHC-AN synaptic processing. The response of an AN fiber to a constant-intensity tone burst is typified by very rapid firing at the onset that declines with time, rapidly at first and then more slowly over a period of tens of milliseconds. The AN response after the offset of the stimulus (cessation of the stimulus) is greatly reduced relative to the spontaneous response and slowly recovers over several tens of milliseconds (Harris and Dallos, 1979; Westerman, 1985). Since the amplitude of the receptor potential of the IHC produced by a constant-amplitude stimulus (for example, a high-frequency tone) is essentially constant (Russell and Sellick, 1978), adaptation is believed to occur at the level of the transmitter release process in the IHC-AN synapse. Recent physiological studies (Moser and Beutner, 2000) provide direct evidence for the idea of adaptation at the IHC-AN synapse. The adaptation process in the IHC-AN synapse results in greater sensitivity to transient stimuli than to steady-state stimuli and underlies other temporal aspects of AN responses. These properties also limit some aspects of temporal coding of stimulus envelopes. For example, offset adaptation may be responsible in part for the psychophysical phenomenon of forward masking (Harris and Dallos, 1979).

IHC-AN synaptic adaptation is very complex. Its characteristics depend on stimulus intensity, duration, and previous stimulation history. Synaptic adaptation (Fig. 1) (Harris and Dallos, 1979) at the onset of AN responses to tone bursts is usually characterized functionally by two exponential components (Westerman and Smith, 1984; Rhode and Smith, 1985; Westerman, 1985):

$$R_{on}(t) = A_{sus} + A_r e^{-t/\tau_R} + A_{st} e^{-t/\tau_{ST}}, \qquad (1)$$

where A_r and A_{st} are the components of rapid and short-term adaptation, τ_R and τ_{ST} are the respective decay time constants, and A_{sus} is a steady-state component. Adaptation at the offset of the stimulus can also be described as an exponential recovery component, with a different time constant than the onset, after a "deadtime" period, or a period of very low probability of firing:

$$R_{off}(t) = \begin{cases} 0; & t < t_0 \\ A_{sp}(1 - e^{-(t-t_0)/\tau_{sp}}) \end{cases}, \qquad (2)$$

where A_{sp} is the spontaneous rate, t_0 is the deadtime period, and τ_{SP} is the recovery time constant of the offset responses. Other measurements that have helped characterize AN adaptation processes include responses to increments or decrements in stimulus intensity with an ongoing tonal background (Smith and Zwislocki, 1975;

---

[a]Author to whom correspondence should be addressed; electronic mail: lacarney@syr.edu
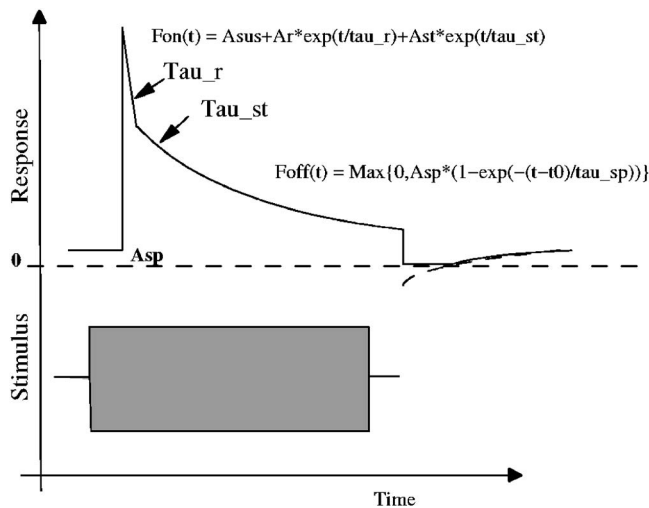
FIG. 1. Schematic diagram of the response of an AN fiber to a toneburst (Harris and Dallos, 1979; Westerman, 1985). The onset response of the AN fiber can be described as a sum of two exponential components [see Eq. (1)], and the offset response can be fit with a single exponential recovery function after a deadtime [Eq. (2)].

FIG. 2. Schematic diagram of two synapse models (Meddis, 1988; Westerman and Smith, 1988). (a) In the Westerman model [see Eqs. (6) and (7)], three transmitter reservoirs are cascaded, each with their own concentration and volume. The transmitter diffuses from global to local and then from local to immediate reservoirs. The output of the synapse model, S_{out}, is determined by the product of the concentration in the immediate reservoir and the diffusion permeability from the immediate reservoir into the synaptic cleft. (b) In the Meddis model [Eqs. (3)–(5)], the transmitter in the cleft is either lost or reprocessed by the reprocessing store (w). The free transmitter pool receives the transmitter from the reprocessing store or from a global factory and releases the transmitter into the cleft (c) with instantaneous release probability density k. The model synapse output is proportional to the amount of transmitter left in cleft (Meddis, 1988) or the rate of transmitter released into the cleft (Sumner et al., 2002).

Smith et al., 1985). Adaptation characteristics also differ for fiber types with different spontaneous rates (Rhode and Smith, 1985). The diversity and complexity of IHC-AN synaptic adaptation provide a challenge for the successful modeling of synapse dynamics (Hewitt and Meddis, 1991).

The mechanisms that give rise to synaptic adaptation can be due either to the depletion of a readily released presynaptic pool of neurotransmitter (Moser and Beutner, 2000) or the desensitization of the postsynaptic receptors (Raman et al., 1994). The adaptation at IHC-AN synapse has often been modeled using multiple reservoirs of a neurotransmitter, with diffusion out of the cell and between reservoirs within the cell (Furukawa and Matsuura, 1978; Schwid and Geisler, 1982; Smith and Brachman, 1982; Meddis, 1986; Westerman and Smith, 1988; Ross, 1996). Each diffusion step is controlled by a permeability parameter, and at least one of the permeabilities in these models is determined by the stimulus (presumed to be controlled by IHC calcium concentration, intracellular IHC voltage, or equivalently, stimulus intensity). Adaptation of the synaptic output in these models depends on the reduction of the driving force for the diffusion of the synaptic material (transmitter) from the cell into the synaptic cleft (Fig. 2). Mathematically, low-pass filters are used to implement the replenishment and diffusion mechanisms between different transmitter reservoirs. These models can be implemented using either a cascade of low-pass filters or parallel low-pass filters, depending on the interconnection of the reservoirs.

The variation of adaptation characteristics across AN fibers makes it desirable to have different sets of model parameters to predict individual AN fiber responses more accurately, but determining these parameters is tedious. Compared to analytical solutions, the exhaustive search of a multidimensional parameter space using numerical methods is very time consuming, and the parameters may be trapped by local minima in the cost function. In addition, simulation

results may not give insight into a particular model; that is, quantitative relationships between the model parameters and adaptation properties may not be revealed by the exercise of fitting the parameters to a single fiber response, and the difference between different models may not be fully understood.

During the past twenty years, two IHC-AN synapse models with different structures have been developed independently and actively in a series of studies (Meddis, 1986, 1988; Westerman and Smith, 1988; Meddis et al., 1990; Carney, 1993; Zhang et al., 2001; Sumner et al., 2002). Both of these models are very successful in simulating the onset adaptation responses of the AN fibers to some extent. For the model proposed by Westerman and Smith (1988), equations were derived to determine the model parameters from the desired shape of the onset of the peri-stimulus time (PST) histogram. A recent study (Zhang et al., 2001) detailed how these parameters can be derived for different types of AN fibers. For the model proposed by Meddis (1986), the analytical relationship between steady-state responses and model parameters was described in a later study (Meddis et

al., 1990). This model was further explored in a recent study (Sumner *et al.*, 2002) to simulate the rate-level functions of low-, medium-, and high-spontaneous-rate AN fibers by changing parameters in the transmitter-release control function (that is, the stage that determines how the model IHC voltage affects the transmitter release permeability in the adaptation model) and synapse model parameters. However, the studies of other model adaptation characteristics (e.g., amplitudes and time constants of adaptation components) are limited by the sets of parameters provided by the authors (e.g., Hewitt and Meddis, 1991; Sumner *et al.*, 2002), which makes it difficult to understand the relationship between these parameters and different adaptation properties. Also, the difference in adaptation responses for these models, and whether the difference is a result of the different model structures or the different model parameter values, have not been clarified.

In this article, we use a unified approach to consider two models, that of Meddis (1986) and of Westerman and Smith (1988) (we will refer to the latter as the Westerman model). Since the essential architecture of the synapse model under investigation (interconnection of the synaptic reservoirs) has not been changed in the newer versions of these models (Zhang *et al.*, 2001; Sumner *et al.*, 2002), the results from this study are also applicable and beneficial to understanding the more recent versions of these models (see Sec. VI). In our approach, both models can be described as circuits of interconnected low-pass filters. In this framework, the responses of each model to a constant-intensity stimulus can be determined analytically, and relationships between the response characteristics and model parameters can be established. This mathematical method clarifies the differences and similarities between the two models, leading us to the conclusion that, despite their different structures, the two models are essentially the same. In addition, because the model equations provide better insight into the model structure, we can easily derive the model parameters on the basis of the desired adaptation properties.

There has been less attention paid to the modeling of offset adaptation at the IHC-AN synapse. It is difficult to find a common physiological explanation for both onset and offset adaptation because of the difference in their respective characteristic functions. Simulation results from models (Sumner *et al.*, 2003) have also not provided insights concerning the relationship between onset and offset adaptation.

We use our analytical approach to study the relationship between onset and offset adaptation in the Westerman and Meddis models. The model response characteristic function of offset adaptation is the same as that of onset adaptation, and is limited by the model parameters used to specify the desired onset adaptation. A modification for both models is then proposed; the modified model has the same onset adaptation but improved offset adaptation. The onset and offset adaptation responses are based on same physiological assumption in this modification. The modified offset adaptation enhances the modulation gain of the model AN fiber responses to modulated stimuli, consistent with AN physiology (Joris and Yin, 1992).

II. PHYSIOLOGICAL INTERPRETATION OF THE MEDDIS AND WESTERMAN SYNAPSE MODELS

In this section, we provide our physiological interpretation of the two models and explain how onset adaptation is achieved in both models. The similarities and differences between the two models are compared qualitatively here; their underlying physiological assumptions are further discussed in Sec. VI.

The model proposed by Meddis (1986, 1988) has three neurotransmitter reservoirs that are specified by the number of synaptic transmitters in the reservoir and arranged in a cycle: The immediate store (q); the synaptic cleft (c); and the reprocessing store (w) (see Fig. 2). New transmitter in the immediate store is either manufactured from the factory at a rate of $y[M-q(t)]$, where M is the amount of neurotransmitter in the global store), provided by a reprocessing store [at a rate of $xw(t)$], or released from the immediate store into the synaptic cleft at a rate of $k(t)q(t)$. The transmitter in the cleft is either lost at a rate of $lc(t)$ or recycled at a rate of $rc(t)$. The transmitter replenishment and release in the model can be described by the following equations:

$$\frac{dq}{dt} = y(M - q(t)) + xw(t) - k(t)q(t), \tag{3}$$

$$\frac{dc}{dt} = k(t)q(t) - (l + r)c(t), \tag{4}$$

$$\frac{dw}{dt} = rc(t) - xw(t). \tag{5}$$

The model synapse output is a value proportional to the rate of transmitter release into the cleft, $k(t)q(t)$. The instantaneous release probability, $k(t)$, is the only stimulus-dependent variable in the model and is usually specified as a function of intracellular IHC voltage (defined as V_{IHC}, referenced to the IHC resting potential). Rapid adaptation after onset is the result of the depletion of the synaptic transmitters in the immediate store, and is mainly limited by the instantaneous release probability, $k(t)$. Short-term adaptation is the result of the depletion of synaptic materials available for recycling process, and is mainly limited by the reuptake rate [$rc(t)$], reprocessing rate [$xw(t)$], and the fraction of synaptic material in the cleft available for recycling, $r/(1+r)$. The replenishment of the synaptic transmitter from the global factory into the immediate store, $y[M-q(t)]$, contributes the model's steady-state response.

In the Westerman model, the transmitter reservoir is characterized by the volume, V, and concentration, C, of the synaptic material in the reservoir. The model (Fig. 2) has three reservoirs (global, local, and immediate) of synaptic material, connected in series by a diffusion path. The diffusion rate of synaptic material between reservoirs (from global to local, and then from local to immediate) is determined by the permeability (P) and by the concentration difference between the reservoirs. The governing equations for the transmitter release between these reservoirs are

$$V_I \frac{dC_I(t)}{dt} = -P_L C_I(t) - P_I(t)C_I(t) + P_L C_L(t), \tag{6}$$

and

$$\frac{dC_L(t)}{dt} = \frac{P_G C_G}{V_L} + \frac{P_L}{V_L}C_I(t) - \left(\frac{P_L + P_G}{V_L}\right)C_L(t), \tag{7}$$

where C_I, C_L, and C_G are the immediate, local, and global reservoir concentrations; V_I and V_L are the immediate and local reservoir volumes; and P_L and P_G are the permeabilities between the local and immediate and between global and local reservoirs, respectively.

The small immediate store (V_I, C_I) in the Westerman model is directly comparable with that in the Meddis model (q). The release permeability $P_I(t)$ is the same as $k(t)$ in the Meddis model, which is the only parameter determined by the input sound stimulus. The synapse output is a value proportional to the rate of transmitter released from the immediate store $[P_I(t)C_I(t)]$ into the synaptic cleft, similar to $k(t)q(t)$ in Meddis model. The Westerman model has the following characteristic function:

$$C_I(t) = \Phi_0' + \Phi_1' e^{-t/\tau_1'} + \Phi_2' e^{-t/\tau_2'}, \tag{8}$$

where the values of τ_1', τ_2', and Φ_0', Φ_1', Φ_2' can be calculated from Eqs. (6) and (7) (see Westerman and Smith, 1988). The primes in the parameter names distinguish these variables from those used in original Meddis model [see Eq. (13) below]. Note that the characteristic function here is the same as that of the onset responses of AN fibers [Eq. (1)].

Rapid adaptation in the Westerman model is the result of the depletion of the synaptic material in the immediate store and is mainly limited by the diffusion permeability of synaptic material from the immediate store to the cleft. Short-term adaptation is the result of the depletion of the synaptic material in the local store, and is mainly limited by the diffusion permeability between the local and immediate stores.

The local store (V_L, C_L) in the Westerman model is similar to the reprocessing store (w) in the Meddis model, which provides synaptic material for short-term adaptation. The synaptic material (w) in the Meddis model comes from very fast reuptake of released synaptic vesicles in the cleft, whereas the synaptic material in the Westerman model for rapid adaptation is stored in local store (V_L, C_L). The steady-state response in both models is attributed to the existence of the global store. In the Meddis model, the synaptic material from the global store is released directly into the immediate store (in parallel with the reprocessing store, w), whereas in the Westerman model, synaptic material enters the immediate store through the local store (which is cascaded with the immediate store).

III. MATHEMATICAL ANALYSIS OF THE MEDDIS AND WESTERMAN MODELS

The instantaneous release probability of a vesicle from the immediate store, $k(t)$, in the Meddis model [or $P_I(t)$ in the Westerman model] is usually a function of intracellular IHC voltage (V_{IHC}), which in turn is determined by the input sound stimulus. For a high-frequency tone burst, the IHC receptor potential V_{IHC} is dominated by the "dc" component (Russell and Sellick, 1978; Cheatham and Dallos, 1993), and it is reasonable to assume that $k(t)$ is a constant after the onset (denoted as k_2; the value of $k(t)$ before the onset is denoted as k_1). For further analysis, Eqs. (3)–(5) can be transformed into the Laplace (complex frequency) domain as follows, for $t > 0$:

$$sQ(s) - q(0^-) = yM/s - yQ(s) + xW(s) - k_2 Q(s), \tag{9}$$

$$sC(s) - c(0^-) = k_2 Q(s) - (l+r)C(s), \tag{10}$$

$$sW(s) - w(0^-) = rC(s) - xW(s), \tag{11}$$

where $q(0^-)$, $c(0^-)$, and $w(0^-)$ are the reservoir concentrations before the onset.

After solving for $Q(s)$,

$$Q(s) = \frac{(sq(0^-) + yM)(s+x)(s+l+r) + c(0^-)rxs + w(0^-)xs(s+l+r)}{s(s+x)(s+y+k_2)(s+l+r) - k_2 rxs}, \tag{12}$$

the characteristic function of $q(t)$ can be represented by

$$q(t) = \Phi_0 + \Phi_1 e^{-t/\tau_1} + \Phi_2 e^{-t/\tau_2} + \Phi_3 e^{-t/\tau_3}, \tag{13}$$

where $-1/\tau_i$ are poles of $Q(s)$. The values of τ_1, τ_2, τ_3, and Φ_0, Φ_1, Φ_2, Φ_3 can be calculated from $Q(s)$ directly.

Based on the parameters given in Meddis (1986, 1988) and Sumner et al. (2002), we can assume that l and $r \to \infty$ (or $l, r \gg x, y, k$), and we will let $u = r/(l+r)$. Thus, the following equations describe a simplified version of the Meddis model:

$$\frac{dq}{dt} = y(M - q(t)) - kq(t) + xw(t), \tag{14}$$

$$\frac{dw}{dt} = kuq(t) - xw(t), \tag{15}$$

$$C(s) = \frac{k_2 Q(s) + c(0^-)}{s+l+r}, \tag{16}$$
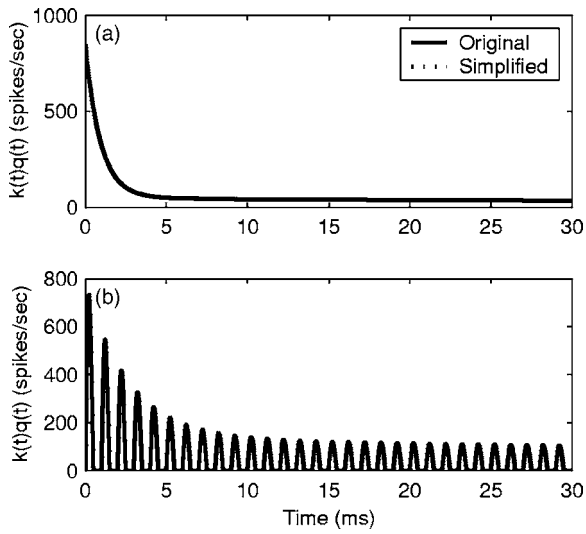
FIG. 3. Comparison of synapse output for the original Meddis model (solid line) and the simplified model (dotted line) described in the current study. Both models used the parameter values given in the Meddis study (1988): $l=2500$, $r=6580$, $x=66.31$, $M=1$, $y=5.05$. The responses from two models are essentially identical, as illustrated by the overlapping curves. (a) The analytical solution of $k(t)q(t)$ for both models for a constant stimulus, with $k_1=40.49$ and $k_2=1660$. (b) Simulation results of $k(t)q(t)$ for both models with a time-varying input $k(t)$ (a half-wave rectified sinusoidal wave form at 1 kHz with amplitude of k_2: $k(t)=0.5k_2[\sin(2000\,\pi t)+|\sin(2000\,\pi t)|]$).

$$Q(s) = \frac{(sq(0^-) + yM)(s + x) + w(0^-)xs}{s(s + x)(s + y + k_2) - k_2uxs}, \quad (17)$$

$$q(t) = \Phi_0' + \Phi_1'e^{-t/\tau_1'} + \Phi_2'e^{-t/\tau_2'}, \quad (18)$$

where again, the primes are meant to distinguish the parameters in the simplified model from those in the original model.

Figure 3(a) shows the comparison of the analytical solution of $k(t)q(t)$ for both the original model [Eq. (13)] and the simplified model [Eq. (18)] using the parameters given in Meddis (1988). Figure 3(b) compares simulation results from both models with the same input $k(t)$, a sinusoid wave form at 1 kHz. Both plots show that solutions from the simplified equations and the original equations are essentially identical, and the accuracy of the prediction does not depend on the input stimulus (e.g., whether it is stationary or nonstationary). These results also hold for other parameter sets given in other studies (Meddis, 1986, 1988; Sumner *et al.*, 2002; not shown). Thus, the simplified equations provide a good description of the Meddis model, and the solution [Eq. (18)] has two components with different time constants that are the same as those used to describe the characteristics of onset adaptation in AN fibers [Eq. (1)].

The accuracy of the simplified equations describing the Meddis model across different values of the sum $l+r$ is illustrated in Fig. 4. Each component of $q(t)$ (t_i and Φ_i) is plotted as a function of $l+r$ for both models (Fig. 4). The values of these components are constant for the simplified model because u is the same for different values of $l+r$. The values of corresponding components in the original model reach the values in the simplified model as the value of $l+r$ increases, and the values are indistinguishable from each
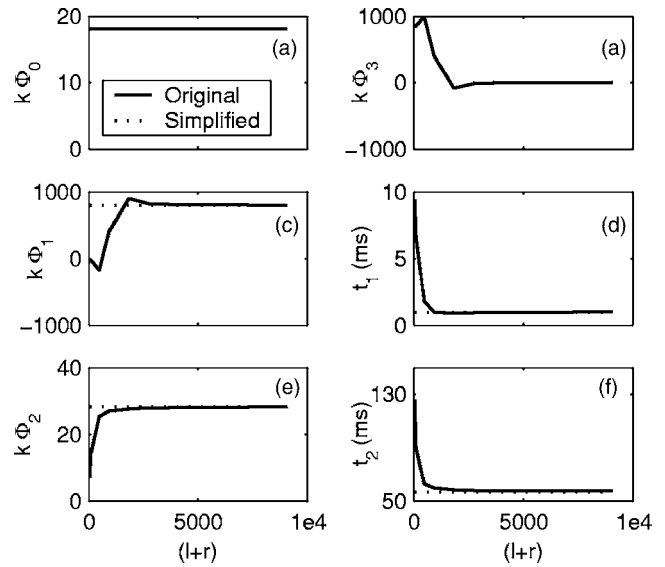
FIG. 4. The difference of each individual component (t_i and Φ_i) of $q(t)$ between the Meddis (1986) model and the simplified model as a function of $l+r$. The values of these components are constant for the simplified model (dotted line) because $u=r/(l+r)$ is kept the same for different values of $l+r$. The amplitude of the component (Φ_i) is scaled by k so it can be interpreted as the instantaneous firing rate of the synapse output. The values of corresponding components in the original model (solid line) all converge quickly to the values in the simplified model as $l+r$ increases, and the values are indistinguishable from each other when $l+r>5000$. The amplitude (Φ_3) of the additional component does not exist for the simplified model; in the original model, this amplitude converges to zero as $l+r$ increases and it can be neglected when $l+r$ is large.

other when $(l+r)>5000$ (the values of $l+r$ used in the Meddis (1986, 1988) were usually greater than 15 000). The amplitude (Φ_3) of the additional component in the original model can be neglected compared to the other components when the value of $l+r$ is large. A detailed interpretation of the relationship between the simplified and original Meddis models is further explained in Sec. VI.

Because $k(t)$ and $P_I(t)$ are the only stimulus-dependent variables in the simplified Meddis and Westerman models, the two models can be compared more directly by substituting $k(t)$ with $P_I^M(t)$ and $q(t)$ with $C_I^M(t)$ in Eqs. (14) and (15). The constant yM in Eq. (14) can be removed by substituting $w(t)$ with $[(y+xu)/x][C_L^M(t) - yM/x - uC_I^M(t)]$. After making these changes, we have the following equations describing the simplified Meddis model:

$$\frac{dC_I^M}{dt} = -(y + xu)C_I^M(t) - P_I^M(t)C_I^M(t) + (y + xu)C_L^M(t), \quad (19)$$

$$\frac{dC_L^M}{dt} = \frac{xyM}{y + xu} + xu\left(\frac{x}{y + xu} - 1\right)C_I^M(t) - x(1 - u)C_L^M(t). \quad (20)$$

Since there is one free parameter in Westerman's model, we can always assume $V_I=1$ (that is, we can decrease V_I, V_L, P_I, P_L, P_G and increase C_I, C_L, C_G by the same scale, V_I, and the synapse output P_IC_I remains the same). It is obvious that the equations describing the two models [Eqs. (6), (7), (19), and (20)] are directly comparable. If we set the correspond-

X. Zhang and L. H. Carney: Analysis of auditory-nerve synapse

ing parameters to the same values (e.g., $P_L = y + xu$), then the synaptic response of the simplified Meddis model $[k(t)q(t)]$ and that of the Westerman model $[P_I(t)C_I(t)]$ to arbitrary inputs should be exactly the same.

IV. DERIVING MODEL PARAMETER VALUES FROM PERI-STIMULUS TIME HISTOGRAM PROPERTIES: THE RELATIONSHIP BETWEEN MODEL PARAMETERS AND ADAPTATION CHARACTERISTICS

The values of the parameters in the onset adaptation function [Eq. (1)] are usually determined by fitting this characteristic equation to the PST histogram of AN fiber responses. The refractoriness in the PST histograms can be removed before fitting (Westerman and Smith, 1988), so the IHC-AN synapse output (instantaneous firing rate) can be more accurately represented by the characteristic function.

With these adaptation parameters $(A_{sus}, A_r, A_{st}, \tau_R, \tau_{ST})$ and the spontaneous rate (A_{sp}) of the AN fibers, the values of parameters in both models can be derived. The analytical solution for the Westerman model is reported in previous studies (Westerman and Smith, 1988; Zhang et al., 2001). The derivation of the analytical solution for the values of the parameters in the simplified Meddis model from these characteristic parameters is given in the Appendix. Since any desired adaptation responses can be obtained from both models with appropriate parameters, comparing adaptation responses for different models with specific parameters will not give us any insight concerning these models.

It is important to note that the mapping from V_{IHC} to $k(t)$ in the Meddis model, or $P_I(t)$ in the Westerman model, should change with the fiber adaptation characteristics: each adaptation characteristic requires different sets of the model parameters (including the permeability at rest and the permeability at any other desired level); the spontaneous rate, rate threshold, and rate-level function curve differ among fiber types that require different V_{IHC}-k (V_{IHC}-P_I) relationships. An appropriate V_{IHC}-k (V_{IHC}-P_I) representation is critical for a composite model that successfully describes AN responses (Heinz et al., 2001; Sumner et al., 2002). The procedure for deriving the appropriate V_{IHC}-k (V_{IHC}-P_I) function combining the results in the present study and recent developments of these models are further discussed in Sec. VI.

Though the same desired adaptation responses can be obtained from both models, the dependence of the parameters in each model upon various adaptation characteristics differs. Also, because the models have different structures, these parameters have specific physical meanings in each model. A survey of how these parameters change with different model-AN fiber types will provide valuable information and further our understanding of both models.

Figures 5 and 6 show an example of how the parameters change systematically for different model-AN fiber types in the simplified Meddis model and Westerman models, respectively. Other possible ways to specify the adaptation across different model-AN fiber types are detailed in Sec. VI. Here, the fiber types are described by the peak-to-sustained ratio $[PTS = (A_r + A_{st} + A_{sus})/A_{sus}]$ of the onset adaptation, which changes with spontaneous rate, A_{sp} (Rhode and Smith, 1985), as described by the following equation:
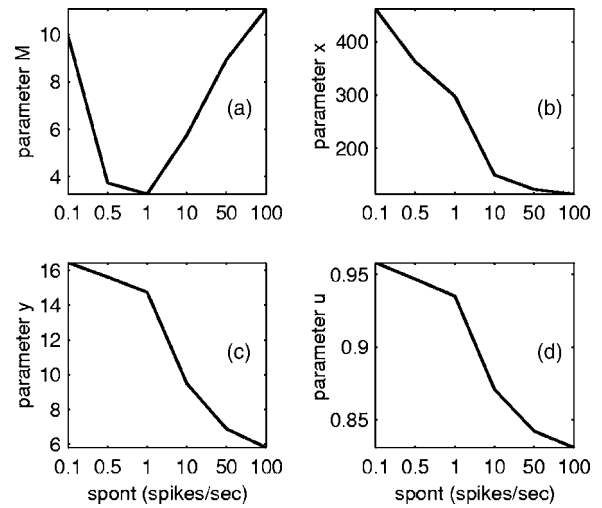
FIG. 5. The relationship between parameter values (M, x, y, and u) and spontaneous rate for the Meddis model. Other adaptation parameters used to derive the model parameters are: $PTS = (A_r + A_{st} + A_{sus})/A_{sus} = 1 + 9A_{sp}/(9 + A_{sp})$, $\tau_R = 2$ ms, $\tau_{ST} = 60$ ms, $A_r/A_{st} = 6$, and $A_{sus} = 350$. The value of M in the Meddis model changes nonmonotonically with increasing spontaneous rate; all other parameters decrease as spontaneous rate increases.

$$PTS = 1 + 9 \times A_{sp}/(9 + A_{sp}). \tag{21}$$

The other adaptation parameters are held constant across model AN fibers ($\tau_R = 2$ ms, $\tau_{ST} = 60$ ms, $A_r/A_{st} = 6$, $A_{sus} = 350$). The parameter values from Fig. 5 for several specified spontaneous rates are given in Table I. Figure 5(a) shows that the value of M in the Meddis model changes nonmonotonically with increasing spontaneous rate; it drops initially and then increases from 3 to 10 as spontaneous rates increase from 1 to 100 spikes/s. All other parameters (x, y, and u) in the Meddis model decrease continuously as the model's spontaneous rate increases. The parameter changes in Westerman's model are plotted in Fig. 6. The global concentration (C_G) was always set to 1
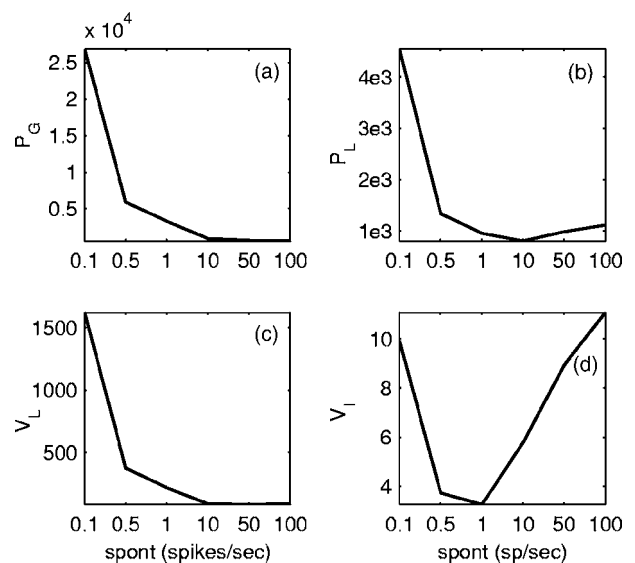
FIG. 6. The relationship between parameters (P_L, P_G, V_L, V_I) and the spontaneous rate in the Westerman model. The other adaptation parameters are the same as in Fig. 5. The parameters were derived with global concentration (C_G) equal to 1.

TABLE I. Parameters in the simplified Meddis model for several fiber types, specified by the spontaneous rate in parentheses in the first column.

	x	y	M	u	$k1$	$k2$
HSR (60)	120.3	6.63	9.4	0.84	7.6	389.7
MSR (10)	149.6	9.48	5.8	0.87	1.78	357.6
LSR (0.1)	461.4	16.43	9.9	0.96	0.01	38.80

before determining the other parameter values. The trends in the value of the immediate volume, V_I [Fig. 6(d)], are similar to those of M observed in Fig. 5(a). Other parameters (P_G, P_L, and V_L) decrease initially and then are insensitive to the value of spontaneous rate above 5 spikes/s.

The A_{sus} defined in the adaptation parameters is the sustained rate at a certain stimulus level. The level is not specified here but is presumably a high level determined by the level-permeability function in the rest of the composite model for the AN response. The values of model parameters derived from these adaptation parameters are specified for this level (a medium or high level for high-spontaneous-rate model fibers, and a very high level for low-spontaneous-rate model fibers). The assumption of both models is that all of these parameters except one (P_I in Westerman model, and k in Meddis model) are constant across different levels. The adaptation characteristics of the models change as the value of the level-dependent parameter changes. When the level-dependent permeability (P_I, or k) increases, the sustained rate in both models saturates at a value:

$$A_{sus\_MAX} = \lim_{P_i \to \infty} P_I C_I = \frac{P_L P_G C_G}{P_L + P_G} = \lim_{k \to \infty} kq = \frac{yM}{1-u}. \quad (22)$$

The value of $A_{sus,\_MAX}$ also changes with the AN fiber's spontaneous rate (see Fig. 7): The value for high-spontaneous-rate model fibers is comparable to A_{sus},
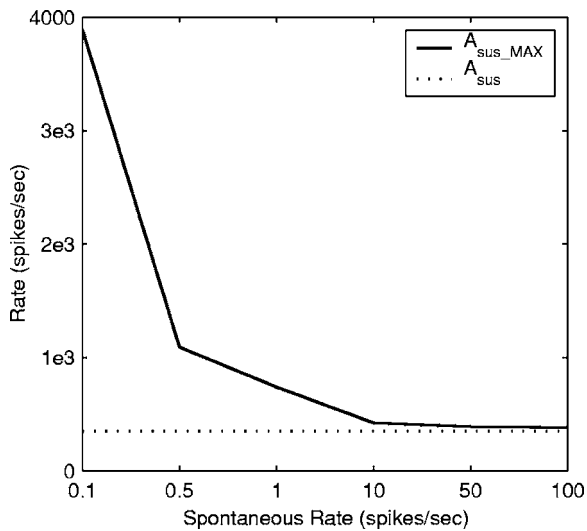
FIG. 7. The maximum sustained rate [A_{sus_MAX}, solid line; see Eq. (22)] for both Meddis and Westerman models as a function of model spontaneous rate. The other adaptation parameters are set as in Fig. 5 (thus, A_{sus} has a constant value, as shown by the dotted line). The value of A_{sus_MAX} for high-spontaneous rate model fibers is comparable to A_{sus}, whereas the value increases rapidly as spontaneous rate decreases.

whereas A_{sus_MAX} increases rapidly as spontaneous rate decreases. The results above may suggest a fundamental difference between low- and high-spontaneous-rate AN fibers: The high-spontaneous rate fibers have a saturated response rate around A_{sus} with an input level higher than specified, and low-spontaneous-rate fibers will not have a saturated response at higher levels due to a much larger A_{sus_MAX} value than that of A_{sus}.

It is also possible to study analytically the effects of the model parameters on adaptation characteristics with the solution given above. For example, Eq. (17) is the analytical solution for the simplified Meddis model. The steady-state value of q is given by:

$$q_\infty = sQ(s)|_{s=0} = \frac{yM}{y + k_2(1-u)}. \quad (23)$$

If we denote k_L as the value of k at a specified sound stimulus level (L), the steady-state response rate, $R(L)$, can be derived:

$$R(L) = k_L q_\infty = \frac{k_L yM}{y + k_L(1-u)}. \quad (24)$$

With the values of the other model parameters fixed, the model's steady-state response $R(L)$ changes proportionally with the value of M. Sumner et al. [2002, Fig. 5(c)] reported the simulation results of rate-level functions with different values of M. Their results and comments are in fact a direct reflection of the relationship explained in Eq. (24).

Since P_I (or k in the Meddis model) is level dependent, changes of the model adaptation response as a function of the value of P_I after onset represent the effect of level on the model onset characteristics. If we set the value of P_I before onset (P_{I1}) to another value, we can also study the model adaptation responses to increments and decrements using the same paradigm (a comprehensive survey of adaptation in the Meddis model with specified parameters is provided by Sumner et al., 2003).

Figure 8 shows the change of components of the model onset adaptation as a function of the value of immediate permeability P_I after onset (or k in the Meddis model) for a model IHC-AN synapse with spontaneous rate of 50 spikes/s. The model has a spontaneous response with permeability of P_{I1} when there is no stimulus input. The sustained rate (A_{sus}) is roughly saturated at permeability P_{I2}, which generates the desired onset adaptation, while the contribution of short-term and rapid components continues to change as the permeability P_I increases [Fig. 8(a)]. Both the rapid decay time constant and short-term decay time constant decrease as P_I increases [Fig. 8(b); note that τ_R is multiplied by 10 in this plot]. For the simplified Meddis model, these two time constants are determined by the poles in Eq. (17), and values of $-1/\tau_R$ and $-1/\tau_{ST}$ are the roots of the polynomial function in the denominator. The values of τ_R and τ_{ST} can be approximated by $1/k$ and $1/(x-ux)$ when $k \gg x, y$ (i.e., at a medium or high stimulus level for this model fiber). For the model illustrated in Fig. 8, the short-term time constant is about 120 ms at rest and decreases to 50 ms at high levels, and the rapid time constant is 8 ms at rest and de-

X. Zhang and L. H. Carney: Analysis of auditory-nerve synapse
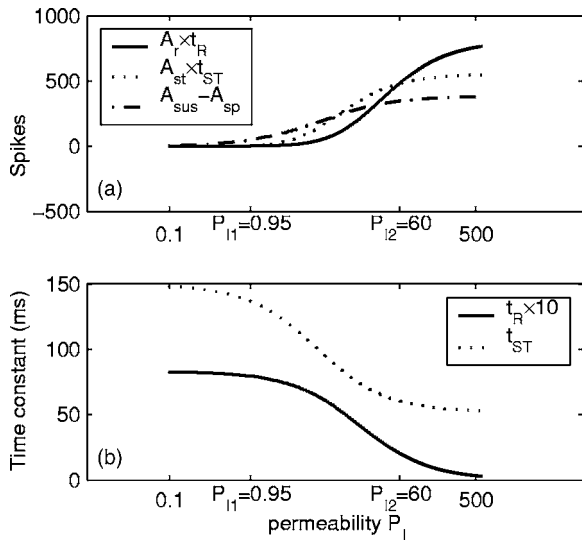
FIG. 8. Effects of the model parameter, P_I, (for the Westerman model) on onset adaptation characteristics. The model parameters were derived from adaptation properties specified in Fig. 5 with a spontaneous rate of 50 spikes/s. The values of P_{I1} and P_{I2} are the values of the immediate permeability before and after the stimulus onset. (a) The contribution of each PST onset component. The contribution of the rapid component (solid line) and short-term component (dotted line) are the integrals of their exponential functions (and thus have units of spikes). The contribution of the constant component (sustained rate–spontaneous rate; dashed line) saturates as P_I increases. (b) Both rapid and short-term time constants decrease as the permeability increases. The rapid time constant (τ_R, solid line) is multiplied by 10 to allow it to be plotted with the short-term time constant (τ_{ST}, dotted line).

creases continuously as the input sound level increases (the value we chose to derive the model parameters was $\tau_{ST} = 60$ ms, $\tau_R = 2$ ms). Since onset adaptation cannot be estimated accurately at low sound levels from physiological responses, these time constants in the model at rest are more related to other adaptation characteristics (e.g., offset adaptation, as discussed in next section, or recovery of onset adaptation).

V. MODIFYING THE SYNAPSE MODEL TO OBTAIN THE DESIRED OFFSET ADAPTATION RESPONSE

The above analysis shows that the model parameters can be determined on the basis of onset adaptation characteristics (at a certain stimulus level). When the model parameters are set, the other adaptation characteristics of the model synapse are automatically determined. If we set the level-dependent permeabilities P_{I1} and P_{I2} (k_1 and k_2 in the Meddis model) to the values before and after the stimulus offset, the analytical solution from Eq. (8) also provides a description of the responses of the model synapse after stimulus offset.

A direct conclusion is that the offset adaptation of the model synapse is also a sum of two exponential functions. The time constants of these two components are determined by P_I after offset and limited by the time constant at the onset. Since only P_{I2} affects the time constant [see Westerman and Smith, 1988, for Westerman's model and Eq. (17) for the Meddis model], Fig. 8(b) also shows how the time constant in the offset adaptation changes with the permeabil-

ity after offset. Physiological studies (Smith and Zwislocki, 1975; Smith, 1977; Westerman, 1985) suggest that AN fibers with medium- or high-spontaneous rate usually stop responding immediately after offset (which begins the deadtime period) and recover slowly with a time constant longer than that of short-term adaptation. However, the rapid component of the model recovery function causes the synapse to recover quickly after stimulus offset. Since $P_I(t)$ and $C_I(t)$ cannot be negative, the magnitude of these recovery components [ϕ'_1, ϕ'_2 in Eqs. (8) and (18); the sign of these components is negative at stimulus offset] is also limited by the value of ϕ'_0 (determined by the spontaneous rate). Because of these limitations, the offset adaptation of the model response is unrealistic as compared to the physiological responses specified in Eq. (2).

The function fitting offset adaptation responses has another general form (Harris and Dallos, 1979):

$$R_{off}(t) = (A_{sp} - A_{min})(1 - e^{-t/\tau_{sp}}) + A_{min} \quad \text{for } R(t) > 0$$
$$= 0 \quad \text{for } R(t) < 0, \quad (25)$$

where A_{min} is a negative value that accounts for the deadtime period of the offset response. Since an AN fiber has a very low probability of response during the deadtime period, the recovery process is not well studied, and there may be a rapid recovery component that corresponds to the onset rapid adaptation and becomes insignificant before the end of the deadtime period. The offset characteristic function can thus be described as:

$$R_{off}(t) = (A_{sp} - A_{min})(1 - e^{-t/\tau_{sp}}) + A_{min}$$
$$+ A_r^{off} e^{-t/\tau_r^{off}} \quad \text{for } R_{off}(t) > 0$$
$$= 0 \quad \text{for } R_{off}(t) < 0, \quad (26)$$

where A_r^{off} is the magnitude (with a negative value) and τ_r^{off} is the time constant of rapid recovery component. Since the model synapse output $P_I C_I$ in the Westerman model cannot be negative, we can assume that the measured synapse output is:

$$R_{off}(t) = P_I(t)C_I(t) - A_{shift} \quad \text{for } R_{off}(t) > 0$$
$$= 0 \quad \text{for } R_{off}(t) < 0, \quad (27)$$

where A_{shift} is a shift value to guarantee that $P_I C_I > 0$, for all t, and that $R_{off}(t)$ is the same offset adaptation function as in Eq. (26) (thus $A_{shift} > |A_r^{off}| + |A_{min}|$). The onset of the model synapse output now becomes

$$P_I(t)C_I(t) = R_{on}(t) + A_{shift}$$
$$= A_{shift} + A_{sus} + A_r e^{-t/\tau_r} + A_{st} e^{-t/\tau_{st}}. \quad (28)$$

Thus, by including this shift, the same equations can describe adaptation of both the onset and offset in the Westerman model. By substituting $P_I(t)C_I(t)$ with $k(t)q(t)$ in Eqs. (27) and (28), the same procedure can be applied to the Meddis model to obtain the desired onset and improved offset adaptation using the same equations.

Figure 9 shows the onset and offset responses of the synapse model with or without the shift. The value of the model parameters was derived to have the same onset adap-
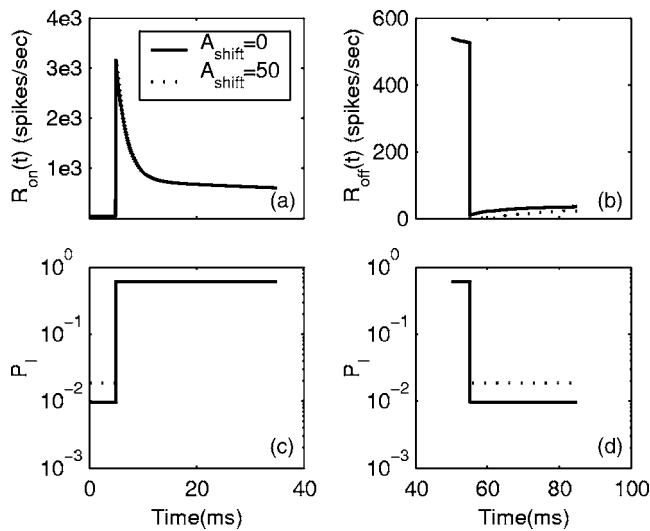
FIG. 9. Onset and offset responses of the Westerman model with (dotted line) or without (solid line) the shift value added. The shift value is set to 50 spikes/s. (a) The model onset responses are unchanged because the model parameters were derived to have the same onset adaptation. (b) The model offset response with the shift recovers more slowly than does the model offset response without the shift and has the desired deadtime period. (c), (d) The permeability $P_I(t)$ jumps from P_{I1} to P_{I2} at the onset (5 ms) and drops back to P_{I1} at the offset (55 ms). The model parameters (e.g., P_{I1}) were changed to maintain the same onset response.

FIG. 10. Effects of the parameter A_{shift} in Eq. (27) on the model offset components. The model parameters are derived so they have the same onset adaptation with different A_{shift} values [spontaneous rate $A_{sp}=50$, $A_{sus}=350$, PTS=$(A_r+A_{st}+A_{sus})/A_{sus}=9$, $\tau_R=2$ ms, $\tau_{ST}=60$ ms, $A_r/A_{st}=6.0$; see Eq. (1)]. The upper plots illustrate how the magnitude and time constant of the offset rapid component changes as a function of shift; the middle plots show changes of the offset short-term component. The time constants for both components decrease as shift increases and are larger than the corresponding onset time constants. The rapid component recovers quickly (usually during the deadtime period) and the recovery time constant is thus determined by the short-term component. The magnitude of this component is always negative and decreases (becomes more negative) as the shift increases. A_{min} [bottom right; see Eqs. (25) and (26)] changes from positive to the desired negative as the shift value increases; A_{min} and the time constant of the short-term component determine the duration of the deadtime period. The plot at the bottom left shows the most negative value (including the contribution of the rapid component, which recovers quickly) of the model response without rectification [$R_{off}(t=0)$ in Eq. (24)] as a function of the shift value.

tation. The permeability $P_I(t)$ jumps from P_{I1} to P_{I2} at 5 ms and drops back to P_{I1} at 55 ms. When the shift ($A_{shift}=50$) was added, the model parameters [e.g., P_{I1} in Figs. 9(c) and 9(d)] were changed to maintain the same onset response. Thus, the original model and modified model have the same onset responses [Fig. 9(a)] while the modified model has a greatly improved offset adaptation response [Fig. 9(b)].

The effect of A_{shift} on the offset components is illustrated in Fig. 10. Since the rapid component recovers quickly during the deadtime period, the offset adaptation process is dominated by the short-term component. The magnitude of the short-term component [$A_{min}-A_{sp}$ in Eq. (25)] determines the duration of the deadtime period, and the time constant (τ_{ST}) determines the value of the recovery time constant τ_{SP}. The recovery time constant is always greater than that of the onset short-term adaptation (60 ms), although the value decreases as the shift is increased. The value of A_{min} [Eq. (25)] becomes more negative with a larger shift, which produces the desired deadtime period.

The offset adaptation properties of the IHC-AN synapse could account for enhanced phase locking to the stimulus envelope in AN fibers. The AN fibers were less responsive during the dip of the envelope because of the offset adaptation, and thus the fibers were more synchronized to the peak of the envelope. Figure 11 illustrates the effects of offset adaptation on the modulation transfer function [(MTF), see Joris and Yin, 1992, for a comprehensive experimental study] for a model AN fiber (Zhang *et al.*, 2001) to a sinusoidally amplitude-modulated stimulus.[1] The IHC-AN synapse component of that model was replaced with the synapse model presented here (the onset adaptation parameters were the same as in previous study), and a different shift value was used to represent offset adaptation (i.e., a shift value of

0 means the synapse model was the same as the original model). The stimulus was always fully modulated (modulation index $m=1$) at 10 dB sound pressure level (SPL), with the carrier frequency set equal to the AN CF of 21 kHz. The model AN fibers with offset adaptation (e.g., with a larger shift value) had increased modulation gain and an enhanced MTF that was more consistent with the data from physiological experiments (Joris and Yin, 1992).

VI. DISCUSSION

In this section, we first discuss the differences and similarities between the different models under investigation. We then discuss how to use the analytical methods developed in the present study to further our understanding of adaptation in models and in the IHC-AN synapse. We consider different adaptation measures that can be used to describe a single AN fiber, as well as the fact that different adaptation properties may be observed for different AN fiber types. Finally, we discuss how to extend the analytical method to other models and identify the candidate model that represents the actual underlying physiological mechanism.
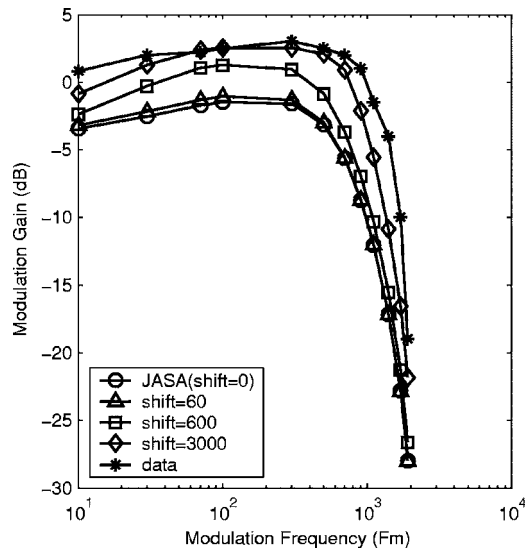
FIG. 11. Modulation transfer function for the model AN fibers. The stimulus was always fully modulated at 10 dB SPL with carrier frequency at AN CF of 21 kHz, gated with a 10-ms cosine-square function. The modulation gain was defined as $20 \log[(\text{modulation of responses})/(\text{modulation of stimulus})]$ (i.e., modulation gain $20 \log(2R/m)$, where R was the vector strength of the AN responses and m was the modulation depth; see Joris and Yin, 1992). The model AN fibers with larger shifts in the synapse had higher modulation gain and were more consistent with the data from physiological experiments (asterisks in the figure, from Joris and Yin, 1992).

A. Interpretation and justification of the simplified Meddis model

The loss rate, l, in the cleft of the original Meddis model, was represented in the simplified model by the fraction of synaptic material recycled, $u=r/(l+r)$, where r is the reuptake rate. The difference between the simplified and original model centers on the following questions: Is the recycling of synaptic material in the cleft a two-stage process, as in the original Meddis model, or a one-stage process, as in the simplified model (by assuming reuptake rate r is much larger than w and removing it)? Or, to put the question another way, is the phenomenological one-stage process of the simplified model sufficient to describe the recycling mechanism instead of the two-stage process in the original Meddis model?

With a value of r much larger than the values of x and k, the one-stage process in the simplified model definitely generates the same response as that of the two-stage process in the original Meddis model (see Fig. 4). Even when the value of r decreases to a value comparable to k (still much larger than x), the responses from the simplified model are still almost the same as those of the original model, since the recycling process contributes to the short-term adaptation, which is mainly determined by the value of x. For values of $l+r$ between 1000 and 5000 (Fig. 4), the amplitude and time constant of short-term adaptation does not change much, and the additional component (ϕ_3) which contributes to the rapid adaptation increases at the expense of another rapid component, ϕ_1. For a very small value of r, both reuptake (r) and reprocessing (x) will affect the short-term adaptation in a complex way for the original Meddis model.

It then becomes a question of whether the model with

small values of r and x could predict the adaptation of the IHC-AN synapse. It is not surprising that the exercise of finding model parameters to provide a satisfactory description of adaptation responses ended up with the large value of r in the original Meddis model (Meddis, 1986, 1988; Sumner et al., 2002). These findings justify our simplification of the original Meddis model.

When the value of k becomes extremely high (infinite) after the onset in the original Meddis model, all of the transmitter, $q(0^-)$, in the immediate pool is released into the cleft instantaneously right after the onset. The amount of transmitter in the cleft increases instantly and becomes $q(0^-)+c(0^-)$. Thereafter, all transmitter replenished from the reprocessing store (w) and the global factory into the immediate pool is released into the cleft instantaneously:

$$kq = yM + xw. \qquad (29)$$

The dynamics of the original Meddis model are now governed by Eqs. (4), (5), and (29). The solution of these equations in the Laplace domain is given by:

$$C(s) = \frac{(yM + c(0^-) + q(0^-))(s+x) + xw(0^-)s}{s(s+r/u)(s+x) - rxs}, \qquad (30)$$

where u is defined as the fraction of transmitters recycled, $u=r/(l+r)$.

If the synapse output is proportional to the amount of transmitter in the cleft $c(t)$, the responses in the original Meddis model can again be characterized by Eq. (18), based on Eq. (30). The steady-state response in the original Meddis model is also specified by Eq. (24), and is independent of the value of r. The rapid adaptation time constant (τ_R) can be approximated by u/r, or $1/(l+r)$, when $r/u \gg x$, and this value of τ_R is independent of k at high stimulus levels (k is very large). Thus, with an appropriate value of $l+r$, the original Meddis model is a better candidate than the simplified Meddis model in predicting the changes of the adaptation time constants at high stimulus levels (Westerman, 1985). In reality, the composite AN model with the original Meddis synapse model usually has k operating within a limited range, and thus the original Meddis model can be perfectly replaced by the simplified Meddis model.

B. Physiological assumptions underlying the Westerman and Meddis models

This study shows that two phenomenological IHC-AN synapse models that were designed to explain the onset adaptation characteristics of AN responses are essentially the same. It is impossible to differentiate these two models by evaluating their adaptation responses. There are many similarities between the Westerman and Meddis models. Rapid adaptation is modeled as a depletion of the synaptic material in the immediate store in both models, which is supported by recent physiological studies (Moser and Beutner, 2000).

The sustained response of the IHC-AN synapse in both models is attributed to the replenishment of synaptic material from the global store to the immediate store, either directly or indirectly. For model fibers with little adaptation but considerable sustained response rate (e.g., some low-

TABLE II. Adaptation characteristics derived from three model fibers specified in Sumner *et al.* (2002). The values of M, spontaneous rate, and sustained rate of fibers are specified in the first three columns. The sustained response rate of fibers is always 10 spikes less than the maximum response described in Eq. (22). The values of other model parameters are from Sumner *et al.* (2002): $y=10$, $x=66.3$, $l=2580$, $r=6580$. The values of $k1$ and $k2$ are then derived from Eq. (24), and the other adaptation characteristics are derived from the simplified Meddis model.

	M	A_{sp}	A_{sus}	$k1$	$k2$	τ_R (ms)	A_r (spikes/s)	τ_{ST}(ms)	A_{st} (spikes/s)
H1	10	60	345	7.2202	1225	0.78	9660	54.5	174.6
M1	13	10	451	0.7863	1603	0.60	19667	54.3	271.6
L1	8	0.1	274	0.0125	972.9	0.97	7340	54.7	167.0

spontaneous-rate fibers), the replenishment rate of synaptic material from the global pool into the immediate store must be comparable to the release rate of synaptic material from the immediate store into the cleft. For the Meddis model, either a small k or large y at high levels is needed for these model fibers (or both; see Table I and Fig. 5). To overcome this limitation, another global store that releases synaptic materials directly into the cleft can be included to generate the sustained response. This parallel global synaptic store separates the sustained response from adaptation, and adaptation across different fibers can be adjusted by changing the value of M (or C_G in the Westerman model) instead of the value of the replenishment rate of synaptic material into the immediate store.

The key difference between two models is that short-term adaptation is accounted for by the recycling of the synaptic material from the cleft in the Meddis model and by depletion in the Westerman model. Recent physiological studies (Moser and Beutner, 2000) show that the time constant of endocytosis is about 7.5 s in the IHC-AN synapse. However, in the Meddis model, the recycling rate of synaptic material in the cleft must be set to a time scale of approximately 10 ms to account for short-term adaptation. Thus, the model time constant is about 100 times faster than the value obtained in physiological studies.

In the Westerman model, it is assumed that short-term adaptation does not come from endocytosis, and the recycling process is not included in this model. Instead, short-term adaptation is modeled as the depletion of synaptic materials in the local store. Of course, it is entirely possible that the physiological assumptions in both models are wrong, but that the models work because they are sufficient models for representing actual physiological mechanisms that are unknown.

C. Different adaptation measures for single auditory-nerve fibers

Other than onset and offset adaptation, there are many other measures of adaptation: Recovery after offset (Harris and Dallos, 1979; Westerman, 1985), responses to increments (Smith and Zwilocki, 1975; Smith *et al.*, 1985), decrements (Smith *et al.*, 1985), and conservation (Westerman, 1985). Sumner *et al.* (2003) provided a comprehensive survey of these adaptation measures in a recently developed version of the Meddis model, with several sets of model parameters specified. These results can be applied to the Westerman model quite naturally since the two models are essentially the same, with the appropriate parameters specified. Our analytical results allow the adaptation characteristics to be derived directly from the specified model parameters (Table II).

Instead of evaluating whether specified model AN fibers can predict different measures of adaptation, our analytical results can provide insights for the limitations of these models, and explain the performance of these models for different adaptation measures. For example, Fig. 8 showed the changes of onset adaptation components with the stimulus-level-dependent variable k in the simplified Meddis model (or P_I in the Westerman model). The rapid and short-term adaptation time constants, τ_R and τ_{ST}, approach $1/k$ and $1/(x-ux)$, respectively, as the value of k increases. The value of τ_R is more affected by the value of k than that of τ_{ST}, especially when they are plotted on a log scale. These results explain the simulation results reported by Sumner *et al.* (2003), and provide insights into the failure of these models in predicting the changes of the rapid adaptation time constant at low sound levels.

For the Westerman model, recovery after stimulus offset of the onset rapid and short-term adaptation can be interpreted as the replenishment of the synaptic material in the immediate and local stores. The recovery time constant of rapid adaptation is thus determined by the permeability, P_L, between the local and immediate stores, which also determines the time constant of short-term adaptation (depletion of synaptic material in the local store) at onset. This explains why the recovery time constant of rapid adaptation is comparable to the short-term adaptation time constant instead of that of rapid adaptation, consistent with both physiological studies (Westerman, 1985; Moser and Beutner, 2000) and simulation studies (Sumner *et al.*, 2003).

The analytical results from these models show that different adaptation measures are highly correlated for a single model IHC-AN synapse. It is of great interest to see how these adaptation measures are correlated in the physiological responses of single AN fibers. The correlation studies of adaptation measures across AN fibers may be limited by the variation of physiological properties across AN fibers.

The values of model parameters can be derived from adaptation characteristics at different stimulus levels (Westerman and Smith, 1988). When more than one model parameter varies with level, the underlying assumptions of the method used here to generate the model parameters (e.g., allowing only P_I or k to change with level) is invalidated. In this situation, it is more mathematically challenging to estab-

lish the relationship between adaptation characteristics and model parameters. On the other hand, when this level-dependent relationship is established using more complex techniques, the model may provide more accurate predictions for different measures of IHC-AN adaptation, and it may also provide insights into the physiological properties specified by the model parameters.

D. Adaptation across different auditory-nerve fiber types and transmitter release functions in model auditory-nerve fibers

Our analytical approach makes it possible to link adaptation characteristics directly to the values of model parameters. Making models to predict adaptation across AN fiber types then becomes a question of how to determine different adaptation characteristics across AN fiber types. In the present study, we considered the example of a description of different fiber types based on the peak-to-sustained ratio $[PTS=(A_r+A_{st}+A_{sus})/A_{sus}]$ of the onset adaptation, which changes as a function of spontaneous rate, A_{sp} (Rhode and Smith, 1985) [Eq. (21)]; other parameters were fixed across different fiber types ($\tau_R=2$ ms, $\tau_{ST}=60$ ms, $A_r/A_{st}=6$, $A_{sus}=350$). For low-spontaneous-rate fibers with PTS close to a value of 1, values of A_r and A_{st} are very small. Since there is little contribution of these adaptation components to AN fiber responses, the values of other parameters (τ_R, τ_{ST}, and A_r/A_{st}) may not be determined with confidence based on a fit of the PST histograms, and models with a wide range of parameter values can provide equally good fits to the AN fiber responses.

It is possible to determine adaptation characteristics using the results of other experimental paradigms. Relkin and Doucet (1991) found that the recovery time constant for low-spontaneous-rate fibers is much larger than that of high-spontaneous-rate fibers, and the small peak-to-sustained ratio observed for low-spontaneous-rate fibers (Rhode and Smith, 1985) may have been partly determined by the short inter-stimulus intervals used to study them (105 ms for Rhode and Smith, 1985). The relationship between adaptation characteristics of different measures and model parameters can be created using our analytical approach, and values of model parameters can be derived accordingly.

When some characteristics of adaptation cannot be determined with confidence by fitting the PST histograms for certain AN fiber types (e.g., low-spontaneous-rate fiber), it is reasonable to fix some model parameters across all model fiber types (e.g., y for the Meddis model, or P_G for the Westerman model), and then derive the values of the other parameters in the model. The assumption underlying this method is that the corresponding physiological property specified by the model parameter does not change across different fiber types. It would be very interesting to see how these corresponding physiological properties varied across the IHC-AN synapse in physiological studies.

In addition to specifying the model parameters to explain adaptation, it is also critical to find an appropriate transmitter release function for the AN model fiber to generate desired rate-level responses. The relationship between IHC voltage (V_{IHC}, referenced to the IHC resting potential),

which is stimulus dependent, and release rate from the immediate store (k in the Meddis model, or P_I in the Westerman model) is highly simplified and usually specified by an *ad hoc* function in most modeling studies.

For the Meddis model, the relationship between the steady-state response rate and k is specified in Eq. (24), and the rate-level function is further determined by the V_{IHC}-k relationship. For a simple linear V_{IHC}-k relationship, the value of k at rest ($V_{IHC}=0$) determines the spontaneous response of the model fiber, and the slope of the V_{IHC}-k function determines how fast the response will approach saturation [A_{sus_MAX} in Eq. (22)] and whether or not the rate-level function will saturate. It is thus not surprising that equivalent parameters in these models varied across different model fiber types (Heinz *et al.*, 2001; Sumner *et al.*, 2002), and that the values of these parameters affected the shape of the rate-level functions in simulations [Figs. 5(a) and 5(b) in Sumner *et al.*, 2002, Fig. 1 in Sumner *et al.*, 2003].

Another useful conclusion from Eq. (22) is that the saturation of the synapse response is largely contributed by the synapse model itself, rather than by saturation of the value of k (or P_I). The fact that P_I does not saturate as V_{IHC} increases at high levels (Carney, 1993; Zhang *et al.*, 2001) is important for the model to predict AN fiber temporal response properties (e.g., synchronization index) across levels.

The V_{IHC}-k (or V_{IHC}-P_I) relationship for model AN fibers should also be characteristic frequency (CF) dependent. The voltage responses of the IHC at various basilar-membrane locations to a tone at CF have different ac/dc components (Cheatham and Dallos, 1993),[2] and attenuation of the AC component at high frequencies suggests that a steeper slope of V_{IHC}-k relationship (or V_{IHC}-P_I, see Zhang *et al.*, 2001) is required to generate similar rate-level functions.

Physiologically, the transmitter release at the IHC-AN synapse is controlled through V_{IHC}-dependent calcium dynamics, and is highly nonlinear and complex. Calcium dynamics have been considered in recent versions of these models. Sumner *et al.* (2002) included a specific calcium-controlled transmitter-release function in the Meddis model, and Zhang *et al.* (2001) used low-pass filters and an *ad hoc* control function to model the effects of calcium dynamics.

E. Understanding the relationship between offset and onset adaptation

The study shows that a simple shift can improve the model's offset adaptation without compromising the model's onset adaptation characteristics. Inclusion of offset adaptation in the model improves the model's prediction of the temporal aspects of the AN fiber responses and thus benefits other modeling studies based on the temporal responses of AN fibers (e.g., modulation studies). A negative adaptation was used in a previous study (Ross, 1996) to produce the deadtime period in the offset adaptation. This negative adaptation was limited to one of the parallel pathways and thus had a limited contribution to the temporal responses of AN fibers. The physiological meaning of the negative adaptation also needs further explanation.

The shift A_{shift} introduced in the present model can be

interpreted as a constant threshold (either concentration or number of neurotransmitter molecules) in the synaptic cleft before the transmitter reaches the postsynaptic site. When the transmitter concentration in the cleft is higher than A_{shift}, the firing probability in the postsynaptic neuron changes linearly as a function of concentration. The firing probability in the postsynaptic neuron approaches zero quickly when the transmitter concentration is less than A_{shift} (so zero in the model output could be interpreted as a very low firing probability). Since the offset adaptation of the AN fiber is related to the spontaneous rate (Harris and Dallos, 1979; Westerman, 1985), it is reasonable for the value of the shift to vary as a function of the spontaneous rate of the fiber as well. Further study of how the value of the shift is related to the other characteristics of the model fiber responses is needed.

Also, the results show that the offset adaptation time constant is closely related to the short-term constant of the onset adaptation for the model explored. In fact, if the model onset adaptation is measured from level L_0 to level L_1 (L_0 is the level before the onset and L_1 is the level after onset) and offset adaptation is measured from level L_2 to L_1 (L_2 is the level before the offset and L_1 is the level after the offset), the time constants for the onset short-term component and for the offset adaptation are the same, based on Eq. (17). Whether this is a limitation of the model or reflects some underlying mechanism of the synapse requires further examination.

Our modification to improve the model's offset adaptation is based on the assumed shape of offset adaptation from physiological studies: A period of deadtime (or period of low firing probability), followed by a single slow recovery function. Whether other characteristic functions can provide a better fit for the physiological data is an interesting topic for future studies. The modification of the model will be further justified if future studies suggest that this characteristic function provides a better fit to the physiological data than that of other models.

F. Extension of analytical approach to other models

The mathematical methods proposed in this study are also useful for analysis of models with other structures. Several models (Smith and Brachman, 1982; Ross, 1996) include parallel pathways of transmitter reservoirs to produce the additivity of responses to increments. These models could also be described by differential equations, and the relationship between adaptation characteristics and model parameters can be established for further understanding of these models.

The analytical approach provided in this study can help us to identify the candidate model that represents the actual underlying physiological mechanism. Values of model parameters can be derived from our analytical approach for different amounts of adaptation in the IHC-AN synapse, due to either different IHC-AN synapses (i.e., different AN fiber types) or to different stimulus levels. By studying the physiological changes of the synapse with different amounts of adaptation, and the correlation between changes of the model parameters and physiological properties, we can validate if the model can represent the actual physiological mechanism. For example, physiological studies (Moser and Beutner, 2000) support the idea that rapid adaptation is due to the depletion of the synapse material in the immediate store that is present in both models considered here.

Modeling the IHC-AN synapse is still a very challenging task because of the synapse structure (how the reservoirs are connected) and the processing dynamics (how the V_{IHC} affects model parameters such as permeabilities). While it is easy to complicate the IHC-AN synapse model with more physiologically realistic structures, the benefit of such complications in providing more accurate responses should be carefully studied. Modeling studies of other adaptation measures (Hewitt and Meddis, 1991; Sumner *et al.*, 2002, 2003), which focused on models with preset parameters, should also be extended. An analysis of the relationship of these adaptation measures and how they are affected by the model parameters will greatly improve our knowledge of both IHC-AN synapse models and of the underlying physical system.

APPENDIX: RELATIONSHIP BETWEEN ADAPTATION CHARACTERISTICS AND THE MEDDIS-MODEL PARAMETERS

The Meddis model with three neurotransmitter reservoirs (q, w, c) can be described by the simplified Eqs. (A1) and (A2) [rewritten from Eqs. (14) and (15)]:

$$\frac{dq}{dt} = y(M - q(t)) - kq(t) + xw(t), \tag{A1}$$

$$\frac{dw}{dt} = kuq(t) - xw(t). \tag{A2}$$

The model parameters M, y, x, u are constant, and k is the level-dependent permeability, which is denoted as k_1 before the onset and as k_2 after the onset.

The onset adaptation of the model synapse is characterized by:

$$R_{on}(t) = A_{sus} + A_r e^{-t/\tau_R} + A_{st} e^{-t/\tau_{ST}}, \tag{A3}$$

with a spontaneous rate of A_{sp}, where A_{sus}, A_r, and A_{st} represent the response amplitude (instantaneous firing rate) of the sustained component, rapid adaptation component, and short-term adaptation component, and τ_R and τ_{ST} represent the time constants of the rapid and short-term adaptation components.

If we assume that $kq(t)$ is the output of the model synapse, the relationship between the model parameters and adaptation characteristics can be described by the following equations [$-1/\tau_R$ and $-1/\tau_{ST}$ are the poles of $Q(s)$ in Eq. (17)]:

$$\frac{k_1}{k_2} = \frac{A_{sp}}{A_r + A_{st} + A_{sus}}, \tag{A4}$$

$$\frac{yMk_1}{y + k_1(1-u)} = A_{sp}, \tag{A5}$$

$$\frac{yMk_2}{y + k_2(1-u)} = A_{sus}, \tag{A6}$$

$$x + k_2 + y = 1/\tau_R + 1/\tau_{ST}, \tag{A7}$$

$$x(y + k_2(1-u)) = 1/(\tau_R \tau_{ST}), \tag{A8}$$

$$\frac{k_2}{k_1}(k_2 - k_1)A_{sp} = A_r/\tau_R + A_{sus}/\tau_{ST}. \tag{A9}$$

Several intermediate parameters are useful for the derivation of the model parameters:

$$A_{on} = A_{sus} + A_r + A_{st}, \tag{A10}$$

$$S_r = 1/\tau_R + 1/\tau_{ST} \tag{A11}$$

$$S_{r2} = A_r/\tau_R + A_{st}/\tau_{ST}, \tag{A12}$$

$$P_r = 1/(\tau_R \tau_{ST}). \tag{A13}$$

From Eqs. (A4) and (A9), k_1 and k_2 can be specified as:

$$k_2 = S_{r2}/(A_{on} - A_{sp}), \quad k_1 = A_{sp}/A_{on}k_2. \tag{A14}$$

Several other intermediate parameters are defined as:

$$\beta = (A_{sus} - A_{sp})k_1 k_2/(A_{sp}k_2 - A_{sus}k_1), \tag{A15}$$

$$a = (\beta + k_2)\beta, \quad b = -(S_r - k_2)(\beta + k_2), \quad c = P_r, \tag{A16}$$

$$z = -\frac{-b + \sqrt{b^2 - 4ac}}{2a}. \tag{A17}$$

The other model parameters then are set as follows:

$$u = 1 - z, \tag{A18}$$

$$y = \beta z, \tag{A19}$$

$$x = S_r - k_2 - y, \tag{A20}$$

$$M = \frac{A_{sp}(y + k_1 z)}{yk_1}. \tag{A21}$$

Computer code related to the material in this Appendix is available at the website http://web.syr.edu/~lacarney/.

[1]The model results presented here are illustrative. More extensive explorations have been conducted in another study (Nelson and Carney, 2004).

[2]We (and others) have assumed that there is no adaptation in the voltage responses of the IHC, but a recent study suggests that there may indeed be some adaptation at this level. It is important to explore the contribution of the IHC adaptation described by Zeddies and Siegel (2004) to AN responses, especially at the onset and offset of tone bursts and in response to AM stimuli.

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**(1), 401–417.

Cheatham, M. A., and Dallos, P. (**1993**). "Longitudinal comparisons of IHC ac and dc receptor potentials recorded from the guinea pig cochlea," Hear. Res. **68**(1), 107–114.

Furukawa, T., and Matsuura, S. (**1978**). "Adaptive rundown of excitatory postsynaptic potentials at synapses between hair cells and eight nerve fibers in the goldfish," J. Physiol. (London) **276**, 193–209.

Harris, D. M., and Dallos, P. (**1979**). "Forward masking of auditory nerve fiber responses," J. Neurophysiol. **42**(4), 1083–1107.

Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). "Auditory-nerve model for predicting performance limits of normal and impaired listeners," ARLO **2**, 91–96.

Hewitt, M. J., and Meddis, R. (**1991**). "An evaluation of eight computer models of mammalian inner hair-cell function," J. Acoust. Soc. Am. **90**(2 Pt 1), 904–917.

Joris, P. X., and Yin, T. C. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**(1), 215–232.

Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**(3), 702–711.

Meddis, R. (**1988**). "Simulation of auditory-neural transduction: Further studies," J. Acoust. Soc. Am. **83**(3), 1056–1063.

Meddis, R., Hewitt, M. J., and Shackleton, T. (**1990**). "Implementation details of a computational model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. **87**, 1813–1818.

Moser, T., and Beutner, D. (**2000**). "Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse," Proc. Natl. Acad. Sci. U.S.A. **97**(2), 883–888.

Nelson, P., and Carney, L. (**2004**). "A phenomenological model of peripheral and central neural responses to amplitude-modulated tones," J. Acoust. Soc. Am. **116**, 2173–2186.

Raman, I. M., Zhang, S., and Trussell, L. O. (**1994**). "Pathway-specific variants of AMPA receptors and their contribution to neuronal signaling," J. Neurosci. **14**(8), 4998–5010.

Relkin, E. M., and Doucet, J. R. (**1991**). "Recovery from prior stimulation. I: Relationship to spontaneous firing rates of primary auditory neurons," Hear. Res. **55**(2), 215–222.

Rhode, W. S., and Smith, P. H. (**1985**). "Characteristics of tone-pip response patterns in relationship to spontaneous rate in cat auditory nerve fibers," Hear. Res. **18**(2), 159–168.

Ross, S. (**1996**). "A functional model of the hair cell-primary fiber complex," J. Acoust. Soc. Am. **99**(4 Pt 1), 2221–2238.

Russell, I. J., and Sellick, P. M. (**1978**). "Intracellular studies of hair cells in the mammalian cochlea," J. Physiol. (London) **284**, 261–290.

Schwid, H. A., and Geisler, C. D. (**1982**). "Multiple reservoir model of neurotransmitter release by a cochlear inner hair cell," J. Acoust. Soc. Am. **72**(5), 1435–1440.

Smith, R. L., and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses of single auditory-nerve fibers," Biol. Cybern. **17**(3), 169–182.

Smith, R. L. (**1977**). "Short-term adaptation in single auditory nerve fibers: Some poststimulatory effects," J. Neurophysiol. **40**(5), 1098–1111.

Smith, R. L., and Brachman, M. L. (**1982**). "Adaptation in auditory-nerve fibers: A revised model," Biol. Cybern. **44**(2), 107–120.

Smith, R. L., Brachman, M. L., and Frisina, R. D. (**1985**). "Sensitivity of auditory-nerve fibers to changes in intensity: A dichotomy between decrements and increments," J. Acoust. Soc. Am. **78**(4), 1310–1316.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). "A revised model of the inner-hair cell and auditory-nerve complex," J. Acoust. Soc. Am. **111**(5 Pt 1), 2178–2188.

Sumner, C. J., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003**). "Adaptation in a revised inner-hair cell model," J. Acoust. Soc. Am. **113**(2), 893–901.

Westerman, L. A., and Smith, R. L. (**1984**). "Rapid and short-term adaptation in auditory nerve responses," Hear. Res. **15**(3), 249–260.

Westerman, L. A. (**1985**). *Adaptation and recovery of auditory nerve responses*, Institute for Sensory Research, Syracuse Univ., Syracuse, N.Y.

Westerman, L. A., and Smith, R. L. (**1988**). "A diffusion model of the transient response of the cochlear inner hair cell synapse," J. Acoust. Soc. Am. **83**(6), 2266–2276.

Zeddies, D. G., and Siegel, J. H. (**2004**). "A biophysical model of an inner hair cell," J. Acoust. Soc. Am. **116**(1), 426–441.

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**(2), 648–670.

Phase of neural excitation relative to basilar membrane motion in the organ of Corti: Theoretical considerations

Masayoshi Andoh, Chihiro Nakajima, and Hiroshi Wada[a)]
Department of Bioengineering and Robotics, Tohoku University, 6-6-01 Aoba-yama,
Sendai 980-8579, Japan

Although the auditory transduction process is dependent on neural excitation of the auditory nerve in relation to motion of the basilar membrane (BM) in the organ of Corti (OC), specifics of this process are unclear. In this study, therefore, an attempt was made to estimate the phase of the neural excitation relative to the BM motion using a finite-element model of the OC at the basal turn of the gerbil, including the fluid-structure interaction with the lymph fluid. It was found that neural excitation occurs when the BM exhibits a maximum velocity toward the scala vestibuli at 10 Hz and shows a phase delay relative to the BM motion with increasing frequency up to 800 Hz. It then shows a phase advance until the frequency reaches 2 kHz. From 2 kHz, neural excitation again shows a phase delay with increasing frequency. From 800 Hz up to 2 kHz, the phase advances because the dominant force exerted on the hair bundle shifts from a velocity-dependent Couette flow-induced force to a displacement-dependent force induced by the pressure difference. The phase delay that occurs from 2 kHz is caused by the resonance process of the hair bundle of the IHC. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2000770]

I. INTRODUCTION

The mammalian cochlea consists of a fluid-filled duct that is coiled like a snail shell. Figure 1 shows a simplified sketch of the cross section of the gerbil cochlea. The cochlea is divided into three compartments, i.e., the scala vestibuli (SV), the scala media (SM), and the scala tympani (ST), by two longitudinal membranes, i.e., Reissner's membrane and the basilar membrane (BM). The vibration of the tympanic membrane and the ossicular chain, which is induced by sound, is transmitted to the BM through the cochlear fluid.

Figure 2 shows a schematic of a cross section of the organ of Corti (OC) of the gerbil. The OC, which sits on the BM, also vibrates in response to the vibration of the BM. When the OC undergoes vibration, it is conventionally thought that the hair bundles of the inner hair cell (IHC) and the outer hair cell (OHC) are deflected due to the shear motion between the tectorial membrane (TM) and the reticular lamina (RL), as shown in Fig. 2. The tip of the hair bundle of the IHC appears to be freestanding, while that of the OHC is embedded in the TM (Engströme and Engström, 1978). The deflection of the hair bundle induces a change of the membrane potential of these cells. As shown in Fig. 2, the deflection in the direction of the tallest hair induces depolarization and that in the direction of the shortest hair induces hyperpolarization. The IHC excites the auditory nerves when its membrane potential is depolarized (Katz and Miledi, 1967) and the OHC contracts and elongates in response to the depolarization and hyperpolarization (Brownell *et al.*, 1985; Kachar *et al.*, 1986; Zenner, 1986; Ashmore, 1987; Santos-Sacchi and Dilger, 1988), respectively. This motility of the OHC generates a force (Hallworth, 1995; Iwasa and Adachi,

1997; Frank *et al.*, 1999) which amplifies the OC vibration, especially at low stimulus intensity. The auditory transduction process is therefore dependent on specific factors of the neural excitation in response to the BM vibration. Knowledge of these factors is important in the investigation of the auditory mechanics and the development of speech processors for cochlear implants.

For an investigation of the phase of the neural excitation relative to the BM motion, considerable research has been conducted. The traditional view of the response of the hair bundle of the IHC to the BM vibration is that the hair bundle of the IHC responds to the velocity of the BM vibration (Pickles, 1988) because the shear force exerted on the hair bundle is thought to be due to the shear flow between the TM and the hair bundle, the so-called Couette flow, as indicated in Fig. 2, this force being proportional to the velocity of the shear flow. As a result, if the maximum velocity of the shear flow is assumed to occur when the velocity of the BM is maximum, the maximum displacement of the tip of the hair bundle of the IHC in the direction of the tallest hair, which induces depolarization, occurs when the velocity of the BM toward the SV is maximum, i.e., when the BM crosses its equilibrium position (zero displacement) from the ST to the SV. Although the phase of the neural excitation relative to the displacement of the hair bundle is influenced by low-pass filtering at the membrane of the IHC, this influence seems to be small if the stimulus frequency is sufficiently lower than the corner frequency of the filter [approximately 200 Hz (Cheatham and Dallos, 1999)]. Therefore, according to this traditional idea, at well below 200 Hz, neural excitation would occur when the velocity of the BM toward the SV is maximum.

Ruggero and Rich (1983) measured the relationship between the neural excitation and the simultaneous BM posi-

[a)]Electronic mail: wada@cc.mech.tohoku.ac.jp
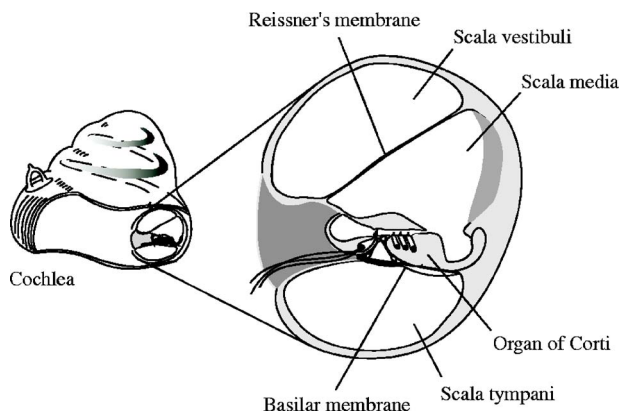
FIG. 1. Schematic of the cross section of the cochlea of the gerbil.

tion, and concluded that neural excitation occurs when the displacement of the BM toward the ST is maximum at the basal turn of the chinchilla cochlea with the characteristic frequency (CF, the most effective stimulus frequency) of 16 kHz for low-frequency stimuli (30 to 100 Hz) at 100 dB SPL. Furthermore, Wada *et al.* (2002) measured the relationship between the neural excitation and the BM motion at the
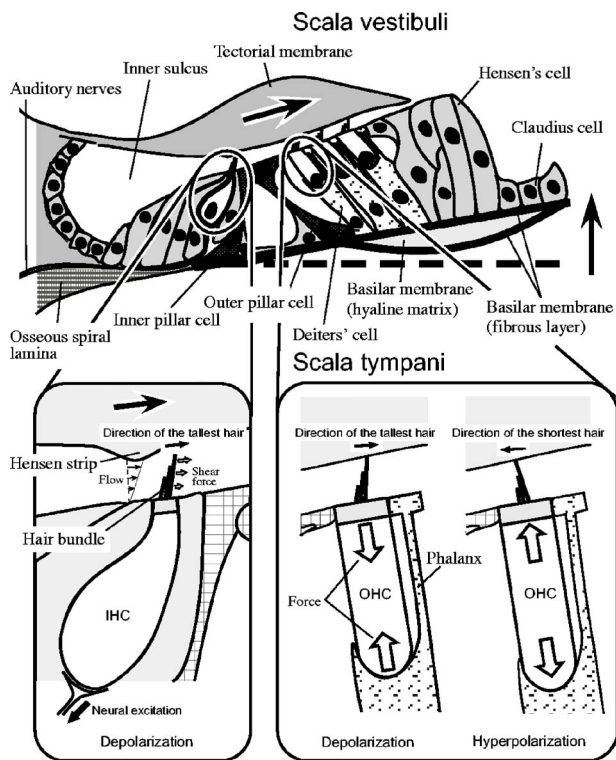
FIG. 2. Structure and mechanoelectrical transduction of the OC of the gerbil. Large arrows on the TM and beside the BM indicate the direction of their displacement. When the BM undergoes oscillatory motion, the hair bundle of the OHC is deflected by the relative displacement between the TM and the RL because its tip is embedded in the TM. On the contrary, according to the traditional idea, the hair bundle of the IHC is assumed to be vibrated by the shear force due to the shear flow between the TM and the hair bundle, and this shear force in the excitatory direction (toward the tallest hair) is thought to become maximum when the velocity of the BM toward the SV is maximum, i.e., the BM crosses its equilibrium position (zero displacement) from the ST to the SV. The deflection of the hair bundle of the IHC causes the excitation of auditory nerves, while the deflection of the hair bundle of the OHC induces the OHC motility that generates the force.

basal turn of the guinea pig cochlea with high CFs (14–22 kHz) to low-frequency stimuli (50 to 2000 Hz) at 100 dB SPL, and demonstrated that neural excitation lags behind the maximum displacement of the BM toward the SV by 20° from 50 to 100 Hz and by 45° at 500 Hz, and leads it by 45° at 2 kHz. These experimental data are therefore different from each other, as well as differing from the above-mentioned conventional idea. Thus, the phase of the neural excitation in relation to the BM motion in the cochlea has remained unclear. Cai *et al.* (2004) constructed a model of the OC by which its dynamic behavior can be analyzed. Although such a micromechanical model is expected to make it possible to investigate the phase relationship between the displacement of the hair bundle of the IHC and that of the BM in detail, they did not conduct this kind of analysis. Steele and Puria (2003) numerically analyzed the force exerted on the hair bundle of the inner hair cell by using local lubrication theory. Their research was a pioneering work, and the three rows of cilia and tip links were taken into account. However, they analyzed the vibration of the OC until 300 Hz, although the CF of their model was higher than 10 kHz. In addition, the mechanism of the force generation was not discussed.

In this study, first, a finite-element model of the OC and finite-element models of the lymph fluid surrounding the OC and of such fluid in the sub-TM space, which is a narrow space between the TM and RL, were constructed. Using these models, the phase of the displacement of the hair bundle of the IHC relative to the BM motion was then calculated. Finally, by considering the phase delay due to the low-pass filtering at the membrane of the IHC, the phase of the neural excitation relative to the BM motion was estimated.

II. MODEL OF THE ORGAN OF CORTI

A. Modification of the previous model

In a previous study (Andoh and Wada, 2004), although the geometry of the OC at the basal turn of the gerbil cochlea, where the CF is approximately 16 kHz, was reproduced based on measurement in the hemicochlea of the gerbil (Edge *et al.*, 1998; Richter *et al.*, 2000), the shapes of the TM, Deiters' cells, Hensen's cells, pillar cells, and the inner sulcus were not realistic. In addition, the homogeneous layer, the so-called hyaline matrix, between the fibrous layers in the BM, was neglected. Furthermore, as the lymph fluid in the sub-TM space was omitted, the dynamic behavior of the hair bundle of the IHC could not be analyzed. Therefore, the following additional features were incorporated in the modified model.

(a) Realistic shapes of the TM, Deiters' cells, Hensen's cells, pillar cells, and the inner sulcus.
(b) Inclusion of the hyaline matrix in the BM.
(c) Modeling of the lymph fluid in the sub-TM space.

Figure 3(a) shows the previous model of the OC, and Fig. 3(b) shows the modified model of the OC. Shapes of the TM, Deiters' cells, Hensen's cells, pillar cells, and the inner sulcus were determined based on the measurement results of

(a)

(b)
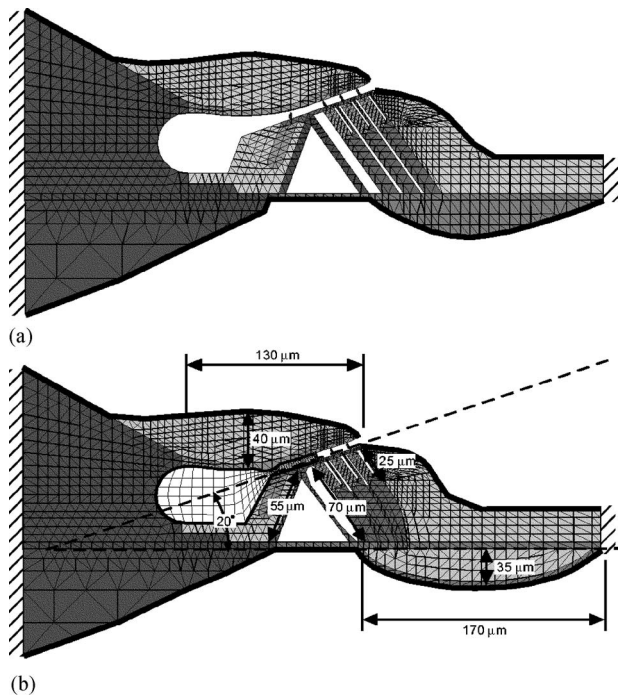
FIG. 3. Model of the OC. (a) Previous model of the OC. (b) Modified model of the OC. Each shade of gray in the model indicates a portion with the same mechanical properties. Thick lines surrounding the model indicate the fluid-structure interface of the model.

FIG. 4. Model of the lymph fluid in the sub-TM space. The sub-TM space is a fluid-filled space below the TM. It was modeled from the inner sulcus to the narrow space between the hair bundle of the IHC and that of the first row of the OHC. The height of the hair bundle of the IHC is 4.3 μm, and its shape is sharply wedged. The number of nodes is 2128 and the number of elements is 1554.

the OC in the hemicochlea of the gerbil (Richter, 2003). Although the constituents of the hyaline matrix in the BM are unknown, its structure is similar to the gel-like structure of the TM. Young's modulus of the hyaline matrix was therefore assumed to be the same as that of the TM. Table I shows all mechanical properties assigned to the models.

The boundary conditions of the models of the OC and the lymph fluid are the same as those in the previous study. They were validated by comparing the experimental data, i.e., the pressure in the ST (Olson, 2001) and the frequency characteristics of the velocity of the BM (Ren and Nuttall,

TABLE I. Mechanical properties assigned to the models of the OC and lymph fluid. The stiffness of the hair bundle relative to the displacement of its tips is 2.5×10^{-3} N/m (Langer *et al.*, 2001). The density and viscosity of the lymph fluid are 1.0×10^3 kg/m^3 and 1.0×10^{-3} Pa·s, respectively, which are equal to those of water.

	Young's modulus (N/m^2)	Poisson's ratio
BM (fibrous layer)	1.0×10^7	0.3
BM (hyaline matrix)	3.0×10^4	0.49
Deiters' cell	1.0×10^7	0.3
Hair bundle	1.0×10^7	0.3
Hensen's cell	5.0×10^3	0.49
IHC	1.0×10^4	0.49
Kimura's membrane	1.0×10^6	0.3
Osseous spiral lamina	2.0×10^{10}	0.3
OHC	1.0×10^4	0.49
Phalanx	1.0×10^7	0.3
Pillar cell	1.0×10^9	0.3
RL	1.0×10^9	0.3
TM	3.0×10^4	0.49

2001), with the numerical results in the previous study (Andoh and Wada, 2004). As the force generated by the OHC motility is not taken into account, the simulation is limited to the passive cochlea, i.e., OHCs are dysfunctional, or to the active cochlea, i.e., OHCs are functional when the level of the input stimulation is greater than 100 dB SPL (Ren and Nuttall, 2001).

B. Lymph fluid in the sub-TM space

1. Model of the sub-TM space

To simulate the dynamic behavior of the hair bundle of the IHC, the pressure around the hair bundle must be analyzed. The model of the sub-TM space was therefore constructed as shown in Fig. 4. The width in the longitudinal direction (normal to the paper) and boundary conditions of the sub-TM model are the same as those of the SV and ST models (Andoh and Wada, 2004), i.e., the width in the longitudinal direction is 48 μm and free-flux boundary conditions are assigned to both longitudinal ends. The distribution of the velocity of the fluid around the hair bundle of the OHC was assumed to be the same as that of the hair bundle itself, due to the fact that its tip is embedded in the TM. In addition, as the hair bundle arranged in a W-shaped formation accounts for 80% of the diameter of the OHC, we assumed that the flow in the gap between the hair bundles of the OHCs in the same row was negligibly small. As a result, the sub-TM space was modeled up to the hair bundle of the first row of OHCs. These assumptions lead to the idea that the lymph fluid beyond the hair bundle of the first row of OHCs does not have an effect on the behavior of the lymph fluid around the hair bundle of the IHC. Although the tip of the hair bundle of the IHC is hemispherical, that of the model employed in this study is sharply wedged due to refinement of the mesh. The geometry of the space between the TM and the tip of the hair bundle of the IHC was determined based on a micrograph taken at the basal turn in the gerbil cochlea (Richter, 2003). The intervals of the mesh in the
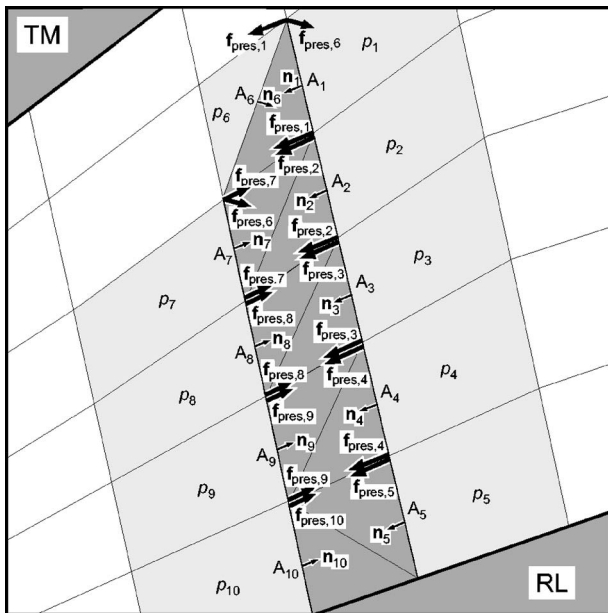
FIG. 5. Force induced by the pressure difference and exerted on the hair bundle. The pressures in the fluid elements (light gray areas) adjacent to the structural elements of the hair bundle were converted to the force vector $\mathbf{f}_{\text{pres},n}$ at each node on the hair bundle. The force vectors at the attachment to the RL were omitted because they have no influence on the dynamic behavior of the hair bundle.

cross section are varied to avoid severe distortion of the cubic shape of the element, and result in 6×37 elements in one cross section, which make it possible to evaluate the pressure around the hair bundle of the IHC. The interval of the mesh in the longitudinal direction is 6 μm, which is the same as that in the SV and ST models.

2. Forces exerted on the hair bundle of the IHC

The pressure difference between the left and right sides of the hair bundle induces force that is exerted on the hair bundle of the IHC. In addition, as the pressure difference between the left and right sides and the relative velocity of the TM in relation to the hair bundle induce fluid flows in the narrow space, shear forces generated by these flows, i.e., Poiseuille flow and Couette flow, are considered to be exerted on the hair bundle. These three types of forces are therefore considered to be exerted on the hair bundle of the IHC.

a. Force induced by the pressure difference Figure 5 shows the force vectors induced by the pressure around the hair bundle and exerted on the hair bundle. The force vector $\mathbf{f}_{\text{pres},n}$ on each node of the structural element of the hair bundle caused by the pressure is obtained by

$$\mathbf{f}_{\text{pres},n} = \frac{1}{2} p_n A_n \mathbf{n}_n, \tag{1}$$

where p_n is the pressure in the fluid element adjacent to the structural element of the hair bundle, A_n is the area of contact between the fluid and structural element, and $\mathbf{n}_n$ is a unit normal vector of the fluid-structure interface. The sum of force vectors over nodes on the hair bundle is the force in-
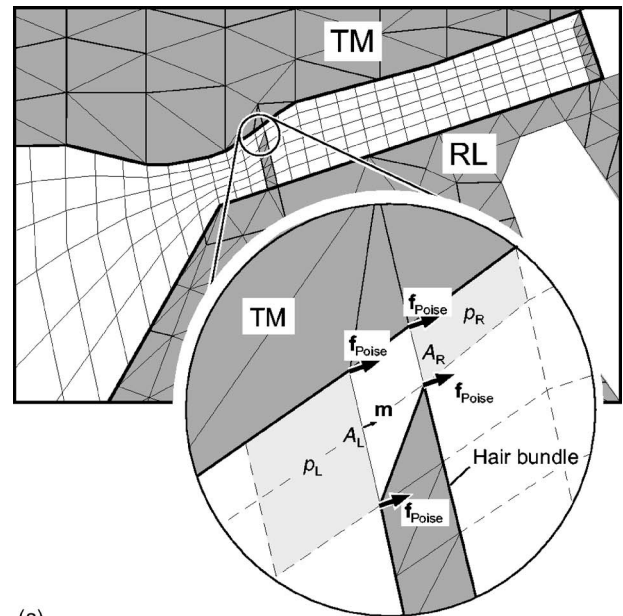
(a)
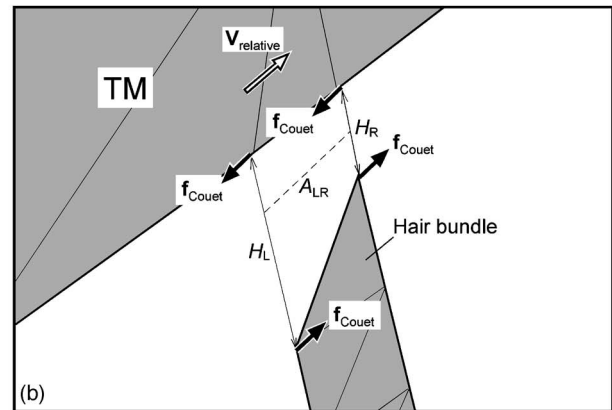
(b)

FIG. 6. Shear forces exerted on the hair bundle. (a) Force induced by Poiseuille flow. Light gray areas indicate the places where the mean pressures were calculated. Poiseuille flow occurs in the gap between the TM and the hair bundle because of the difference between the mean pressure at the left side of the gap p_L and that at the right side of the gap p_R. The force vector induced by Poiseuille flow $\mathbf{f}_{\text{Poise}}$ is in equilibrium with the difference between the product of the pressure and the cross-sectional area at the left side of the gap between the TM and the hair bundle and that at the right side of the gap. (b) Force induced by Couette flow. Couette flow in the gap between the TM and the hair bundle results from the relative velocity V_{relative} of the TM in relation to the hair bundle. The force vector induced by Couette flow $\mathbf{f}_{\text{Couet}}$ is a product of the viscosity and the velocity gradient between the TM and the hair bundle.

duced by the pressure difference between the left and right sides of the hair bundle.

b. Analytical deviation of the shear forces Due to computer memory size limitation, the mesh intervals in the sub-TM model are insufficiently fine for an analysis of the distribution of the fluid velocity in the gap between the TM and the RL, which is the space bounded by the TM, RL, and light gray areas shown in Fig. 6(a). Therefore, the shear forces induced by the fluid flow and exerted on the hair bundle were derived analytically.

Figure 6(a) shows Poiseuille flow-induced shear force exerted on the hair bundle and the TM. The sum of Poiseuille flow-induced shear force exerted on the hair bundle and the

TM was assumed to be the same as the difference between the product of the pressure and the cross-sectional area at the left side of the gap between the TM and the hair bundle and that at the right side. As there are two structural element nodes on each surface facing the gap between the TM and the hair bundle, the force vector $\mathbf{f}_{\mathrm{Poise}}$ on each node of the structural element caused by Poiseuille flow is given by

$$\mathbf{f}_{\mathrm{Poise}} = \frac{1}{4}(p_{\mathrm{L}}A_{\mathrm{L}} - p_{\mathrm{R}}A_{\mathrm{R}})\mathbf{m}, \tag{2}$$

where p_{L} and p_{R} are the mean pressures in the fluid elements adjacent to the left and right sides of the gap between the TM and the hair bundle shown by light gray areas in Fig. 6(a), A_{L} and A_{R} are the cross-sectional areas at the left and right sides of the gap, respectively, and $\mathbf{m}$ is a unit normal vector of the constant pressure line. Figure 6(b) shows Couette flow-induced shear force exerted on the hair bundle and the TM. After Arnell *et al.* (1991), the force vector $\mathbf{f}_{\mathrm{Couet}}$ on each node of the structural element facing the gap between the TM and the hair bundle caused by Couette flow is given by

$$\mathbf{f}_{\mathrm{Couet}} = \frac{1}{2}\mu\frac{\mathbf{V}_{\mathrm{relative}}}{(H_{\mathrm{L}} - H_{\mathrm{R}})}A_{\mathrm{LR}}\ln\frac{H_{\mathrm{L}}}{H_{\mathrm{R}}}, \tag{3}$$

where $\mathbf{V}_{\mathrm{relative}}$ is the relative velocity vector of the TM in relation to the hair bundle, H_{L} and H_{R} are heights at the left and right sides of the gap between the TM and the hair bundle, respectively, A_{LR} is the area of the cross section between the left and right sides of the gap and parallel to $\mathbf{V}_{\mathrm{relative}}$, and μ is the viscosity of the lymph fluid.

   c. Numerical analysis of the shear forces To validate the analytically obtained shear forces exerted on the hair bundle represented by Eqs. (2) and (3), the shear forces were numerically analyzed by using the model of the gap between the TM and the hair bundle of the IHC, as shown in Fig. 7(a). The width in the longitudinal direction (normal to the paper) and boundary conditions of this model are the same as those of the other fluid models, i.e., the width in the longitudinal direction is 48 μm and free-flux boundary conditions are assigned to both longitudinal ends. In the case of Poiseuille flow, the pressures at the left and right sides of the gap between the TM and the hair bundle, i.e., p_{L} and p_{R}, are applied to the elements at the left and right sides of this model [light gray areas in Fig. 7(a)]. In the case of Couette flow, the relative velocity vector $\mathbf{V}_{\mathrm{relative}}$ of the TM in relation to the hair bundle is applied as the boundary condition at the top of this model. The distribution of the velocity vector between the TM and the hair bundle is obtained according to these inputs, and the shear force vector $\mathbf{f}_{\mathrm{shear}}$ induced by Poiseuille flow and Couette flow is then calculated as follows based on the parameters shown in Fig. 7(b):

$$\mathbf{f}_{\mathrm{shear}} = \frac{1}{2}\mu\sum_{n=1}^{16}\frac{\mathbf{v}_n + \mathbf{v}_{n+1}}{h_n + h_{n+1}}A_{\mathrm{top},n}, \tag{4}$$

where $\mathbf{v}_n$ is the velocity vector at nodes proximate to the hair bundle, h_n is the height of the fluid element adjacent to the hair bundle, and $A_{\mathrm{top},n}$ is the area of contact between the fluid element and the hair bundle. By comparing this nu-
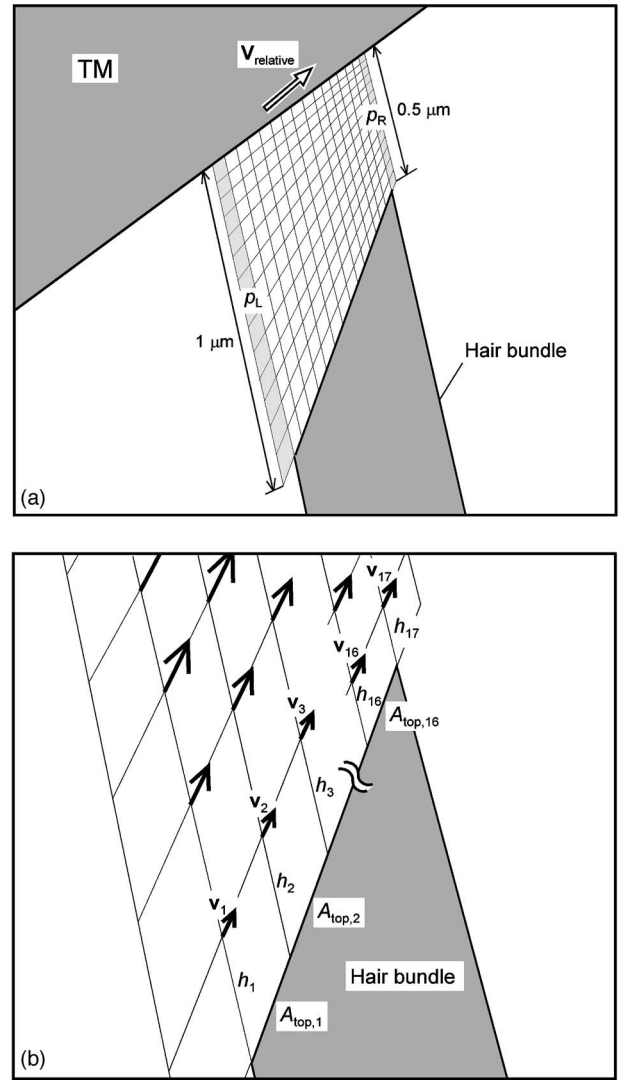
FIG. 7. Numerical analysis of the shear forces exerted on the hair bundle of the IHC. (a) Model of the gap between the TM and the hair bundle of the IHC. The number of nodes is 2040 and the number of elements is 1568. p_{L} and p_{R} are mean pressures at the left and right sides of the gap (light gray areas), respectively, and $\mathbf{V}_{\mathrm{relative}}$ is the relative velocity of the TM in relation to the hair bundle. (b) Magnification of the model around the top of the hair bundle. Arrows indicate the velocity vectors $\mathbf{v}_n$ at nodes proximate to the hair bundle. h_n is the height of the fluid element adjacent to the hair bundle and $A_{\mathrm{top},n}$ is the area of contact between the fluid element and the hair bundle. Based on these velocity vectors, the shear force exerted on the hair bundle is obtained from Eq. (4) for validation of the analytically obtained shear forces represented by Eqs. (2) and (3).

merically obtained shear force with that given by Eqs. (2) and (3), the validity of the analytically obtained shear forces is examined.

3. Shear force exerted on the TM and RL

   In the narrow space between the TM and the RL, Couette flow is assumed to occur because of the shear motion between these two components, and the force caused by this shear motion is considered to be exerted on the TM and the RL. In this case, the force vector $\mathbf{f}_{\mathrm{TM,RL}}$ is given by

Andoh *et al.*: Neural excitation vs basilar membrane motion
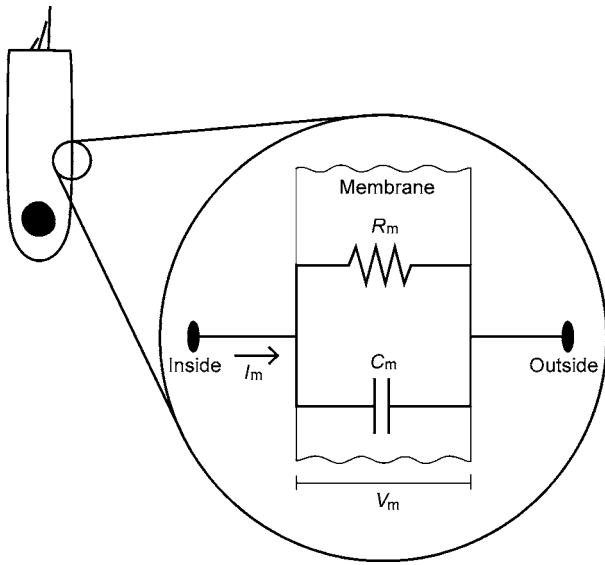
FIG. 8. Electrical circuit model of the membrane of the IHC. R_m and C_m are the membrane resistance and membrane capacitance, respectively. I_m is the current across the membrane and V_m is the membrane potential. The membrane impedance consists of a parallel connection of R_m and C_m. The membrane behaves as a low-pass filter on the membrane potential as indicated by Eq. (6).

FIG. 9. A schematic of the model of the OC showing the stimulation of the model and the method of measurement. Hatching under the BM indicates the site where a sinusoidal pressure of 2 Pa is applied to the model of the OC as an initial pressure. The white circle on the BM indicates the point of displacement measurement. Positive displacement of the BM and that of the hair bundle were defined as that toward the ST and that in the excitatory direction (toward the tallest hair), respectively. A positive driving force exerted on the hair bundle was defined as that in the excitatory direction of the hair bundle.

D. Numerical procedure

Using the models mentioned above, the dynamic behavior of the OC, including the hair bundle of the IHC, was obtained from 10 Hz to 30 kHz when a sinusoidal pressure of 2 Pa was applied to the bottom of the OC, as shown in Fig. 9. The basic procedure for coupling the fluid and structure models is the same as that in the previous study (Andoh and Wada, 2004), but the forces exerted on the hair bundle of the IHC are newly taken into consideration. The revised procedure is as follows: First, the stimulus frequency is fixed to a specific frequency. As shown in Fig. 10, in time Step 1, initial pressure p_{INT}^1 is applied to the model of the OC and the velocity vector of the OC $\partial \mathbf{u}_s^1 / \partial t$ is obtained, where p_{INT} is the pressure initially applied to the model of the OC. Then, applying this obtained velocity vector $\partial \mathbf{u}_s^1 / \partial t$ to the SV, ST, and sub-TM models as a fluid velocity vector $\mathbf{v}_f^0$ over a fluid-structure interface, the pressure p_{OC}^1 in the SV, the ST and the sub-TM space caused by the movement of the OC is obtained. In time Step 2, the force vector $\mathbf{f}_{pres,n}^1$ induced by the pressure difference between the left and right sides of the hair bundle and Poiseuille flow-induced force vector $\mathbf{f}_{Poise}^1$ are given by substituting the pressure p_{OC}^1 into Eqs. (1) and (2), and Couette flow-induced force vector $\mathbf{f}_{Couet}^1$ is given by Eq. (3) by calculating the relative velocity vector $\mathbf{V}_{relative}^1$ of the TM in relation to the hair bundle of the IHC. The force vector $\mathbf{f}_{TM,RL}^1$ caused by the shear motion between the TM and the RL and exerted on them is given by substituting the relative velocity vector $\mathbf{V}_{RELATIVE}^1$ of the TM in relation to the RL into Eq. (5). The previously obtained pressure p_{OC}^1, the initial pressure p_{INT}^2 in time Step 2, the force vectors on the hair bundle, i.e., $\mathbf{f}_{pres,n}^1$, $\mathbf{f}_{Poise}^1$, and $\mathbf{f}_{Couet}^1$, and the force vector $\mathbf{f}_{TM,RL}^1$ exerted on the TM and the RL are applied to the model of the OC. By repeating the above procedure, the displacement of the OC, three types of forces exerted on the hair bundle and the displacement of the hair bundle at a

$$\mathbf{f}_{TM,RL} = \mu \frac{\mathbf{V}_{RELATIVE}}{h} A_{TM,RL}, \tag{5}$$

where $\mathbf{V}_{RELATIVE}$ is the relative velocity vector of the TM in relation to the RL, h is the height of the narrow space between the TM and the RL, and $A_{TM,RL}$ represents the area of the TM and that of the RL facing the lymph fluid.

C. Low-pass filtering at the membrane of the IHC

Although the phase of the neural excitation relative to the BM position has not been well understood, in this study, it is assumed to be as follows: the low-pass filtering at the membrane of the IHC is effective for the entire frequency and intensity range. As shown in Fig. 8, the electrical impedance of the membrane Z_m is represented as the resistance R_m and capacitance C_m in parallel (Russell and Sellick, 1983). In this case, the magnitude $|V_m|$ and the phase θ of the membrane potential relative to the current I_m across the membrane are given by

$$|V_m| = \frac{R_m}{\sqrt{1 + \left(\frac{\omega}{\omega_0}\right)^2}} |I_m|, \tag{6}$$

$$\theta = -\arctan\left(\frac{\omega}{\omega_0}\right), \tag{7}$$

where ω is the angular frequency of the stimulus and $\omega_0 = 1/R_m C_m$ represents the corner frequency, which is assumed to be 200 Hz (Cheatham and Dallos, 1999). Equation (6) represents the behavior of the membrane as a low-pass filter of the membrane potential.

FIG. 10. Scheme of the fluid-structure interaction using the staggered approach. In time Step 1, initial pressure p_{INT}^1 is applied to the model of the OC and the velocity vector of the OC $\partial \mathbf{u}_s^1 / \partial t$ is obtained, where p_{INT} is an applied sinusoidal pressure of 2 Pa at the bottom of the BM, as shown in Fig. 9. Then, applying this obtained velocity vector $\partial \mathbf{u}_s^1 / \partial t$ to the SV, ST, and sub-TM models as a fluid velocity vector $\mathbf{v}_f^0$ over a fluid-structure interface, the pressure p_{OC}^1 in each scala, and sub-TM space caused by the movement of the OC is obtained at the same time step. In time Step 2, this obtained pressure p_{OC}^1, the initial pressure p_{INT}^2 in time Step 2, the force vectors exerted on the hair bundle, i.e., the force vector $\mathbf{f}_{pres,n}^1$ induced by the pressure difference between the left and right sides of the hair bundle, Poiseuille flow-induced force vector $\mathbf{f}_{Poise}^1$ and Couette flow-induced force vector $\mathbf{f}_{Couet}^1$, and the force vector $\mathbf{f}_{TM,RL}^1$ exerted on the TM and the RL are applied to the model of the OC. By repeating the above procedure, the time history of the displacement of the OC and that of the hair bundle of the IHC are obtained.

specific frequency are obtained. This analysis is conducted for stimulus frequencies from 10 Hz to 30 kHz to obtain results for this frequency range.

A numerical analysis of the shear forces exerted on the hair bundle of the IHC is also conducted for each specific frequency. The relative velocity vector $\mathbf{V}_{relative}^1$ of the TM in relation to the hair bundle or the pressures at the left side p_L and right side p_R of the gap between the TM and the hair bundle are applied to the model shown in Fig. 7(a). A Navier-Stokes equation is used to analyze the dynamic behavior of the lymph fluid for these inputs, and it is solved by a Marker-and-Cell method (Harlow and Welch, 1965). The velocity vectors $\mathbf{v}_n^1$ in this space are then obtained. By substituting the obtained velocity vectors $\mathbf{v}_n^1$ into Eq. (4), the shear force $\mathbf{f}_{shear}^1$ induced by Couette flow or Poiseuille flow at a specific frequency is obtained. This analysis is also conducted for stimulus frequencies from 10 Hz to 30 kHz. Consequently, the shear force for this frequency range is obtained.

III. RESULTS

A. Validation of the analytically obtained shear forces exerted on the hair bundle

Analytically obtained shear forces from Eqs. (2) and (3) are compared with those obtained numerically by using the model shown in Fig. 7(a) and from Eq. (4). Figure 11 shows
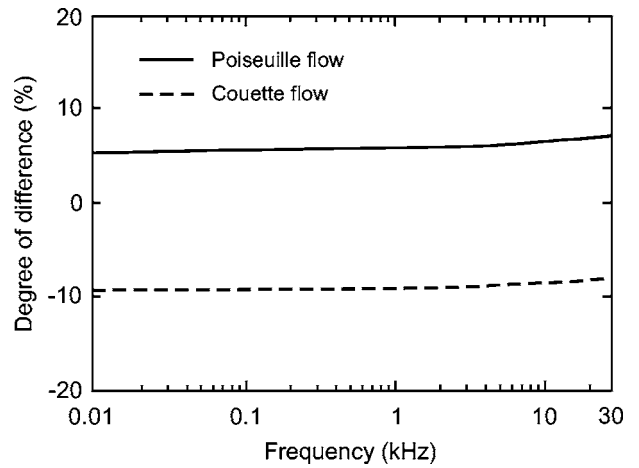
FIG. 11. Degree of difference between the analytically obtained shear forces from Eqs. (2) and (3) and those obtained numerically from Eq. (4) when a pressure of 2 Pa was applied to the bottom of the BM. The analytically obtained shear forces are in the 10% range of those obtained numerically.

the degree of difference between the analytically obtained shear forces and those obtained numerically when a pressure of 2 Pa was applied to the bottom of the BM, as shown in Fig. 9. As the degree of difference is in the 10% range for the entire frequency range, the analytically obtained results are used hereafter.

B. Driving force exerted on the hair bundle and its displacement versus BM motion

Figure 12 shows the frequency characteristics of the displacement of the BM when a pressure of 2 Pa was applied to the bottom of the BM. The displacement of the BM increases with increasing frequency toward the CF (16 kHz), and then decreases. The frequency characteristics of the magnitude of the forces exerted on the hair bundle are shown in Fig. 13. The force induced by Couette flow is larger than the other forces below 1 kHz. This relationship changes around 1 kHz, and the force induced by the pressure difference between the left and right sides of the hair bundle is larger than those induced by Poiseuille flow and Couette flow above 1 kHz.
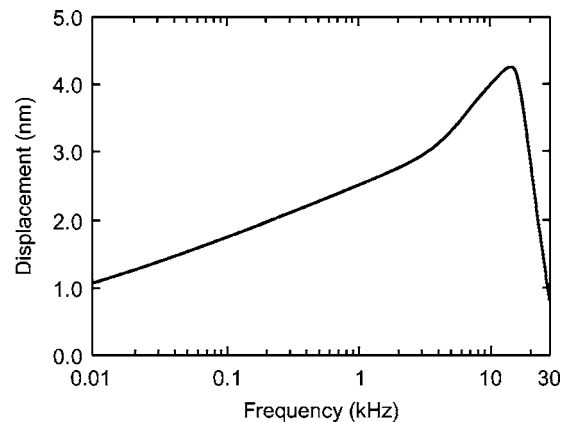
FIG. 12. Frequency characteristics of the displacement of the BM when a pressure of 2 Pa was applied to the bottom of the BM. The displacement of the BM was analyzed as the transversal displacement of the BM at the point indicated in Fig. 9 because the radial displacement of the BM was substantially small relative to the transversal one.
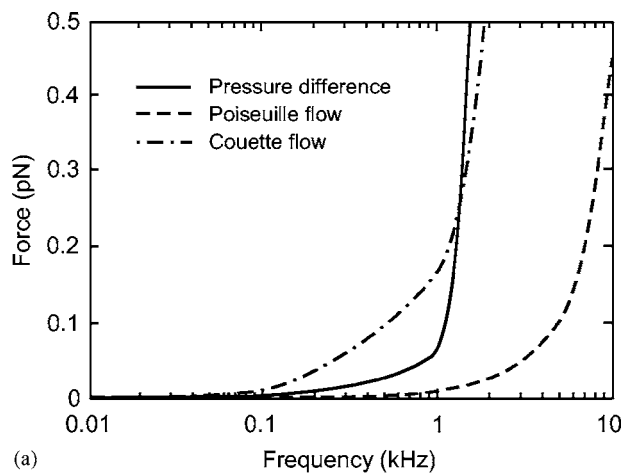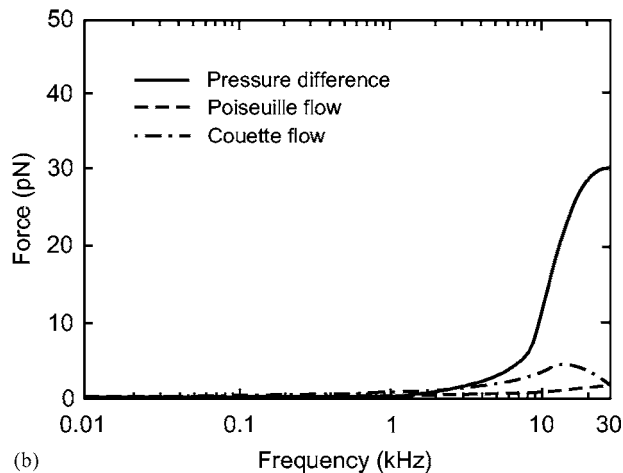
FIG. 13. Magnitudes of the three types of forces exerted on the hair bundle of the IHC versus frequency when a pressure of 2 Pa was applied to the bottom of the BM. (a) Forces plotted up to 10 kHz. (b) Forces for the entire frequency range. Below 1 kHz, the force induced by Couette flow is dominant over the forces exerted on the hair bundle. On the other hand, above 1 kHz, the force induced by the pressure difference between the left and right sides of the hair bundle is larger than the forces induced by Poiseuille flow and Couette flow.

FIG. 14. Frequency characteristics of the driving force exerted on the hair bundle of the IHC and the displacement of the hair bundle when a pressure of 2 Pa was applied to the bottom of the BM. (a) The amplitude of the driving force and the displacement of the hair bundle. (b) Phases of the driving force and the displacement of the hair bundle relative to the displacement of the BM. The zero phases of the driving force and the displacement of the hair bundle relative to the displacement of the BM mean that the driving force and the displacement of the hair bundle are maximum in the excitatory direction when the BM exhibits a maximum displacement toward the ST. The increase of these phases means that their responses lag behind the BM motion. The right ordinate shows positions of the BM when the driving force or the displacement of the hair bundle is maximum in the excitatory direction. ST and SV mean positions of the BM when the BM exhibits a maximum displacement toward the ST and the BM exhibits a maximum velocity toward the SV, i.e., the BM crosses its equilibrium position (zero displacement) from the ST to the SV, respectively. As shown in Fig. 9, the displacement of the hair bundle d was analyzed based on the relative displacement of the tip of hair bundle in the radial direction x and that in the transversal direction y to those of basal end of the hair bundle on the apical surface of the IHC, i.e., $d=\sqrt{x^2+y^2}$. The phase of the driving force according to the traditional idea is plotted for reference by the dotted line.

Figure 14(a) shows the frequency characteristics of the magnitude of the driving force exerted on the hair bundle (solid line), which is a sum of the three types of forces exerted on the hair bundle, as well as those of the displacement of the hair bundle (dashed line) when a pressure of 2 Pa was applied to the bottom of the BM. The increase of the driving force is approximately proportional to the second power of the stimulus frequency. By contrast, the displacement of the hair bundle reaches a maximum at 12 kHz and then decreases. The frequency characteristics of their phases relative to the displacement of the BM are shown in Fig. 14(b). The right ordinate shows positions of the BM when the driving force or the displacement of the hair bundle is maximum in the excitatory direction. The driving force in the excitatory direction (toward the tallest hair and to the right, as indicated in Fig. 9) takes a maximum when the BM exhibits a maximum velocity toward the SV below 1 kHz. It then shows a significant phase advance relative to the BM motion around 1 kHz and takes a maximum when the BM exhibits a maximum displacement toward the ST above 1 kHz. The maxi-

mum displacement of the hair bundle in the excitatory direction occurs when the BM exhibits a maximum velocity toward the SV at 10 Hz, and according to the phase shift of the driving force, it then shows a significant phase advance relative to the BM motion up to 2 kHz, where the maximum displacement of the hair bundle in the excitatory direction lags by 13° behind the maximum displacement of the BM toward the ST. From 2 kHz, unlike the driving force, it shows a phase delay relative to the BM motion with increasing frequency, and becomes 90° at 12 kHz.

C. Neural excitation versus BM motion

Figure 15 shows the phase of the membrane potential of the IHC relative to the current across the membrane versus the stimulus frequency derived from Eq. (7). Assuming that this phase is equivalent to that of the neural excitation relative to the displacement of the hair bundle, this result implies
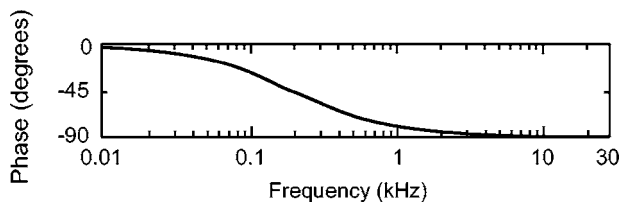
FIG. 15. Phase of the membrane potential of the IHC relative to the current across the membrane versus stimulus frequency derived from Eq. (7). The corner frequency $\omega_0 = 1/R_m C_m$ was taken to be 200 Hz.

that the neural excitation has a slight phase delay relative to the displacement of the hair bundle until 50 Hz, e.g., 3° at 10 Hz and 14° at 50 Hz. This phase delay becomes 45° at the corner frequency of 200 Hz, and then gradually approaches 90° with increasing frequency.

Figure 16 shows the frequency characteristics of the phases of the displacement of the hair bundle [dashed line: the same as the dashed line in Fig. 14(b)] and the neural excitation (solid line) relative to the displacement of the BM. The phase of the neural excitation (solid line) was obtained by adding the phase delay of the membrane potential due to the low-pass filtering at the membrane of the IHC (shown in Fig. 15) to the phase of the displacement of the hair bundle (dashed line). The right ordinate shows positions of the BM when the displacement of the hair bundle is maximum in the excitatory direction or when neural excitation occurs. Neural excitation occurs when the BM exhibits a maximum velocity toward the SV at 10 Hz. It shows a phase delay relative to
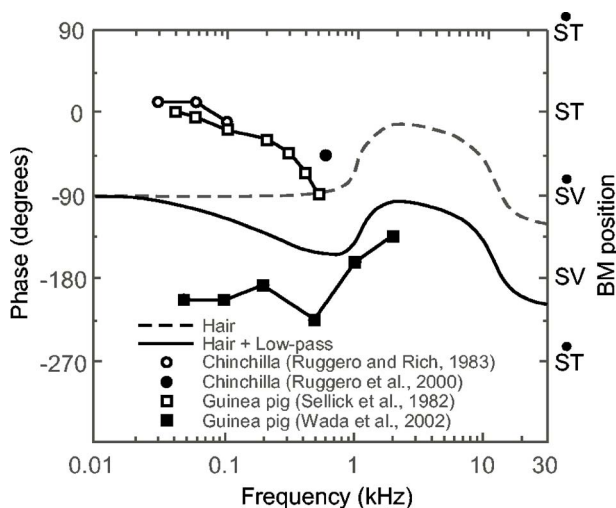
FIG. 16. Frequency characteristics of the phases of the displacement of the hair bundle (dashed line) and the neural excitation (solid line) relative to the displacement of the BM. The dashed line is the same as that in Fig. 14(b). The right ordinate shows positions of the BM when the displacement of the hair bundle is maximum in the excitatory direction or when neural excitation occurs. ST, ṠT, ṠV, and SV mean positions of the BM when the BM exhibits a maximum velocity toward the ST, the BM exhibits a maximum displacement toward the ST, the BM exhibits a maximum velocity toward the SV, and the BM exhibits a maximum displacement toward the SV, respectively. The phase of the neural excitation (solid line) was obtained by adding the phase delay of the membrane potential due to the low-pass filtering at the membrane of the IHC (shown in Fig. 15) to the phase of the displacement of the hair bundle (dashed line). Other data were taken from the literature for purposes of comparison, and sources are indicated in the legend.

the BM motion with increasing frequency up to 800 Hz and then exhibits a phase advance until the frequency reaches 2 kHz, where the neural excitation lags by 7° behind the maximum velocity of the BM toward the SV. From 2 kHz, neural excitation again shows a phase delay with increasing frequency, and at the characteristic frequency of 16 kHz, neural excitation lags by 18° behind the maximum displacement of the BM toward the SV.

IV. DISCUSSION

A. Source of the driving force exerted on the hair bundle

The traditional idea regarding the mechanism of the hair bundle oscillation in relation to the OC vibration is shown in Fig. 2. When the OC undergoes vibration, shear motion occurs between the tectorial membrane (TM) and the reticular lamina (RL). This shear motion causes Couette flow in the sub-TM space. As indicated in Eq. (3), the driving force exerted on the hair bundle due to Couette flow becomes maximum when the relative velocity of the TM in relation to the hair bundle is maximum. As a result, if the maximum relative velocity is assumed to occur when the velocity of the BM is maximum, the maximum driving force exerted on the hair bundle in the excitatory direction occurs when the velocity of the BM is maximum toward the SV at any frequency, as shown by the dotted line in Fig. 14(b). However, the solid line in Fig. 14(b) shows that the driving force in the excitatory direction becomes maximum when the BM exhibits a maximum displacement toward the ST above 1 kHz, although the numerical result in this study shows the same tendency of the driving force as that in the traditional idea below 1 kHz. This result implies that the hair bundle is driven by a different mechanism from that posited by the traditional idea.

In this study, the forces induced by the pressure difference between the left and right sides of the hair bundle and by Poiseuille flow were considered in addition to that induced by Couette flow. As shown in Fig. 13, Couette flow-induced force, which is caused by the velocity difference between the TM and the hair bundle, is larger than the other forces exerted on the hair bundle below 1 kHz. Therefore, similar to the traditional idea, the maximum driving force in the excitatory direction occurs when the velocity of the BM is maximum toward the SV below 1 kHz. By contrast, above 1 kHz, the force induced by the pressure difference is larger than the forces induced by Poiseuille flow and Couette flow. To discuss the phase of the force induced by the pressure difference relative to the BM motion, these frequency characteristics should be considered.

The pressure in an incompressible viscous fluid flow is described by the Navier-Stokes equation as follows:

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v} \cdot \nabla)\mathbf{v} + \nabla p - \mu \Delta \mathbf{v} = 0, \tag{8}$$

where $\mathbf{v}$ is the velocity vector, p is the pressure, ρ is the density, μ is the viscosity, t is the time, and the gradient operator ∇ and the Laplacian operator Δ are defined in the following form:

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right), \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{9}$$

Equation (8) means that the pressure gradient ∇p is determined by the product of the density and the acceleration (first term), the product of the density and the velocity gradient and velocity, i.e., the momentum change (second term), and the product of the viscosity and the Laplacian of the velocity, i.e., the shear stress (fourth term). Assuming that the vibration of the OC is sinusoidal, the acceleration is 180° out of phase with the displacement. Therefore, the first term in Eq. (8) is regarded as a displacement-dependent term, and in Eq. (8), it means that the pressure is proportional to the displacement. Substituting the value of the fluid velocity (17 μm/s) around the hair bundle at 1 kHz, which was obtained from FEM analysis, into Eq. (8), the pressures in the space between the TM and the RL at 1 kHz due to the first term (displacement), the second term (momentum), and the fourth term (viscosity) were determined to be 2.2 mPa, 0.33 μPa, and 3.4 mPa, respectively. If the amplitude of the fluid velocity is assumed to be constant for any frequency, only the first term in Eq. (8) increases with increasing frequency because only this term is differentiated by time. These facts mean that, above 1 kHz, the pressure around the hair bundle is dominated by the displacement of the fluid, and the driving force is therefore synchronized with the displacement of the fluid.

Figure 17 shows the motion of the structures around the hair bundle and displacements of the fluid in the space to the right (cross-hatched area) and that to the left (hatched area) of the hair bundle. In the space to the right of the hair bundle (cross-hatched area), differences were found between the inward and outward displacements of the fluid when the BM is displaced. As the inward displacement of the fluid is defined as being positive, the sum of the inward and outward displacements of the fluid in the cross-hatched area takes maximum positive and negative values when the displacements of the BM are maximum toward the SV and the ST, respectively. The cause of the difference between the inward and outward displacements of the fluid is the volume change of the fluid above the cross-hatched area, which is indicated by the thick line in Fig. 17, due to the change of the distance between the TM and the RL. The increase of this volume results in the positive value of the sum of the inward and outward displacements of the fluid in the cross-hatched area, and *vice versa*. As mentioned in the previous paragraph, the pressure is proportional to the displacement of the fluid above 1 kHz so that the maximum positive and negative pressures in the cross-hatched area occur when the displacements of the BM are maximum toward the SV and the ST, respectively. On the other hand, in the space to the left of the hair bundle (the hatched area in Fig. 17), the pressure exerted on the hair bundle does not have a profound effect on the behavior of the hair bundle because the sum of the inward and outward displacements of the fluid is less than 14% of that in the right space (cross-hatched area) when the displacement of the BM is maximum. Consequently, above 1 kHz, the maximum driving force exerted on the hair bundle in the excitatory direction occurs when the BM exhibits a
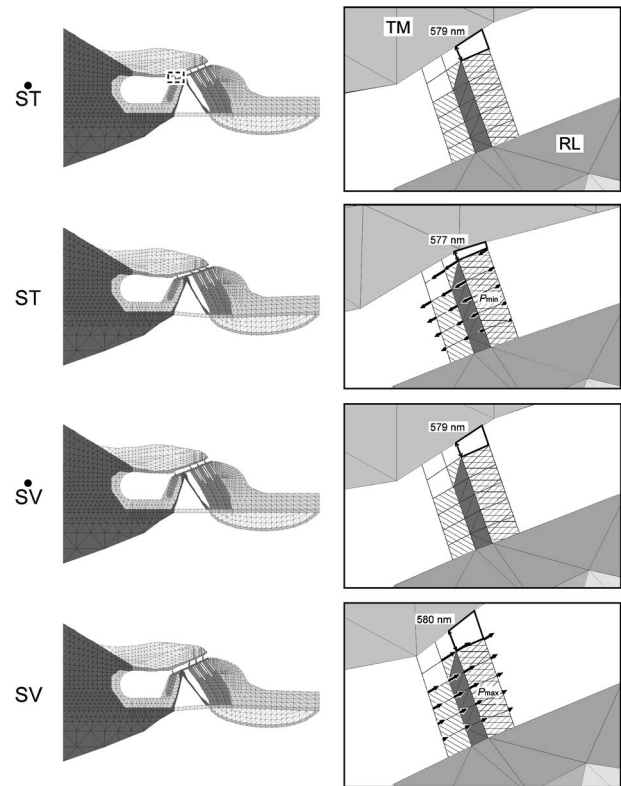
FIG. 17. Motion of the structures around the hair bundle and displacements of the fluid in the space to the right (cross-hatched area) and that to the left (hatched area) of the hair bundle when a pressure of 2 Pa was applied to the bottom of the BM at 1 kHz. ST, $\dot{\text{ST}}$, SV, and $\dot{\text{SV}}$ indicate positions of the BM as in Fig. 16. The images on the left are the overall appearance of the model of the OC at each position of the BM and are displayed for the sake of facilitating comprehension. Those on the right are enlarged images around the hair bundle of the IHC corresponding to the dashed square in the top image on the left. Arrows indicate displacements of the fluid. However, they are not displayed when the BM exhibits maximum velocities because they are negligibly small compared with those when the BM exhibits maximum displacements. The thick line surrounding the space above the cross-hatched area is drawn on the model for clarity of its volume change caused by the change of the vertical distance between the TM and the RL, i.e., the minimum and maximum value is 577 and 580 nm, respectively. P_{max} and P_{min} mean the averaged maximum positive and negative pressures at the space to the right of the hair bundle (cross-hatched area), respectively. Displacements of the left and right images are magnified 2000 and 1000 times, respectively.

maximum displacement toward the ST. For the above-mentioned reasons, the phase relationship between the driving force and the displacement of the BM significantly changes around 1 kHz.

The phase of the displacement of the hair bundle is shown by the dashed line in Fig. 14(b). In a frequency range below 1 kHz, the displacement of the hair bundle follows the driving force. However, from 1 kHz, the displacement of the hair bundle gradually lags behind the driving force with increasing frequency, and its phase delay becomes 90° at 12 kHz. Assuming that the hair bundle is a simple mass-spring system and by considering the added mass of fluid to the hair bundle, the resonance frequency of the hair bundle f_{res} is analytically given by

Andoh *et al.*: Neural excitation vs basilar membrane motion   1563

$$f_{\text{res}} = \frac{1}{2\pi}\sqrt{\frac{k_{\text{hair}}}{m_{\text{hair}} + m_{\text{fluid}}}}, \tag{10}$$

where k_{hair} and m_{hair} are the stiffness and mass of the hair bundle, respectively, and m_{fluid} is the added mass of the fluid. k_{hair} and m_{hair} assigned to the model of the OC are 2.5×10^{-3} N/m (Langer *et al.*, 2001) and 24 pg. If the hair bundle is assumed to be a simple cylindrical structure, the added mass of the fluid is a product of the volume of the hair bundle and the density of the fluid, and so is assumed to be 20 pg. In this case, the resonance frequency of the hair bundle obtained from Eq. (10) is 12.1 kHz. Therefore, the phase delay of the displacement of the hair bundle relative to the driving force appears to occur due to the resonance process of the hair bundle. As this phase shift directly changes the timing of the depolarization of the membrane potential of the IHC, the resonance frequency of the hair bundle, as well as the phase of the driving force, is highly important when investigating the phase of the neural excitation relative to the BM motion.

Cai *et al.* (2004) constructed a model of the OC by which its dynamic behavior could be analyzed. As they assumed that the height of the gap between the TM and the RL was constant, the pressure caused by the volume change in the sub-TM space was neglected. By contrast, Hu *et al.* (1999) measured the vibration trajectories of measured points on the TM and the top of the OHCs by applying an oscillatory force via a glass paddle to the OC and reported that the direction of the vibration at the TM is approximately vertical whereas that at the top of the OHCs is horizontal. As shown in Fig. 17, the geometrical configuration of the sub-TM space, which consists of the TM, the RL, and the hair bundle of the OHC, clearly demonstrates that the distance between the TM and the tip of the hair bundle of the IHC is 579 nm when the displacement of the BM is zero and that it varies from 577 to 580 nm depending on the displacement of the BM. This causes the volume change of the sub-TM space, which must be considered when analyzing the dynamic behavior of the hair bundle of the IHC.

B. Comparison with experimental data

As the motility of the OHC was not included in the present model, and as the OC at the basal turn of the cochlea was modeled, experiments conducted at a high stimulus level and at the basal turn are appropriate for purposes of comparison. A comparison of the phase of the neural excitation relative to the BM position is presented in Fig. 16. As the data shown in the figures (Ruggero and Rich, 1983; Ruggero *et al.*, 2000; Sellick *et al.*, 1982; Wada *et al.*, 2002) had already been corrected by subtracting synaptic delay and neural travel times, i.e., 1 ms delay, from the experimental data, our numerically obtained results were compared with their data without correction. Ruggero and Rich (1983) measured the relationship between the neural excitation and the BM position at the basal turn of the chinchilla cochlea with a CF of 16 kHz to low-frequency stimuli (30 to 100 Hz) at 100 dB SPL, and concluded that the neurons from IHCs may fire in response to the displacement of the BM toward the ST. Sel-

lick *et al.* (1982) measured the neural excitation at the basal turn of the guinea pig cochlea with CFs between 15 and 18 kHz to low-frequency stimuli (40 to 500 Hz) at 100 dB SPL, and demonstrated that the excitation of the IHC was synchronous with the maximum displacement of the BM toward the ST at 40 Hz and with the maximum velocity of the BM toward the SV at 500 Hz. Although these two experimental data show the phase delay of the neural excitation likely induced by the low-pass filtering at the membrane of the IHC as considered in the present study, they are different from our results by approximately 90°. Wada *et al.* (2002) measured the relationship between the neural excitation and the BM vibration at the basal turn of the guinea pig cochlea with high CFs (14–22 kHz) to low-frequency stimuli (50 to 2000 Hz) at 100 dB SPL and reported that neural excitation lags by 20° behind the maximum displacement of the BM toward the SV until 100 Hz and shows a phase delay relative to the BM motion with increasing frequency up to 500 Hz, where it lags by 45° behind the maximum displacement of the BM toward the SV. Neural excitation then shows a phase advance with increasing frequency and leads the maximum displacement of the BM toward the SV by 45° at 2 kHz. The phase advance with increasing frequency from 500 Hz in their data is similar to the present findings; however, the phase values are different from our results, especially at low frequencies. Although there is a hypothesis that at low frequencies the membrane potential of the IHC is modulated by the extra-cellular potential, which is the potential caused by the electrical activities of OHCs (Cheatham and Dallos, 1999), qualitative and quantitative influences of this extra-cellular potential on the membrane potential of the IHC are still unknown. A full understanding of the phase of the neural excitation relative to the BM vibration awaits further empirical investigations or refinement of our model in order to include an electrical field analysis.

V. CONCLUSIONS

In this study, first, a finite-element model of the OC and models of the lymph fluid surrounding the OC and in the sub-TM space were constructed. Using these models, the phase of the displacement of the hair bundle of the IHC relative to the BM motion was then calculated. Finally, by considering the phase delay due to the low-pass filtering at the membrane of the IHC, the phase of the neural excitation relative to the BM motion was estimated. Conclusions are as follows.

(1) Unlike the conventional idea, the phase of the driving force relative to the BM motion shows a significant phase shift around 1 kHz, i.e., although the maximum driving force in the excitatory direction occurs when the BM exhibits a maximum velocity toward the SV below 1 kHz, it also occurs when the BM exhibits a maximum displacement toward the ST above 1 kHz. This is because the dominant force exerted on the hair bundle changes from a velocity-dependent Couette flow-induced force to a displacement-dependent force induced by the pressure difference around 1 kHz.

(2) Neural excitation occurs when the BM exhibits a maximum velocity toward the SV at 10 Hz and shows a phase delay relative to the BM motion with increasing frequency up to 800 Hz. It then shows a phase advance until the frequency reaches 2 kHz, at which point the neural excitation lags by 7° behind the maximum velocity of the BM toward the SV. From 2 kHz, neural excitation again shows a phase delay with increasing frequency due to the resonance of the hair bundle, and at the characteristic frequency of 16 kHz, neural excitation lags by 18° behind the maximum displacement of the BM toward the SV.

ACKNOWLEDGMENTS

Andoh, M., and Wada, H. (**2004**). "Prediction of the characteristics of two types of pressure waves in the cochlea: Theoretical considerations," J. Acoust. Soc. Am. **116**, 417–425.

Arnell, R. D., Davis, P. B., Halling, J., and Whomes, T. L. (**1991**). *Tribology: Principles and Design Applications* (Macmillan, London).

Ashmore, J. F. (**1987**). "A fast motile response in guinea-pig outer hair cells: The cellular basis of the cochlear amplifier," J. Physiol. (London) **388**, 323–347.

Brownell, W. E., Bader, D., and Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.

Cai, H., Shoelson, B., and Chadwick, R. S. (**2004**). "Evidence of tectorial membrane radial motion in a propagating mode of a complex cochlear model," Proc. Natl. Acad. Sci. U.S.A. **101**, 6243–6248.

Cheatham, M. A., and Dallos, P. (**1999**). "Response phase: A view from the inner hair cell," J. Acoust. Soc. Am. **105**, 799–810.

Edge, R. M., Evans, B. N., Pearce, M., Richter, C.-P., Hu, X., and Dallos, P. (**1998**). "Morphology of the unfixed cochlea," Hear. Res. **124**, 1–16.

Engströme, H., and Engström, B. (**1978**). "Structure of hairs on cochlear sensory cells," Hear. Res. **1**, 49–66.

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.

Hallworth, R. (**1995**). "Passive compliance and active force generation in the guinea pig outer hair cell," J. Neurophysiol. **74**, 2319–2329.

Harlow, F. H., and Welch, J. E., (**1965**). "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface," Phys. Fluids **8**, 2182–2189.

Hu, X., Evans, B. N., and Dallos, P. (**1999**). "Direct visualization of organ of Corti kinematics in a hemicochlea," J. Neurophysiol. **82**, 2798–2807.

Iwasa, K. H., and Adachi, M. (**1997**). "Force generation in the outer hair cell of the cochlea," Biophys. J. **73**, 546–555.

Kachar, B., Brownell, W. E., Altschuler, R., and Fex, J. (**1986**). "Electrokinetic shape changes of cochlear outer hair cells," Nature (London) **322**, 365–368.

Katz, B., and Miledi, R. (**1967**). "A study of synaptic transmission in the absence of nerve impulses," J. Physiol. (London) **192**, 407–436.

Langer, M. G., Fink, S., Koitschev A., Rexhausen, U., Hober, J. K. H., and Ruppersberg, J. P. (**2001**). "Lateral mechanical coupling of stereocilia in cochlear hair bundles," Biophys. J. **80**, 2608–2621.

Olson, E. S. (**2001**). "Intracochlear pressure measurements related to cochlear tuning," J. Acoust. Soc. Am. **110**, 349–367.

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing*, 2nd ed. (Academic Press, London).

Ren, T., and Nuttall, A. L. (**2001**). "Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea," Hear. Res. **151**, 48–60.

Richter, C.-P. (**2003**) (private communication).

Richter, C.-P., Edge, R. M., He, D. Z. Z., and Dallos, P. (**2000**). "Development of the gerbil inner ear observed in the hemicochlea," J. Assoc. Res. Otolaryngol. **1**, 195–210.

Ruggero, M. A., Narayan, S. S., Temchin, A. N., and Recio, A. (**2000**). "Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla," Proc. Natl. Acad. Sci. U.S.A. **97**, 11744–11750.

Ruggero, M. A., and Rich, N. C. (**1983**). "Chinchilla auditory-nerve response to low-frequency tones," J. Acoust. Soc. Am. **73**, 2096–2108.

Russell, I. J., and Sellick P. M. (**1983**). "Low-frequency characteristics of intracellularly recorded receptor potentials in guinea-pig cochlear hair cells," J. Physiol. (London) **338**, 179–206.

Santos-Sacchi, J., and Dilger, J. P. (**1988**). "Whole cell currents and mechanical responses of isolated outer hair cells," Hear. Res. **35**, 143–150.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Modulation of responses of spiral ganglion cells in the guinea pig cochlea by low frequency sounds," Hear. Res. **7**, 199–221.

Steele, C. R., and Puria, S. (**2003**). "Analysis of forces on inner hair cell cilia," *Biophysics of the Cochlea: From Molecule to Model* (World Scientific, Singapore), pp. 359–365.

Wada, H., Takeda, A., and Kawase, T. (**2002**). "Timing of neural excitation in relation to basilar membrane motion in the basal region of the guinea pig cochlea during the presentation of low-frequency acoustic stimulation," Hear. Res. **165**, 165–176.

Zenner, H. P. (**1986**). "Motile responses in outer hair cells," Hear. Res. **22**, 83–90.

Effects of aspirin on distortion product otoacoustic emission suppression in human adults: A comparison with neonatal data

Carolina Abdala[a)]

House Ear Institute, 2100 West Third Street, Children's Auditory Research and Evaluation Center, Los Angeles, California 90057

One of the distortion product otoacoustic emission (DPOAE) paradigms used to study cochlear function is DPOAE $(2f1-f2)$ ipsilateral suppression. Newborns do not have adultlike DPOAE suppression. At 6000 Hz, infants show excessively narrow DPOAE suppression tuning and shallow growth of suppression for low-frequency suppressor tones. The source of this immaturity is not known but the outer hair cell (OHC) is one possible locus. In the present study, DPOAE suppression was measured at $f2=1500$ and 6000 Hz from two groups with impaired OHC function in an attempt to model the observed immaturity in neonates: adults with aspirin-induced OHC dysfunction and subjects with sensorineural hearing loss (SNHL). Their DPOAE suppression results were compared to those obtained from a group of term newborns to address whether infant DPOAE suppression resembles suppression from individuals with known OHC dysfunction. Results indicate that aspirin systematically alters DPOAE suppression in adults at $f2=6000$ Hz, but not 1500 Hz. However, neither aspirin-induced OHC dysfunction nor naturally occurring SNHL produces "neonatal-like" DPOAE suppression at either test frequency. This finding does not support the hypothesis that non-adultlike DPOAE suppression characterizing newborns can be explained by minor impairments or alterations of OHC function. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1985043]

PACS number(s): 43.64.Jb, 43.64.Kc [BLM] Pages: 1566–1575

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) are produced by the cochlea when the ear is presented with two simultaneous pure tones $(f1,f2)$ relatively close in frequency and level. Middle ear function and auditory thresholds must be normal or near normal and outer hair cells must be present and motile at the distortion generation site to detect a DPOAE in the ear canal (Dorn *et al.*, 1999). When outer hair cells are damaged, DPOAEs are reduced or eliminated (Brown *et al.*, 1989). The inner hair cell, in contrast, can be disabled or damaged with little or no impact on the generation of DPOAEs (Trautwein *et al.*, 1996). Clearly, distortion product otoacoustic emissions offer unique insight into outer hair cell function and integrity (Lonsbury-Martin and Martin, 2002; Ryan, 2002).

One of the DPOAE paradigms applied to study cochlear function and, more recently, cochlear maturation, is DPOAE $(2f1-f2)$ ipsilateral suppression. DPOAE suppression is not adult-like in human infants. Neonatal DPOAE suppression tuning curves are excessively narrow and have a steep low-frequency flank, particularly at $f2=6000$ Hz (Abdala *et al.*, 1996; Abdala, 2003; Abdala, 2004). Additionally, the rate of suppression growth is excessively shallow in newborns for suppressor tones lower in frequency than $f2$. These nonadult-like features of DPOAE suppression may indicate a cochlear immaturity in young infants, however, the source or locus of this immaturity within the cochlea is not known.

The outer hair cell (OHC) is one possible locus of functional immaturity in the newborn cochlea. Although anatomically mature, if the OHC is not *functioning* in an adult-like manner and does not exhibit normal motility at birth, the consequence may be apparent in measures of DPOAE suppression. Outer hair cells show a protracted developmental course in the human cochlea and changes in their pre- and postsynaptic specialization have been observed late into the third trimester of pregnancy in humans (Lavigne-Rebillard and Pujol, 1988). It is possible that human OHC maturation continues into the early post-term weeks and months. In laboratory animals, correlations have been found between the developmental sequence of changes in OHC subcellular organization and development of electromotile properties in vitro (He *et al.*, 1994). Comparable data on the development of OHC motility in humans are not available.

To test the hypothesis of OHC immaturity in newborns, naturally occurring OHC damage or benign manipulations of OHC function must be utilized as models. One such model is aspirin (acetylsalicylic acid). It has been safely used in humans to temporarily impair OHC function and simulate sensorineural hearing loss (SNHL) (McFadden *et al.*, 1984; Wier *et al.*, 1988). Aspirin's effect on the cochlea is well established; it typically causes temporary threshold elevation, tinnitus and a reduction, but not complete elimination of DPOAEs (Cazals, 2000; Long and Tubis, 1988; McFadden, 1982; McFadden *et al.*, 1984; Takahashi *et al.*, 1999; Wier *et al.*, 1988). Research with laboratory animals has clearly shown that aspirin produces these reversible effects by impairing OHC electromotility (Dieler *et al.*, 1991; Oliver *et al.*, 2001). Mild-moderate sensorineural hearing (SNHL) provides a second model. This degree and type of loss is associated with OHC dysfunction. DPOAE amplitude is re-

TABLE I. Hearing impaired subject summary.

	Subject	Age (Yr)	Sex	Ear	PTA (dB HL)	Threshold at f_2 (dB HL)
$f_2=1500$ Hz	HI01	12.5	F	R	35	20
	HI02	10.5	F	L	47	35
	HI03	15	M	L	43	25
	HI04	9	F	L	25	25
$f_2=6000$ Hz	HI05	7	M	L	48	15
	HI06	13	F	R	53	15
	HI07	7	M	L	30	20
	HI08	17	M	L	45	20

duced by even mild amounts of SNHL, presumably due to OHC damage (Abdala and Fitzgerald, 2003; Gorga *et al.*, 1997; Martin *et al.*, 1990).

In the present study, DPOAE suppression data from subjects with aspirin-induced OHC dysfunction and mild-moderate SNHL were compared to DPOAE suppression data from a group of healthy newborns to address the following research question: *Is the nonadult-like DPOAE suppression consistently observed in newborns similar to suppression measured from individuals with known impairment of OHC function?* If so, it may suggest that immature DPOAE suppression in newborns is due to OHC alterations (e.g., immaturities) similar to those produced by aspirin. Conversely, if neonatal DPOAE suppression remains different from adult suppression during aspirin, it would not support the hypothesis that the source of functional immaturity in the newborn cochlea is OHC motility.

II. METHODS

A. Subjects

A total of 41 subjects participated in this study: 10 normal hearing adults, 23 healthy term-born neonates, and 8 school-aged children with mild sensorineural hearing loss (SNHL). The ten adults (7 females and 3 males) had a mean age of 29.1 years (range=22–37 years) with no history of otologic disease or exposure to damaging noise. Four left ears and six right ears were tested. Prospective adult subjects filled out a medical questionnaire and received a medical examination at the House Ear Clinic.

All adult subjects had audiometric thresholds ⩽15 dB HL and normal middle ear function as measured by immittance audiometry. Their DPOAEs had signal to noise ratios (SNRs) ranging from 15 to 33 dB at the two test frequencies (mean=22.6 dB at 1500 Hz; 25.1 dB at 6000 Hz). Of the 24 interested adults who contacted the laboratory to participate, 10 met criteria and were included.

Twenty-three term-born neonates (14 males; 9 females) were tested on average within 2.5 days (59.8 h) of birth. Their mean 1 and 5 min APGAR scores were 8.75 and 9.5 (10=optimal). Fifteen left and eight right ears were tested. Average birth weight was 3216 g. All infant subjects passed a click-evoked ABR hearing screening (35 dB HL) and a screening DP-gram at birth to be included in the study. Neo-

nates who met subject inclusion criteria were randomly assigned to either the $f2=1500$ ($n=11$) or $f2=6000$ Hz ($n=12$) condition.

Eight children with mild SNHL were recruited from the Children's Auditory Research and Evaluation (CARE) Center of the House Ear Institute. Children rather than adults were selected as subjects because they have generally experienced less traumatic noise and/or ototoxic insult than adults, thus, increasing the likelihood of measurable DPOAEs despite mild hearing loss. Four girls and four boys (two right and six left ears) were tested. Each provided DPOAE data at either 1500 Hz ($n=4$) or 6000 Hz ($n=4$). The ear showing the greater DPOAE SNR at the test frequency was chosen as the test ear. Descriptions of the 8 HI subjects are shown in Table I.

B. Instrumentation and signal processing

An Ariel DSP16+ signal processing and acquisition board housed within a Compaq Prolinea 590 personal computer with Pentium processor was used to generate stimuli and acquire data. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12 dB/oct; 710 Hz high-pass cut-off). The ER-10C probe contains two output transducers and a low-noise microphone. The two primary tones and the suppressor tone were generated by the DSP processor. The primary tone at $f1$ was generated by one D/A converter and delivered via one transducer. The primary tone at $f2$ and the suppressor tone (fs) were produced by the second D/A-converter and output through the second transducer. Twenty-five sweeps of the microphone signal were added and comprised one block at $f2=6000$ Hz. At $f2=1500$ Hz, fifty sweeps comprised one block.

C. Data acceptance criteria

Acceptance criteria were as follows: (1) noise measurements for three frequency bins (12.2 Hz wide) on either side of the $2f1-f2$ frequency had to be <0 dB SPL to assure appropriate subject state; and (2) the measured DPOAE level had to be at least 5 dB above the average noise measured in the same six bins around the distortion product frequency to be accepted into the grand average. In addition, sweeps were

accepted into a block of data only when the estimated rms level in that sweep did not exceed a user-controlled artifact rejection threshold.

D. Calibration

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured with the probe in a Zwislocki coupler for all test conditions. The mean level of distortion was -21 dB SPL. The recording system noise floor was determined using a similar method with no tones present. The level of system noise floor ranged between -22 and -27 dB SPL.

An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. Tones of fixed voltage were presented to the transducers at 250 Hz intervals from 500 to 10 000 Hz and the resulting SPL of these tones recorded in each ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

E. Procedure

1. Adults

Adult subjects were tested within an IAC sound-attenuated booth at the House Ear Institute in the Children's Auditory Research Laboratory while sitting comfortably in a padded easy chair, reading or watching a sub-titled video. The following procedures were performed during each session: (1) air and bone conduction audiogram, including interoctave frequencies, (2) DP-gram ($f2=1500–12\,500$ Hz), (3) immittance tympanometry (226 Hz probe tone), (4) spontaneous otoacoustic emissions, (5) DPOAE suppression tuning curves (STCs), and (6) DPOAE Input/output function at both $f2=1500$ Hz and 6000 Hz. An $f2/f1$ ratio of 1.21 and level separation of 10 dB ($L1>L2$) were used for all DPOAE testing. During suppression, primary tones were presented at 65/55 dB SPL.

To record DPOAE STCs at $f2=1500$ Hz, 12 different suppressor tones were presented ranging from 850 to 1890 Hz. At $f2=6000$ Hz, 15 suppressor tones were presented ranging from 3044 to 7232 Hz. Suppressor tones were presented at intervals of 25 to 150 cents (1 octave $=1200$ cents) with smaller intervals in the tip region. The suppressor tone was presented simultaneously with the primary tones and its level increased in 2 dB steps from 40 to 84 dB SPL. Unsuppressed DPOAE amplitude was measured at the start of data collection and again prior to the presentation of each new suppressor frequency.

Completion of this protocol yielded a series of suppression growth functions (DPOAE amplitude $\times$ suppressor level) for each of the two $f2$ conditions. To generate iso-suppression tuning curves (STCs), the suppressor level reducing DPOAE amplitude by 6 dB was determined from the suppression growth function using linear interpolation and was plotted as a function of suppressor frequency. Suppression growth slope was measured for each suppressor tone by fitting a regression line to this function. For DPOAE input/output (I/O) functions, stimuli ranged from 30 to 85 dB SPL and were decreased in 5 dB increments. A successful test session lasted approximately $1\frac{1}{2}–2$ h.

Following the baseline test, adult subjects were given a bottle of aspirin, verbal instructions about dosing and an information sheet with the same written instructions to take home. For any given subject, all aspirin was taken from the same bottle and all subjects took aspirin with an expiration date at least one year from their test date (±1 month). The dosage of aspirin was as follows: Three 325 mg tablets four times a day at 6 h intervals, for a total of 3.9 g daily. The schedule was maintained for four days (McFadden *et al.*, 1984).

The second test session (Session B) was conducted 72 h after initiating the aspirin regimen and the third test session (Session C) was conducted 96 h after initiation (and within 24 h of Session B) with an error range of ±1 h. A blood draw was scheduled either immediately before or after Session C, depending on availability of the phlebotomist. Serum salicylate levels were measured from the blood sample to ensure subject compliance with the drug regimen. Serum salicylate levels for all subjects were within the therapeutic range (11.92–31.87 mg/dl mean$=21.83$) indicating subject compliance with dosing. Only one subject did not complete Session C because of an illness unrelated to aspirin. A final DP-gram and audiogram were conducted within two weeks of completing the aspirin regimen to confirm a return to normal DPOAE amplitude and hearing thresholds.

2. Term neonates and hearing-impaired children

Following the acquisition of informed parental consent, term-born neonates were tested at Women and Children's Hospital Los Angeles County + USC Medical Center prior to discharge, in the Infant Auditory Research Laboratory. Newborns were swaddled and fed if necessary, then placed in a hospital isolette for testing. The DP-gram, DPOAE suppression data and I/O functions were collected in the same manner described for adult subjects with two exceptions: (1) 5 dB increments (not 2 dB) were used to increase suppressor levels during DPOAE ipsilateral suppression and (2) each infant provided data at one $f2$ frequency only. These procedural modifications were required because of time constraints intrinsic to newborn testing.

DPOAE testing with hearing-impaired (HI) subjects was conducted with the child seated comfortably in an easy chair inside a single-walled sound-attenuated booth. Subjects watched a silent video with subtitles during testing. The test procedure is as described for adults except that each of the eight subjects provided data at one test frequency only. Test sessions with HI subjects lasted approximately 2 h.

F. Data analysis

DPOAE suppression data were quantified in the following manner: (1) DPOAE suppression tuning curve width was quantified 10 dB above the tip by calculating a $Q10$; If the STC was so broad that a $Q10$ could not be measured, a low Q value was assigned to reflect this broadness: 1.0 for $f2=1500$ Hz and 1.4 for $f2=6000$ Hz; (2) *slope* of the low- and high-frequency flank of the STC was determined by fitting a

regression line from the tip of the curve to end points on the low and high-frequency flank; (3) *STC tip* frequency and tip level were measured; (4) *DPOAE suppression growth* was quantified by measuring slope of the DPOAE amplitude ×suppressor level function using a linear regression equation. Any initial amplitude plateau was eliminated in this calculation and only the linear portion of this function was included.

DPOAE I/O functions were quantified in the following manner: (1) *DPOAE threshold* was measured as the lowest stimulus level (L1) at which a valid DPOAE was detected, (2) *Saturation* was measured as the second point along a marked plateau in DPOAE amplitude. A plateau was defined as a sequence of three data points with <1 dB increase in amplitude or 4 points with <2 dB increase; (3) *slope* of the I/O function was measured by fitting a regression equation to the monotonically growing portion of the function.

There were two stages of data analysis: (1) Analyses were conducted within the adult group to compare pre-aspirin baseline data to data collected from whichever session (B or C) showed the greater shift in DPOAE values. This analysis was conducted at each $f2$ frequency separately (1500 and 6000 Hz) using paired t-tests; (2) the second stage of analysis compared DPOAE data recorded in the adult group during the maximum aspirin effect with DPOAE data collected from newborns. These comparisons were conducted using unpaired t-tests at 1500 and 6000 Hz separately. Data from HI subjects were included in the figures although not included in the statistical analyses due to small numbers at each frequency.

III. RESULTS

A. The effects of aspirin on adult DPOAEs

Spontaneous otoacoustic emissions (SOAEs) were measured in all adult subjects because a SOAE near the DPOAE frequency could influence DPOAE suppression characteristics, especially during aspirin use. Only 40% of adults had present SOAEs (1 per ear). None of these SOAEs were greater in frequency than 3076 Hz and none were within 300 Hz of either target DPOAE frequency (1000 or 4000 Hz).

Audiometric thresholds increased by a mean of 12.7 dB at 1500 Hz and by 14.3 dB at 6000 Hz when comparing adult data during the Baseline Session to Session C. DPOAE mean amplitude was reduced by 3.8 dB at 1500 Hz and 7.6 dB at 6000 Hz when comparing Baseline and Session C data. Figure 1 shows DPOAE suppression tuning curves from ten adult subjects at Baseline and during the two post-salicylate sessions (Sessions B and C). Additionally, the Appendix includes individual data from each adult subject at each test session.

1. f2=1500 Hz

At $f2=1500$ Hz, aspirin did not shift any DPOAE feature significantly. There was a trend for the STC to become broader (Baseline mean $Q=2.51$; Session C mean $Q=2.07$), but it did not reach significance. The only significant change noted at $f2=1500$ Hz, was a reduction in DPOAE amplitude from Baseline compared to Session C ($p=0.001$).
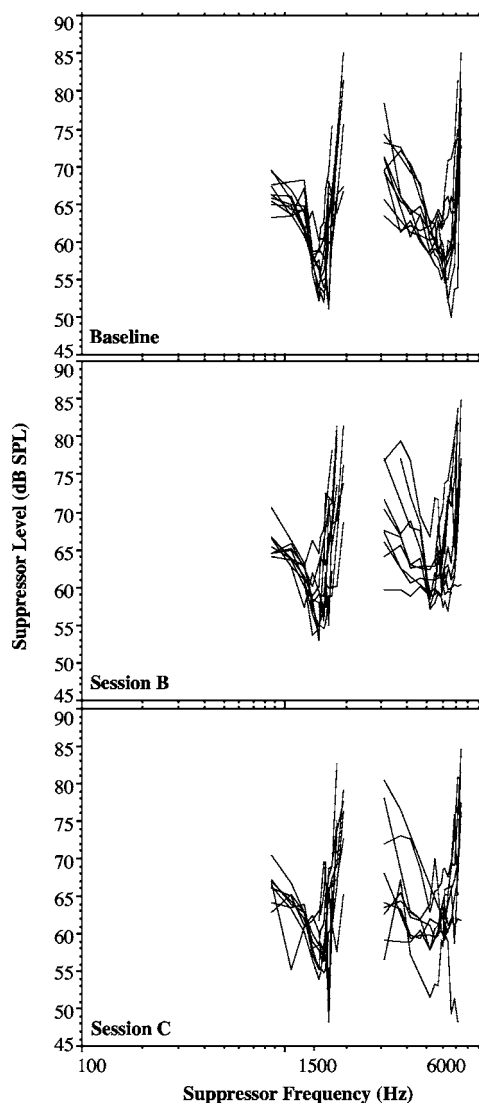
FIG. 1. Individual DPOAE suppression tuning curves for $f2=1500$ and 6000 Hz in 10 normal-hearing adult subjects during three test sessions: Baseline presalicylate, Session B (72 h after beginning a four-day salicylate regimen), Session C (96 h after beginning the salicylate regimen). STCs were generated using primary tones at 65/55 dB SPL and a 6 dB suppression criterion.

2. f2=6000 Hz

The effect of aspirin on DPOAE amplitude, suppression and I/O characteristics was systematic for $f2=6000$ Hz. During Session C, DPOAE unsuppressed amplitude was significantly decreased ($p=0.01$). As seen in Fig. 1, DPOAE STCs at $f2=6000$ Hz showed a general degradation of suppression tuning as the aspirin effect became stronger. Seven out of the 10 adults showed broadened STCs during Sessions B and C. Mean $Q10$ shifted from 2.38 during Baseline to 1.56 when measured after 4 days of aspirin. Paired t-tests comparing baseline DPOAE features to data collected during the maximum aspirin effect (either Session B or C) showed significantly smaller $Q10$ values ($p=0.05$), elevated tip level ($p=0.02$) and tip frequency shifted downward ($p=0.01$).

The slope of suppression growth showed aspirin-related changes restricted to low-frequency side suppressor tones. Figure 2 shows the mean slope of suppression growth, for

Carolina Abdala: Aspirin and distortion product suppression   1569
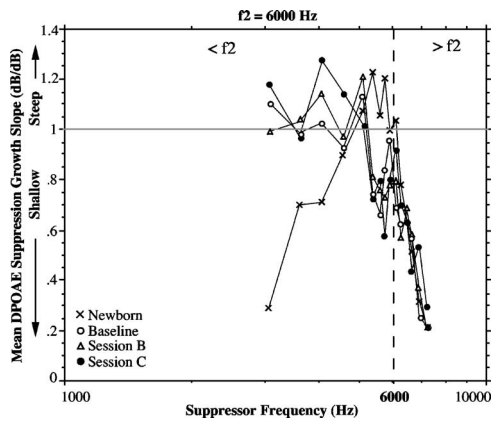
FIG. 2. Mean slope of suppression growth for $f2=6000$ Hz at each suppressor frequency. Results are shown for adult subjects during Baseline, Sessions B and C and for newborns during one test session. Data points that fall to the left of the vertical dashed line represent low-frequency side suppressor tones ($<f2$); those to the right are high-frequency side suppressor tones ($>f2$). Points above the horizontal line correspond to slopes of suppression growth >1 dB/dB and those below the line to slopes <1 dB/dB.

$f2=6000$ Hz, plotted as a function of suppressor frequency for adults (also shown, but discussed later, are newborn data). Adult data in Fig. 2 illustrate the typical pattern of steep suppression growth for low-frequency side ($<f2$) suppressors as they are increased in level, and shallow suppression growth for high-frequency side ($>f2$) suppressors. During Sessions B and C, suppression growth became steeper for three of the four lowest frequency suppressor tones.

The I/O function at $f2=6000$ Hz showed only one significant change during the drug regimen. Mean threshold values shifted from 50 to 57.8 dB SPL ($p=0.006$). Figure 3 shows an example of one adult subject with progressively more elevated DPOAE threshold during aspirin usage.

B. Newborn data

DPOAE suppression tuning curves recorded from 23 term-born neonates are shown in Fig. 4 (See Abdala, 1998 for detailed report of DPOAE suppression in newborns). There was a trend for newborn DPOAE STCs to be slightly narrower at 1500 Hz (Adult $Q=2.5$; Newborn $Q=3.0$) although this effect did not reach statistical significance. Otherwise, at 1500 Hz, DPOAE suppression appears to be gen-
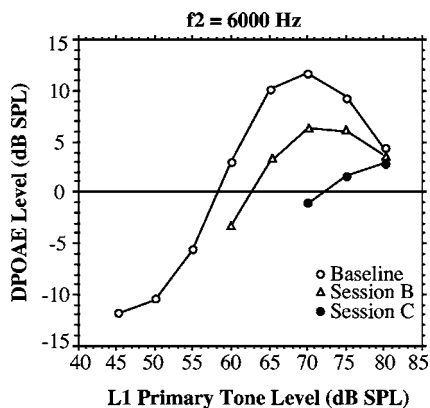
FIG. 3. DPOAE input/output functions at $f2=6000$ Hz for one adult subject during Baseline, Session B, and Session C.
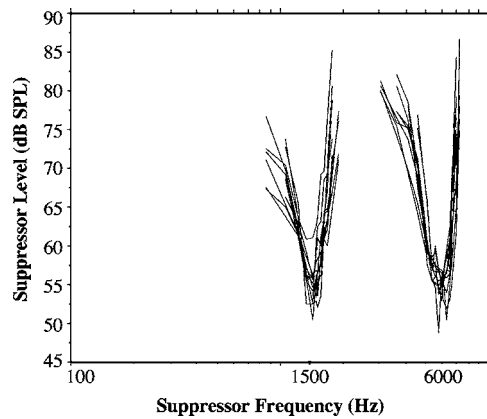
FIG. 4. Individual DPOAE suppression tuning curves for $f2=1500$ (11 ears) and 6000 Hz (12 ears) in term-born neonates. STCs were generated using primary tones at 65/55 dB SPL and a 6 dB suppression criterion.

erally adult-like by term birth. At $f2=6000$ Hz, consistent with previous reports, newborn DPOAE STCs were narrower and steeper than adult STCs collected during baseline (Adult $Q=2.3$; Newborn $Q=3.3$: $p=0.02$). Additionally, suppression growth slope was characteristically shallow in newborns for low-frequency suppressor tones only (see Fig. 2: "x" symbols only).

C. Comparison of aspirin-altered DPOAEs with neonatal DPOAEs

To address the primary question of interest, analyses were conducted to determine whether DPOAE suppression during Sessions B and C (while altered by aspirin) was similar to DPOAE suppression measured from neonates. In all cases, the adult value showing the greater shift from baseline was used in analyses. As previously reported, adult DPOAE data at $f2=1500$ Hz did not shift significantly during the aspirin regimen.

Figure 5 shows mean 6000 Hz DPOAE features measured from adults during their three test sessions and from term neonates tested during one session. Newborn data are plotted under "Session C" (also shown, but discussed later, are HI data). Unpaired t-tests were conducted between adult data from Sessions B/C and neonatal data. Results indicate that newborns and adults tested during aspirin use were significantly different with respect to: Q10 ($p=0.0001$), low and high-frequency STC slope ($p=0.0004$ and 0.0001, respectively), tip level ($p=0.027$), tip frequency ($p=0.017$), and DPOAE amplitude ($p=0.0001$). Tuning curves from neonates were narrower, with steeper flanks, showed a lower tip level, higher tip frequency and a larger amplitude unsuppressed DPOAE relative to adult DPOAEs recorded during the aspirin effect. From Fig. 5 it is also apparent that the shift produced by aspirin did not move adult DPOAE data toward neonatal values; rather, adult tuning curves became less, not more "neonatal-like," during aspirin-induced alterations in OHC function.

Mean adult and newborn DPOAE suppression growth at $f2=6000$ Hz is shown in Fig. 2. Clear neonatal and adult differences are present at baseline, as expected, but more important, these differences are preserved during test ses-
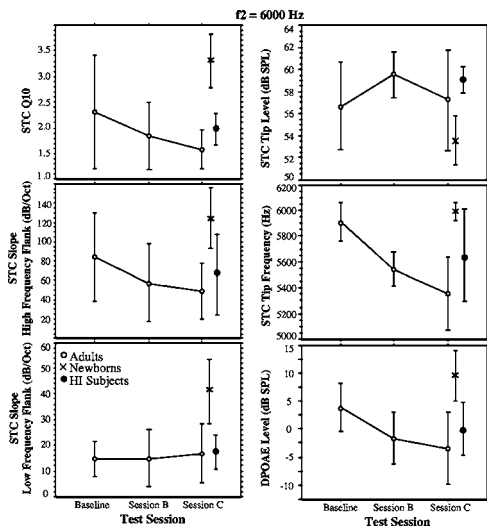
FIG. 5. Mean DPOAE features at $f2=6000$ Hz for adults, term neonates, and a group of hearing impaired subjects. Adult data are plotted during three sessions: pre-salicylate Baseline, Session B (72 h later) and Session C (96 h after beginning the salicylate regimen). Newborn and hearing-impaired data are plotted under Session C for comparison. Error bars represent ± one standard deviation.

sions B and C. Neonates have significantly more shallow suppression growth slope than adults for the four lowest suppressor frequencies. Aspirin did not shift adult suppression growth toward neonatal values, nor minimize differences between these two groups.

DPOAE threshold increased by a mean of 7.8 dB (shifted from 50 to 57.8 dB SPL) during aspirin. Given this shift, adult DPOAE threshold measured during Session C, appears to have become more similar to newborn threshold (55 dB SPL). However, similarity between adult and infant thresholds seen here cannot be attributed to aspirin, because these thresholds were similar both before and after the drug regimen.

D. Comparison of newborn DPOAEs with DPOAEs from hearing-impaired subjects

For a detailed account of how SNHL affects DPOAE suppression, please see Abdala and Fitzgerald, 2003. Although DPOAE STCs recorded from HI subjects showed variability in shape and width, a general degradation or disruption of DPOAE suppression tuning was apparent at both frequencies, but particularly at $f2=6000$ Hz (Fig. 6). At $f2=1500$ Hz, STC width was in the normal range for the HI group; however, at 6000 Hz, their suppression tuning was broadened ($Q=1.98$ vs 2.36 for normal hearing). Tip frequencies were shifted downward at both $f2$ frequencies and STC tip was elevated, only at 6000 Hz.

There does not appear to be any systematic similarity between DPOAE suppression data from these individuals with SNHL and newborn data. The broader suppression tuning at 6000 Hz for HI subjects is in contrast to the excessively narrow STCs observed in newborns (HI $Q10=1.98$ vs 3.3 in newborns). One of the most notable differences between individuals with SNHL and newborns, is STC tip
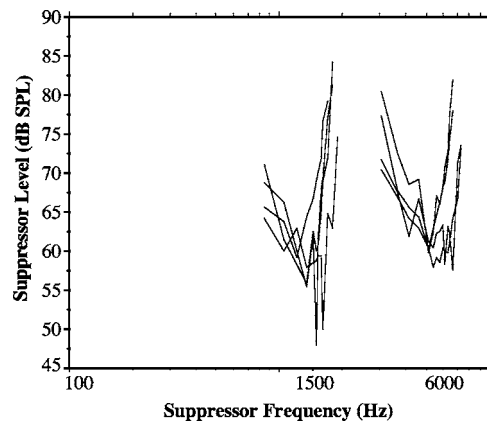
FIG. 6. Individual DPOAE suppression tuning curves for $f2=1500$ and 6000 Hz in eight children with mild-moderate SNHL. STCs were generated using primary tones at 65/55 dB SPL and a 6 dB suppression criterion.

level. At $f2=6000$ Hz, the tip was markedly elevated in the HI group, as is clearly evident by comparing 6000 Hz STCs in Fig. 6 (HI) to those in Fig. 4 (Newborn).

Newborns and HI subjects also show dissimilar DPOAE suppression growth at $f2=6000$ Hz (Fig. 7). In the HI group, the pattern of suppression growth slope is more erratic and shows large fluctuations across suppressor frequency. Additionally, for the lowest three suppressor frequencies, newborns show more shallow suppression growth than HI subjects. These observations suggest that newborn DPOAE suppression does not approximate suppression data from individuals with hearing loss.

The DPOAE I/O functions generated from HI subjects were normal in slope and saturation characteristics but showed elevated DPOAE thresholds (50 dB SPL at 1500 Hz; 55 dB SPL at 6000 Hz). An elevated DPOAE threshold is an expected result when SNHL is present and has been described in the literature previously (Stover *et al.*, 1996). At $f2=1500$ Hz, neonates also had an elevated mean DPOAE threshold of 52 dB SPL. Newborns and these four subjects with hearing impairment were similar with respect to this DPOAE feature at $f2=1500$ Hz.
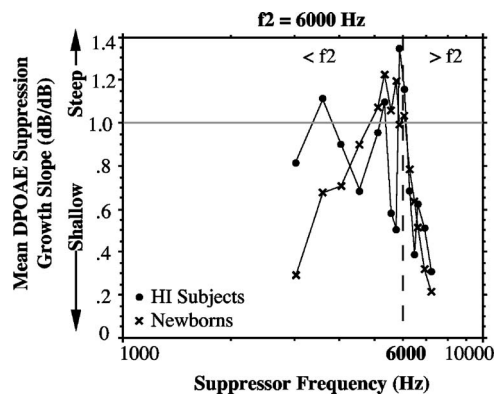
FIG. 7. Mean slope of suppression growth for $f2=6000$ Hz at each suppressor frequency. Data are shown for hearing-impaired subjects and newborns. Data points to the left of the vertical dashed line represent low-frequency side suppressor tones ($<f2$); those to the right are high-frequency side suppressor tones ($>f2$). Points above the horizontal line correspond to slopes of suppression growth >1 dB/dB and those below the line to slopes <1 dB/dB.

E. Serum salicylate levels

Pearson Product Moment Correlations were conducted between the serum salicylate blood level measured following the last adult test session and changes in: (1) audiometric threshold, (2) DPOAE amplitude, and (3) STC Q10 values. This analysis was conducted to examine whether variability in serum salicylate levels among subjects influenced the DPOAE results in a systematic fashion. Only one correlation was significant. At 1500 Hz, there was a significant correlation between serum salicylate blood levels and the decrease in audiometric threshold ($r=0.75$; $p=0.019$). The higher the salicylate level, the greater the shift in audiometric threshold. The lack of correlation between serum salicylate blood level and changes in the DPOAE suggests that it did not influence the results in a systematic fashion.

IV. DISCUSSION

This is the first investigation of the effects of aspirin on DPOAE ipsilateral suppression in humans. Direct studies of the effects of aspirin on the OHC have shown a reduction in cell motility and turgor (Shehata *et al.*, 1991). OHC membrane characteristics are altered when exposed to salicylate, showing increased permeability to K+ on the basolateral membrane, which causes an increased flow of resting current through the cell (Tunstall *et al.*, 1995).

The present study found that salicylate-induced OHC dysfunction in human adults produced significant and systematic effects on DPOAE suppression only for high-frequency primary tones. It has also been show that aspirin-induced hearing loss is more pronounced in high, rather than low frequencies (Cazals, 2000). At $f2=6000$ Hz, aspirin produced broadened DPOAE STCs with elevated tips shifted toward lower frequencies. This frequency effect is consistent with findings that OHC-related phenomenon, such as cochlear nonlinearity and cochlear signal amplification, are strongest in the basal region of the cochlea (Cooper and Rhode, 1995; Rhode and Cooper, 1996).

Aspirin-induced degradation of frequency tuning has been documented previously. Frequency selectivity measured psychoacoustically becomes poorer in human subjects after aspirin usage (Carlyon and Butt, 1993; Beveridge and Carlyon, 1996). Interestingly, Carlyon and Butt (1993) reported that the largest shifts in hearing threshold did not correlate well with the largest degradations in tuning. This was observed in the present study as well. The mean audiometric threshold shifts induced by aspirin at 1500 and 6000 Hz are not greatly different from one another (6000 Hz=14.5 dB; 1500 Hz=12.5 dB) yet DPOAE suppression tuning was only disrupted at $f2=6000$ Hz. It appears that a shift in hearing threshold does not accurately predict the degree or level of disruption in DPOAE suppression tuning. In addition to the effects of aspirin on DPOAE suppression tuning at 6000 Hz, the DPOAE threshold at this frequency was also elevated during the aspirin effect. Combined, these changes in DPOAE suppression tuning and DPOAE threshold suggest that electromotility of OHCs was temporarily disrupted in adult subjects and that the DPOAE suppression paradigm was able to measure or reflect this induced dysfunction.

A. Does OHC dysfunction provide a good model for neonatal cochlear immaturity?

Infants through 3 months of age do not have mature DPOAE suppression at $f2=6000$ Hz, although the locus of this immaturity is not known (Abdala, 2004). The research question in the present study was whether aspirin-induced or naturally occurring OHC dysfunction produces DPOAE suppression that mimics neonatal suppression. Do aspirin-induced alterations in the OHC produce suppression tuning that becomes more narrow and steep on the low-frequency flank with more compressive rates of suppression growth for low-frequency side suppressor tones? If observed, this result would indicate that cochlear function in newborns is like cochlear function in individuals with minor impairments of OHC function. However, adult DPOAE suppression tuning *did not* become more sharp and neonatal-like during the aspirin effect. Adult DPOAE STCs generally became broader and less steep, producing DPOAE data that not only remained significantly different from newborn data but *became more different* during the aspirin effect. It is difficult to argue that dysfunction (or immaturity) of a common cochlear source, such as the OHC, could produce two completely different and opposite patterns of DPOAE suppression.

Neonates do not have DPOAE STCS like those recorded from individuals with OHC dysfunction due to mild-moderate sensorineural hearing loss either. Instead, the disruption of DPOAE suppression in both groups with altered OHC function is consistent with degradation of tuning and cochlear non-linearity observed at many levels of the auditory system when OHCs are damaged or destroyed (Liberman and Dodds, 1984; Evans and Harrison, 1976; Hicks and Bacon, 1999). Hearing-impaired subjects and adults with aspirin-induced OHC dysfunction demonstrate many shared DPOAE characteristics but neither group produces data mimicking the unique patterns of DPOAE suppression consistently observed in neonates.

The dissimilarity between neonatal data and data from either subject group with OHC dysfunction, suggests that an immaturity in OHC motility is not an adequate or likely explanation for nonadult-like DPOAE suppression in newborns. We cannot unequivocally rule out that other aspects of the OHC, not directly related to motility, might be immature and contribute to the DPOAE suppression results seen in infants.

B. DPOAE threshold elevation at 1500 Hz

At 1500 Hz, subjects with SNHL and newborns had comparably elevated DPOAE thresholds (50 and 52.2 dB SPL respectively; normal hearing=38 dB SPL). DPOAE threshold elevation in HI subjects is related to elevated audiometric thresholds produced by hearing loss (Stover *et al.*, 1996). Likewise, DPOAE threshold elevation in newborns may suggest elevated hearing thresholds in the low-frequency audiometric range, possibly due to immaturity of middle ear function. The inefficient transfer of energy through the neonatal middle ear may attenuate stimulus tones entering the cochlea (Keefe *et al.*, 1993, 1994). Elevation of DPOAE threshold at $f2=1500$ Hz could also be unrelated to

auditory phenomena and associated only with the generally high infant noise floor recorded in the low-frequency range. These possibilities should be further explored. It is likely that this threshold immaturity at 1500 Hz comes from a different source than the nonadult-like DPOAE suppression consistently observed in neonates for high-frequency primary tones.

V. CONCLUSION

Numerous studies have documented nonadult-like DPOAE suppression in infants for high-frequency primary tones (Abdala *et al.*, 1996; Abdala, 1998; Abdala and Chatterjee, 2003). This immaturity has been hypothesized to arise from various sources: the outer hair cell, the medial efferent system, the middle ear or from passive vibratory characteristics of basilar membrane. The present study tested one of these hypotheses by investigating whether newborn DPOAE suppression was similar to suppression recorded from two groups with known OHC dysfunction. The findings do not support the hypothesis that nonadult-like DPOAE suppression results that characterize this age group may be explained by minor alterations of OHC function. Neither aspirin-induced changes in OHC function or OHC dysfunction associated with SNHL provide a good model for neonatal immaturity, as measured by DPOAE suppression.

Although a functionally immature middle ear produces attenuation of sound that could influence DPOAE suppression results to some extent, simple modeling of this effect cannot easily account for neonatal immaturities in DPOAE suppression (Abdala, 2004). It may explain the elevation of DPOAE threshold in the low frequencies however. An experiment is currently under way in our laboratory to more directly examine the relationship between middle ear function and DPOAE suppression/threshold during the first months of postnatal life.

Finally, evidence suggests that nonadult-like cochlear function in neonatal gerbils may be due to constraints imposed by immature mass and stiffness of the basilar membrane (Mills and Rubel, 1996). If the physical properties of the basilar membrane cannot sustain or support sound-initiated vibration in a normal way, it may limit effectiveness of the "cochlear amplifier," even if the amplifier is fully developed and functional. The hypothesis that passive vibratory characteristics of the basilar membrane may be immature in human newborns and may contribute to nonadult-like DPOAE suppression results, warrants further investigation.

ACKNOWLEDGMENTS

APPENDIX: INDIVIDUAL ADULT DATA DURING ALL TEST SESSIONS

		$f2=1500$ Hz			$f2=6000$ Hz		
ID	Variables	Baseline	Session B	Session C	Baseline	Session B	Session C
S02	Audio thresh (dB HL)	5	15	⋯	5	15	⋯
	DP ampl (dB SPL)	−2	−5	⋯	8	−1	⋯
	STC Q	1.0	1.0	⋯	1.7	1.4	⋯
	STC tip lev (dB SPL)	59.0	62.9	⋯	58.0	63.0	⋯
	STC tip freq (Hz)	1436	1208	⋯	6087	5350	⋯
	STC LF slope (dB/oct)	71.2	4.9	⋯	10.7	18.7	⋯
	STC HF slope (dB/oct)	70.1	28.6	⋯	85.8	51.3	⋯
	DP thresh (dB SPL)	35	55	⋯	45	50	⋯
S03	Audio threshold	15	25	25	15	30	30
	DP amplitude	5	1	3	0	−7	−7
	STC Q	2.8	2.8	1.0	1.4	1.4	1.4
	STC tip level	54.2	53.7	55.4	63.0	62.5	62.2
	STC tip freq	1436	1345	1436	5914	5119	4560
	STC LF slope	24.3	31.3	38.7	2.7	6.1	7.2
	STC HF slope	68.4	52.9	65.6	41.3	38.2	24.5
	DP threshold	30	40	40	50	60	60
S05	Audio threshold	0	15	15	5	20	20
	DP amplitude	3	0	1	0	−4	−5
	STC Q	4.6	2.2	2.5	3.0	1.6	2.4
	STC tip level	52.0	54.0	54.3	54.7	60.0	51.8
	STC tip freq	1522	1436	1436	6087	5746	5119
	STC LF slope	17.8	16.0	16.0	19.2	14.0	29.3
	STC HF slope	85.4	73.8	98.3	67.7	44.9	43.8
	DP threshold	35	30	30	60	65	70
S09	Audio threshold	0	15	10	5	15	10

ID	Variables	f2=1500 Hz			f2=6000 Hz		
		Baseline	Session B	Session C	Baseline	Session B	Session C
	DP amplitude	14	8	9	9	3	3
	STC Q	1.7	1.0	1.0	3.5	1.6	1.4
	STC tip level	55.4	59.0	58.8	52.0	57.6	59.8
	STC tip freq	1566	1345	1436	6266	5914	6087
	STC LF slope	13.2	11.7	12.6	15.6	8.4	21.3
	STC HF slope	68.3	51.7	51.8	112.6	94.5	89.8
	DP threshold	30	30	35	45	60	40
S11	Audio threshold	10	25	35	0	15	20
	DP amplitude	4	−4	−13	6	−4	−16
	STC Q	1.8	2.7	1.0	3.1	2.6	1.4
	STC tip level	55.9	55.2	⋯	55.2	57.2	⋯
	STC tip freq.	1478	1612	⋯	6266	5119	⋯
	STC LF slope	11.1	19.7	⋯	20.9	23.8	⋯
	STC HF slope	77.8	102.1	⋯	143.4	33.8	⋯
	DP threshold	45	40	75	50	60	75
S12	Audio threshold	5	20	30	0	25	30
	DP amplitude	3	−2	−5	2	−7	−9
	STC Q	1.0	1.0	1.0	1.4	1.6	1.4
	STC tip level	58.4	63.3	50.6	58.9	57.8	59.6
	STC tip freq	1345	1522	1612	5119	5119	5350
	STC LF slope	11.9	6.5	⋯	14.3	12.1	8.9
	STC HF slope	48.3	63.6	168.2	24.0	56.8	75.7
	DP threshold	35	60	55	40	55	60
S14	Audio threshold	0	15	15	0	15	15
	DP amplitude	0	−5	−5	11	3	2
	STC Q	3.8	1.0	4.9	1.4	1.4	1.4
	STC tip level	51.2	55.9	48.5	57.8	60.5	59.6
	STC tip freq	1612	1522	1612	6087	6087	4560
	STC LF slope	22.9	8.8	12.8	6.6	4.1	9.1
	STC HF slope	69.1	33.4	59.1	74.4	141.2	56.3
	DP threshold	50	50	55	45	60	70
S17	Audio threshold	0	10	10	5	25	25
	DP amplitude	9	7	6	1	2	1
	STC Q	1.0	1.0	1.0	1.8	1.4	1.4
	STC tip level	60.5	53.2	57.8	55.3	59.1	57.9
	STC tip freq	1522	1436	1522	5746	5746	5119
	STC LF slope	11.2	21.4	14.7	13.3	0.6	0.6
	STC HF slope	55.2	28.9	64.6	77.4	1.2	6.4
	DP threshold	50	60	35	60	60	50
S21	Audio threshold	5	15	10	5	15	15
	DP amplitude	7	5	5	4	5	2
	STC Q	4.2	2.2	3.8	1.7	3.1	1.9
	STC tip level	51.4	55.9	52.0	61.2	58.6	58.1
	STC tip freq	1612	1478	1612	5119	5119	5119
	STC LF slope	16.3	13.0	14.3	20.7	35.9	26.1
	STC HF slope	53.1	51.3	78.0	42.4	50.8	45.0
	DP threshold	30	30	35	40	45	45
S23	Audio threshold	5	10	5	0	0	0
	DP amplitude	11	12	13	−1	−6	−3
	STC Q	3.2	2.5	2.4	4.9	1.4	1.4
	STC tip level	53.5	54.7	55.1	50.3	59.2	48.7
	STC tip freq	1522	1522	1522	6449	6087	6932
	STC LF slope	20.6	20.6	20.2	24.6	24.5	30.6
	STC HF slope	82.8	73.9	70.5	166.5	⋯	⋯
	DP threshold	40	30	40	65	60	50
Mean	Audio threshold	4.5	16.5	17.2	4	17.5	18.3

Carolina Abdala: Aspirin and distortion product suppression

ID	Variables	$f2 = 1500$ Hz			$f2 = 6000$ Hz		
		Baseline	Session B	Session C	Baseline	Session B	Session C
	DP amplitude	5.4	1.7	1.6	4	−1.6	−3.6
	STC Q	2.5	1.7	2.1	2.4	1.7	1.6
	STC tip level	55.1	56.8	54.1	56.6	59.5	57.2
	STC tip freq	1505.1	1442.6	1523.5	5914	5540.6	5355.7
	STC LF slope	22.1	15.4	18.5	14.9	14.8	16.6
	STC HF slope	67.8	56.0	82	83.5	57	48.8
	DP threshold	38	42.5	44.4	50	57.5	57.8

Abdala, C. (**1998**). "A developmental study of DPOAE (2f1−f2) suppression in humans," Hear. Res. **121**, 125–138.

Abdala, C. (**2003**). "A longitudinal study of distortion product otoacoustic emission ipsilateral suppression and input/output characteristics in human neonates," J. Acoust. Soc. Am. **114**, 3239–3250.

Abdala, C. (**2004**). "DPOAE (2f1−f2) suppression in three-month-old infants: Evidence for postnatal maturation of human cochlear function?," J. Acoust. Soc. Am. **116**, 3572–3580.

Abdala, C. and Chatterjee, M. (**2003**). "Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans," J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C. and Fitzgerald, T. (**2003**). "Ipsilateral distortion product otoacoustic emission (2f1−f2) suppression in children with sensorineural hearing loss," J. Acoust. Soc. Am. **114**, 919–931.

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F-G. (**1996**). "Distortion product otoacoustic emission suppression tuning curves in human adults and neonates," Hear. Res. **98**, 38–53.

Beveridge, H. and Carlyon, R. (**1996**). "Effects of aspirin on human psychophysical tuning curves in forward and simultaneous masking," Hear. Res. **99**, 110–118.

Brown, A., McDowell, B., and Forge, A. (**1989**). "Acoustic distortion products can be used to monitor the effects of chronic gentamicin treatment," Hear. Res. **42**, 143–156.

Carlyon, R. and Butt, M. (**1993**). "Effects of aspirin on human auditory filters," Hear. Res. **66**, 233–244.

Cazals, Y. (**2000**). "Auditory sensori-neural alterations induced by salicylate," Prog. Neurobiol. **62**, 583–631.

Cooper, N. and Rhode, W. (**1995**). "Nonlinear mechanics at the apex of the guinea-pig cochlea," Hear. Res. **82**, 225–243.

Dieler, R., Shehata-Dieler, W., and Brownell, W. (**1991**). "Concomitant salicylate-induced alterations of outer hair cell subsurface cisternae and electromotility," J. Neurocytol. **20**, 637–653.

Dorn, P., Piskorski, P., Gorga, M., Neely, S., and Keefe, D. (**1999**). "Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses," Ear Hear. **20**, 149–163.

Evans, E. and Harrison, R. (**1976**). "Correlation between outer hair cell damage and deterioration of cochlear nerve tuning properties in the guinea pig," J. Physiol. (London) **256**, 43–44.

Gorga, M., Neely, S., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**1997**). "From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss," Ear Hear. **18**, 440–455.

He, D., Evans, B., and Dallos, P. (**1994**). "First appearance and development of electromotility in neonatal gerbil outer hair cells," Hear. Res. **78**, 77–90.

Hicks, M. and Bacon, S. (**1999**). "Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking," J. Acoust. Soc. Am. **106**, 1436–1450.

Keefe, D., Bulen, J., Arehart, K., and Burns, E. (**1993**). "Ear-canal impedance and reflection coefficient in human infants and adults," J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D., Bulen, J., Arehart, K., and Burns, E. (**1994**). "Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal," J. Acoust. Soc. Am. **95**, 355–371.

Lavigne-Rebillard, M. and Pujol, R. (**1988**). "Hair cell innervation in the fetal human cochlea," Acta Oto-Laryngol. **105**, 398–402.

Liberman, M. and Dodds, L. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.

Long, G. and Tubis, A. (**1988**). "Modification of spontaneous and evoked otoacoutic emissions and associated psychoacoustic microstructure by aspirin consumption," J. Acoust. Soc. Am. **84**, 1343–1353.

Lonsbury-Martin, B. and Martin, G. (**2002**). "Distortion product otoacoustic emissions," in *Otoacoustic Emissions, Clinical Applications*, edited by M. Robinette and T. Glattke (Theime Medical, New York).

Martin, G., Ohlms, L., Franklin, D., Harris, F., and Lonsbury-Martin, B. (**1990**). "Distortion product emission in humans. III. Influence of sensorineural hearing loss," Ann. Otol. Rhinol. Laryngol. **99**, 30–41.

McFadden, D. (**1982**). *Tinnitus: Facts, Theories, and Treatments*. (National Academy Press, Washington, DC).

McFadden, D., Plattsmier, H., and Pasanen, E. (**1984**). "Aspirin-induced hearing loss as a model of sensorineural hearing loss," Hear. Res. **16**, 251–260.

Mills, D. and Rubel, E. (**1996**). "Development of the cochlear amplifier," J. Acoust. Soc. Am. **100**, 428–441.

Oliver, D., He, D., Klöcker, N., Ludwig, J., Schulte, U., Waldegger, S., Ruppersberg, J., Dallos, P., and Fakler, P. (**2001**). "Intracellular anions as the voltage sensor of prestin, the outer hair cell motor protein," Science **292**, 2340–2343.

Rhode, W. and Cooper, N. (**1996**). "Nonlinear mechanics in the apical turn of the chinchilla cochlea in vivo," Aud. Neurosci. **3**, 101–121.

Ryan, A. F. (**2002**). "New views of cochlear function," in *Otoacoustic Emissions, Clinical Applications*, edited by M. Robinette and T. Glattke (Theime Medical, New York).

Shehata, W., Brownell, W., and Dieler R. (**1991**). "Effects of salicylate on shape, electromotility and membrane characteristics of isolated outer hair cells from guinea pig cochlea," Acta Oto-Laryngol. **111**, 707–718.

Stover, L., Gorga, M., Neely, S., and Montoya, D. (**1996**). "Toward optimizing the clinical utility of distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **100**, 956–967.

Takahashi, S., Ohyama, K., Ikeda, K., Takasaka, T., and Wada, H. (1999). "Salicylate effects on the distortion products in live guinea-pig cochlea," Abstr. Assoc. Res. Otolaryngol., St. Petersberg, FL, Vol. **22**, p.97.

Trautwein, P., Hofstetter, P., Wang, J., Salvi, R., and Nostrant, A. (**1996**). "Selective inner hair cell loss does not alter distortion product otoacoustic emissions," Hear. Res. **96**, 71–82.

Tunstall, M., Gale, J., and Ashmore, J. (**1995**). "Action of salicylate on membrane capacitance of outer hair cells from the guinea-pig cochlea," J. Physiol. (London) **485**, 739–752.

Wier, C., Pasanen, E., and McFadden, D. (**1988**). "Partial dissociation of spontaneous otoacoustic emissions and distortion products during aspirin use in humans," J. Acoust. Soc. Am. **84**, 230–237.

Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates

Arturo Moleti[a]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

Renata Sisto[b]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

Gabriella Tognola,[c] Marta Parazzini, Paolo Ravazzani, and Ferdinando Grandori
Istituto di Ingegneria Biomedica CNR, c/o Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano, Italy

Transient evoked otoacoustic emissions of a large population of neonates (466 ears) are analyzed, with the aim of establishing if a significant latency difference can be observed between "pass" (333) and "fail" (133) ears, discriminated with a screening protocol based on band and global reproducibility. The ears that did not pass the test in at least one frequency band are named "fail," for simplicity, but they should be more appropriately defined as "partial pass." In a previous study, significantly different average latencies had been observed in adult subjects, comparing normal hearing and hearing impaired ears [J. Acoust. Soc. Am. **111**, 297–308 (2002)], but no similar study has been conducted on neonates yet. An improved wavelet technique was applied to transient evoked otoacoustic emission data, to get accurate experimental measurements of the otoacoustic emission latency. The results show statistically significant differences between the latency distributions of the "pass" and "fail" populations, with the "fail" latencies longer. However, non-Gaussianity of the distributions and systematic errors associated with low reproducibility levels suggested using a conservative approach in the data analysis and interpretation. New otoacoustic estimates of cochlear tuning in neonates are also provided. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000769]

I. INTRODUCTION

Otoacoustic emissions (OAEs) provide important information about cochlear physiology (Probst *et al.*, 1991). The transient evoked OAE (TEOAE) latency is defined as the delay between the onset of the click and the maximum of the evoked OAE amplitude. For each frequency within the stimulus bandwidth, latency essentially consists of the traveling time along the cochlear membrane from the base to its tonotopic site and back (cochlear round trip delay). A much smaller contribution, of order 0.5 ms (Sisto and Moleti, 2002) comes from the transmission through the outer and middle ear. As tonotopic sites are distributed along the cochlea according to the Greenwood frequency map (Greenwood, 1990), latency is intrinsically frequency dependent. In particular, the relation between latency and frequency is dependent on the quality factor of the cochlear filter through the dependence of the traveling wave velocity on the sharpness of the resonance. Thus, measurements of the transient evoked otoacoustic emission (TEOAE) latency provide an indirect estimate of the tuning curve, which is important for subjects that, like neonates, cannot be studied with psychoacoustical techniques. More generally, as with other OAE measurable parameters (DPOAE phase-gradient delays, TEOAE stimulus-response curves, TEOAE decay time, etc.), TEOAE latency provides important information about the cochlear transmission and amplification mechanisms. Cochlear transmission and amplification are related, because both the amplifier gain and the traveling wave velocity are dependent on the sharpness of the cochlear resonances.

Clinical dichotomic "pass-fail" criteria discriminate between presumed normal-hearing and presumed sensorineural hearing loss ears on the basis of the TEOAE reproducibility (Vohr and Maxon, 1996), which is equivalent to the signal to noise ratio or to the absolute response level, given a typical instrumental noise level. Latency and tuning curves could also be used, in principle, to test the cochlear functionality, with the advantage of using differential measurements, insensitive to absolute calibration uncertainties.

Time-frequency analysis techniques are now available (Wit *et al.*, 1994; Tognola *et al.*, 1997; Tognola *et al.*, 1998; Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2004), which are capable of providing reliable estimates of the OAE spectral latency.

[a]Electronic mail: arturo.moleti@roma2.infn.it
[b]Electronic mail: r.sisto@dil.ispesl.it
[c]Electronic mail: gabriella.tognola@polimi.it

Previous time-frequency estimates of the OAE latency in adults found a clear frequency dependence, in good agreement with a simple power law:

$$\tau(f) = af^b, \tag{1}$$

with f in kHz. The experimental estimates were consistent with a value of a of order 10 ms and b in the range -0.4 to -0.7 (Tognola *et al.*, 1997; Sisto and Moleti, 2002). As the Greenwood map $x(f)$ is a logarithmic function of the frequency, this power law is approximately equivalent to the exponential function of the cochlear distance that has been proposed to fit the spectral latency versus cochlear distance relationship in studies of the derived band acoustic brainstem response (ABR) latency (Donaldson and Ruth, 1993). A recent study (Jedrzejczak *et al.*, 2004) confirmed earlier OAE observations, and demonstrated that very good time-frequency resolution may be achieved using Matching Pursuit (MP) iterative algorithms.

Cochlear latency is also related to the $2f_1-f_2$ distortion product OAE (DPOAE) group delay (more correctly named the phase-gradient delay) (Moulin and Kemp, 1996; Prijs *et al.*, 2000; Schoonoven *et al.*, 2001). The DPOAE phase-gradient delay is a characteristic time associated with the slope of the functional relation between the DPOAE phase and the DPOAE frequency $f_{DP}=2f_1-f_2$. This measure can be performed by either sweeping f_2 (fixed-f_1 paradigm, yielding the characteristic delay τ_1) or f_1 (fixed-f_2 paradigm, yielding a conceptually different delay, τ_2). The DPOAE group delays are related to the round trip delay introduced by the forward transmission along the cochlear membrane of the primary tones from the source to the DPOAE generation place and the backward transmission of the DPOAE to the microphone.

The interpretation of the DPOAE phase-gradient delay data in terms of cochlear tuning is rather difficult (Shera *et al.*, 2000), mainly because the forward and backward transmitted signals are of different frequency, and because the DPAOE signal itself comes from minimally two sources at two different cochlear positions, the tonotopic place $x(f_{DP})$ and the region between $x(f_1)$ and $x(f_2)$. For the same reason, phase-gradient delays cannot be directly compared to TEOAE latencies.

Shera *et al.* (2002) measured stimulus-frequency OAE (SFOAE) group delays, in humans and small mammals, also providing an interpretation of the results in terms of cochlear tuning. Human cochlear tuning resulted sharper than previously assumed, and in agreement with recent behavioral measurements (Oxenham and Shera, 2003). More recently, Shera and Guinan (2003) applied these results to successfully test the predictions of the theory of coherent reflection filtering (Zweig and Shera, 1995; Talmadge *et al.*, 1998).

We have remarked that the measure of the cochlear characteristic times is important because they are related to the quality factor (tuning) of the cochlear filter. We also note that the cochlear filter parameters are, in turn, responsible for the sensitivity and frequency selectivity of the cochlear amplifier, and they are very sensitive to outer hair cell (OHC) damage. Thus OAE characteristic times are expected to be sensitive, as OAE amplitudes are, to mild cochlear damage,

with the advantage of being estimated with differential measurements, so avoiding the systematic errors associated with the measurement of absolute amplitudes.

From a clinical point of view, many authors reported a correlation between hearing loss and the degradation of many OAE parameters (Attias *et al.*, 1995; Kowalska and Sulkowski, 1997; Hussain *et al.*, 1998; Hall and Lutman, 1999; Tognola *et al.*, 1999; Lucertini *et al.*, 2002).

With regard to OAE latency, Sisto and Moleti (2002) and Lucertini *et al.* (2002) showed a significant correlation between anomalously long OAE spectral latency and hearing impairment in adults exposed to impulsive noise. A significant latency difference was also found between healthy neonates and healthy adults (Moleti and Sisto, 2003). As high-level click stimuli had been used in both studies (80 dB peak SPL), the response of the cochlear amplifier is strongly nonlinearly saturated. Thus, a possible interpretation of these results is that the longer latencies observed in damaged ears and in neonate ears, which would imply a higher quality factor, if a resonant transmission line cochlear model is assumed, are due to reduced nonlinear saturation of the cochlear amplifier. In the case of neonates a similar effect on DPOAE suppression tuning curves has been attributed indeed to immaturity of the efferent feedback system (Abdala, 2001). In impaired ears, this could be due to some compensative modification of the nonlinear characteristics of the cochlear amplifier, whose interpretation could provide important clinical information. Hence, the interest of comparing the latency measured in populations of neonates discriminated according to their score in standard hearing screening tests.

Our aim in this work is indeed to establish if a statistically significant difference can be observed between OAE latency and cochlear tuning in populations of neonate ears discriminated with a standard screening protocol based on TEOAE reproducibility. We will carefully discuss the possible sources of systematic errors, associated with the proposed latency estimate methods, which could bias the comparison between different populations.

These accurate measurements of the latency-frequency relationship on a large dataset of neonates, may also serve as a reference for future clinical studies using the same acquisition and analysis techniques.

II. METHOD

TEOAEs were obtained from 466 ears of full-term babies on the third day after delivery during a universal hearing screening program at the Service of Neurophysiopathology of the Mangiagalli Clinic, (Milan, Italy). Responses were recorded using an ILO88 system (Otodynamics, Ltd.), digitized at a rate of 25 000 samples/s. A standard neonatal probe was used. For each ear, two sets of replicate responses (A and B replicates recordings in ILO equipment) to 260 repetitions of the click-train (4 clicks per train) were averaged, according to the "nonlinear" mode of operation (Bray and Kemp, 1987; Kemp *et al.*, 1986). The averaged data were windowed and digitally filtered using the default ILO window (2.5–20 ms, cosine tapered from 2.5 to 5 ms) and

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Moleti *et al.*: Otoacoustic emission latency in neonates   1577

filter (second-order digital bandpass set at 600–6000 Hz). All the responses included in the examined dataset were checked to have a stimulus stability greater than 75% and a stimulus level ranging from 77 to 83 dB peak SPL.

Hearing functionality was assessed accordingly to the following criteria (Vohr and Maxon, 1996): an emission was defined as "pass" when the overall reproducibility (i.e., the zero-lag cross-correlation coefficient between the A and B replicates in the 2.5–20 ms window) was ⩾0.7 and the reproducibility in 0.8 kHz wide frequency bands was ⩾0.5 at 1.6 kHz, ⩾0.7 at 2.4, 3.2, and 4.0 kHz; "partial pass" when not all the five values met the criterion; "refer" when all values did not meet it. The criterion for the first frequency band is lower than the others to take into account the effects of the residual noise, which mostly affects the low-frequency region of TEOAEs. In the following we will call "fail" these "partial pass" emissions, for brevity, but one has to remember that they mostly include relatively high reproducibility OAEs, which have not passed the test in only one or two frequency bands. In fact, all these subjects passed an ABR test, so they are all "normal hearing," with different degrees of TEOAE reproducibility. We are just assuming that a systematically lower reproducibility (i.e., SNR) is likely to be statistically associated with a lower effectiveness of the cochlear active mechanisms.

The 20 ms TEOAE waveforms were analyzed by a time-frequency wavelet transform (WT) technique (for more details, see Tognola *et al.*, 1997; Tognola *et al.*, 1998). The WT at the generic time t and frequency f of signal $x(t)$ is defined as (Mallat, 1989)

$$\mathrm{WT}(\tau,f) = \int_t x(t)\left(\frac{f}{f_0}\right)^{-2}\sqrt{\frac{f}{f_0}}\psi^*\left(\frac{f}{f_0}(t-\tau)\right)dt, \qquad (1')$$

where * denotes complex conjugation. Basically, the WT decomposes a signal $x(t)$ into elementary components by means of a bank of bandpass filters—$\sqrt{f/f_0}\cdot\psi[f/f_0(t)]$—that are iteratively derived from a unique prototype function $\psi(t)$, called the mother wavelet. The mother wavelet $\psi(t)$ is a function with finite energy and centered on time $t = 0$; its Fourier transform is a bandpass function centered on frequency f_0. The bandwidth of the bandpass filters derived from the mother wavelet is proportional to the central frequency of the filter, whereas the duration of their impulse response in the time domain is inversely proportional to the central frequency of the filter [for a more detailed mathematical treatment, see Tognola *et al.* (1997, 1998)]. Roughly speaking, this type of analysis provides a decomposition of a signal into subbands with a bandwidth linearly increasing with frequency. In this form, the WT can be viewed as a special case of a spectral analyzer.

The original signal $x(t)$ can be synthesized by adding all the contributions $\mathrm{WT}(\tau,f)$ as follows (Mallat, 1989):

$$x(t) = c\iint \mathrm{WT}(\tau,f)\sqrt{\frac{f}{f_0}}\psi\left(\frac{f}{f_0}(t-\tau)\right)\left(\frac{f}{f_0}\right)^2 d\tau\,df, \qquad (2)$$

where c is a constant that depends on the mother wavelet, and the frequency integration is extended over the whole frequency range. By integration of the coefficients $\mathrm{WT}(\tau,f)$ over 500 Hz-wide frequency bands we get the wavelet band coefficients $W(f_k,t)$, which represent the time evolution of the evoked signal in seven frequency bands, centered at the frequency $f_k(k=1,\ldots,7)$ equal to 1, 1.5, 2, 2.5, 3, 3.5, and 4 kHz. Each function $W(f_k,t)$ is normalized to its maximum value, and describes the time evolution of the signal belonging to a given spectral band.

As in Sisto and Moleti (2002), two alternative definitions of latency were considered. For each ear j, the time at which the function $W_j(f_k,t)$ reaches its maximum value is defined as the latency $\tau_j(f_k)$ of the jth ear at frequency f_k. By doing so, one gets a statistical distribution of the latency for each frequency, for the two populations (in this study, the "pass" and "fail" ears). An average of the $\tau_j(f_k)$ is performed over the ears, yielding

$$\langle\tau^*(f_k)\rangle = \frac{1}{N_{\mathrm{ears}}}\sum_{j=1}^{N_{\mathrm{ears}}}\tau_j(f_k), \qquad (3)$$

which is an estimate of the average spectral latency of each population in the kth frequency band.

The alternative average latency estimate was obtained by computing the average of the band coefficients over the ears of each population:

$$\langle W(f_k,t)\rangle = \frac{1}{N_{\mathrm{ears}}}\sum_{j=1}^{N_{\mathrm{ears}}}W_j(f_k,t), \qquad (4)$$

and defining the average latency $\langle\tau(f_k)\rangle$ as the time for which the averaged band coefficient $\langle W(f_k,t)\rangle$ reaches its maximum value. This method has the advantage of being less sensitive to the fact that high-frequency wavelet band coefficients sometimes have large amplitudes at very long delays, but does not give an estimate of individual latencies; thus a quantitative evaluation of the significance of the difference between the two populations is more difficult. In the next section we will show the results obtained with both methods.

For approximately Gaussian distributions, the two methods tend to give coincident results, as shown in Sisto and Moleti (2002), and the student's t-test procedure applied to the distributions of individual latencies would give a reasonable estimate of the statistical significance of a difference between distributions of the two populations. Unfortunately, the statistical distributions turned out to be non-Gaussian and, in particular, the high-frequency latency distributions of the "fail" population showed long tails corresponding to unreasonably long latency values. This turned out to be due to a systematic error, which required the introduction of a data selection criterion, as will be discussed in Sec. III.

The average latencies were best fitted to a power law of the form

$$\tau(f) = \tau_{\mathrm{nc}} + af^b, \qquad (5)$$

with f in kHz and $\tau_{\mathrm{nc}}=0.5$ ms. This estimate was first obtained by Sisto and Moleti (2002), by analyzing the TEOAE wavelet band coefficients as functions of time, in ears showing intense individual resonances in their TEOAE spectrum. In such cases, a quasiperiodic time behavior was observed, which was attributed to subsequent

reflections of the same wave packet between the tonotopic place and the base. The noncochlear contribution was roughly estimated as the difference between the absolute delay of the first echo and the differential delay between subsequent echoes. This procedure had also been applied to the neonate data analyzed in Moleti and Sisto (2003), finding again $\tau_{nc}=0.5$ ms.

We recall that the measured latency $\tau(f)$ at a given frequency f is made up of a cochlear contribution, $\tau_c(f)$, due to the round-trip delay associated with the transmission of the acoustic signal along the cochlear membrane, and of a much smaller contribution from the outer and middle ear, τ_{nc}. The previously described technique had been successfully used to show significant differences between the latencies of normal-hearing and impaired ears (Sisto and Moleti, 2002), and also between those of normal-hearing ears of subjects exposed to impulsive firearm noise and of nonexposed subjects (Lucertini *et al.*, 2002).

In a previous study (Sisto and Moleti, 2002), it was noted that the use of octave bands, which was a consequence of the filterbank cascade algorithm used to perform the wavelet transform, could lead to systematic errors in the estimate of the slope of the latency-frequency relation, depending on the typical frequency of the main TEOAE lines in each octave band. This source of systematic error is reduced, in the present study, by performing the analysis with better frequency resolution.

The method described in Moleti and Sisto (2003) was applied to the average latency data, to get an estimate of the neonatal tuning curve. This method is based on a simple transmission line cochlear model, in which the round trip cochlear delay is computed as twice the integral of the inverse traveling wave velocity (TWV) from the base to the tonotopic place:

$$\tau_c(\omega) = 2\int_0^{x(\omega)} dx\, \mathrm{TWV}^{-1} = 2\int_0^{x(\omega)} dx\, \frac{\partial k}{\partial \omega}. \quad (6)$$

The TWV is a function of the relation between the wavenumber k, angular frequency $\omega=2\pi f$, and cochlear position x, and drops to a minimum near the resonant place. As this slowing down is a function of the resonance quality factor Q, it is possible to get an estimate of the tuning curve from the measured latency:

$$Q(\omega) \approx \left(\frac{\omega k_\omega}{\alpha k_0}[\tau(\omega) - \tau_{nc}] - \frac{\sqrt{2}}{\alpha} \right)^2, \quad (7)$$

where $\tau(\omega)$ is the experimental average latency value, $k_\omega = 1.382$ cm^{-1} is the scale parameter of the Greenwood map, $k_0=31$ cm^{-1} is a constant associated with the geometry and density of the basilar membrane and of the cochlear cavity (Talmadge *et al.*, 1998), and $\alpha \approx 0.9$.

III. RESULTS

The wavelet analysis method was applied to all the data collected in this experiment. The two definitions of average latency described in the previous section failed to give coincident results. In Fig. 1 we show the average latency values obtained for the two populations using the two methods. Two
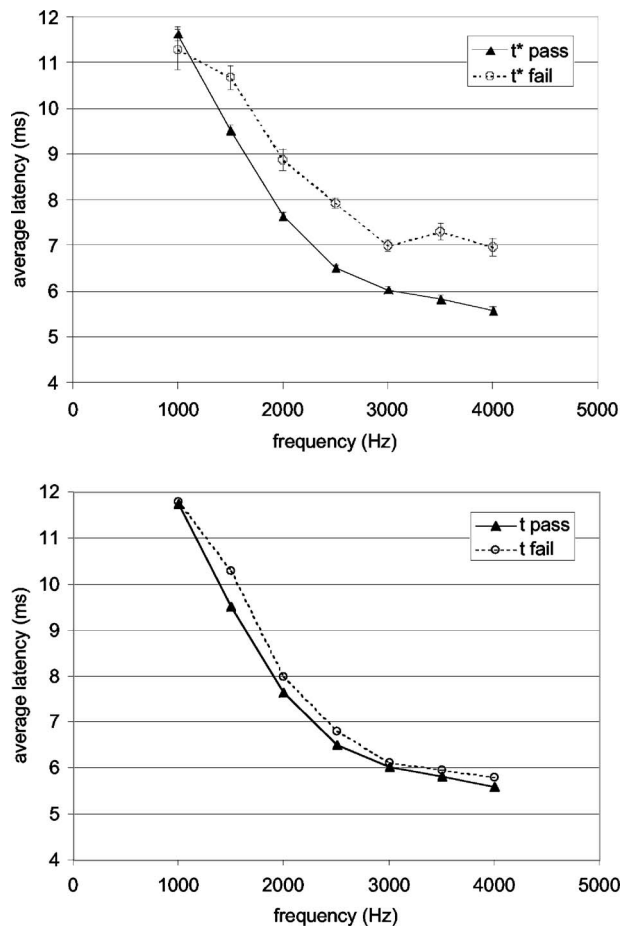
FIG. 1. A comparison between the average latency of the 333 "pass" (black triangles) and of the 133 "fail" (white circles) neonate ears, using the two alternative definitions of average latency $\tau^*(f)$ (top), defined as the average of the individual latencies, and $\tau(f)$ (bottom), defined as the time of the maximum of the average wavelet band coefficients.

main facts are evident: (1) in both cases the average latencies of the "fail" population are longer than those of the "pass" population; (2) the difference is much larger using the first definition of the average latency $\langle \tau^*(f) \rangle$ [see Eq. (3)], which is the average of the individual latencies. The large discrepancy between the two results in the high-frequency range can be explained by looking more carefully at the distributions of the individual latency data. As an example, in the bottom panel of Fig. 2 the histogram of the 3 kHz individual latency data is shown for the "fail" population. Long tails extending up to unreasonably long latencies are clearly visible, which give a significant contribution to the average latency. The histograms of the "pass" population latency do not show such abnormal tails, as shown in the top panel of Fig. 2. The same individual data of the bottom panel of Fig. 2 are plotted in Fig. 3 (top panel) as a function of the individual band reproducibility. This plot shows that the abnormally long latencies are typically associated with low reproducibility in the same 500 Hz wide band. The same pattern may be observed in all mid-high-frequency bands, as shown in Fig. 3 (bottom panel) for the 2.5 kHz band. The seven average wavelet band coefficients are plotted in Fig. 4, showing a plateau on the long latency side of the high-frequency band coefficients, including a large contribution from the ears

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Moleti *et al.*: Otoacoustic emission latency in neonates   1579
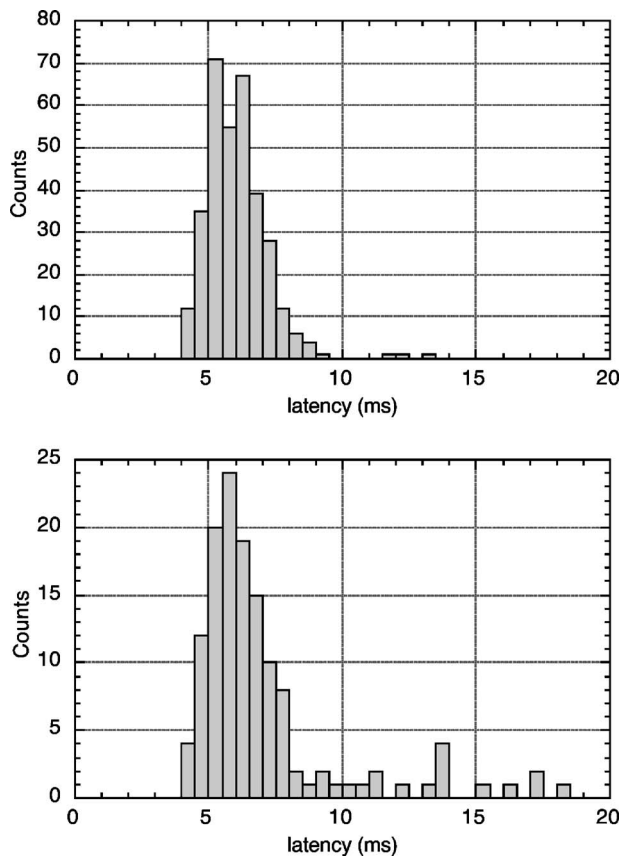
FIG. 2. Experimental histograms of the 3 kHz latency for the "pass" population (top) and for the "fail" population (bottom). The "fail" histogram shows a significant tail extending to unreasonably long latency values.

FIG. 3. Individual latency at 3 kHz (top) of the "fail" population as a function of reproducibility in the same 500 Hz band. Latency data are the same, as shown in previous Fig. 2, bottom panel. Abnormally long latency values are systematically associated with low reproducibility. The same pattern is shown for the 2.5 kHz band (bottom).

whose individual latency is unreasonably long, which does not affect the position of the maximum of the function. In other words, there is evidence for a systematic error associated with the determination of high-frequency individual latencies, more frequent for low reproducibility data, which seriously affects the average latency defined as the average of the individual latency distribution, leaving almost unchanged the peak time of the corresponding average wavelet band coefficient. On the other hand, if we could recover quasi-Gaussian distributions for the individual latencies of the "fail" population, we would get a meaningful statistical comparison of the two populations. These abnormal individual latency values are mainly due to the fact that the individual high-frequency wavelet coefficients often show a periodic structure associated with subsequent reflections in the cochlea (Sisto and Moleti, 2002), and the first peak may sometimes be lower than the following ones, due to the combined effect of acquisition window and noise. Very low reproducibility data are most sensitive to these effects, thus we have filtered the individual latency data, cutting out the lowest reproducibility data of the "fail" population. By doing so, we have restricted our data analysis to the comparison of two populations discriminated by a much smaller, yet significant, difference in band reproducibility. After having computed the time at which each wavelet band coefficient reaches its maximum value for each ear, which is our definition of individual spectral latency, we have discarded the latency values relative to bands with poor reproducibility in that band. This

selection has not significantly affected the "pass" data, whose reproducibility is very high in all bands by definition, but turned out to be important to get meaningful results for the "fail" population, as we will discuss in detail in the next section. The band reproducibility threshold was fixed at 0 at 1 kHz, at 0.35 in the frequency range 1.5–3 kHz, and at 0.7 for the 3.5 and 4 kHz bands. This particular choice is motivated by the spectral shape of the band reproducibility, whose median is plotted in Fig. 5 for the two populations of ears, with and without the selection, and by the particularly high sensitivity of the high-frequency data to the above mentioned systematic error. From Fig. 5, one can see that, after this selection, the reproducibility difference between the two populations is very small in the two highest frequency bands.

This selection rule was sufficient to eliminate most of the unreasonably long latency values, producing quasisymmetrical histograms also for the selected "fail" population, named "fail sel," as shown in Fig. 6. In this way meaningful statistical evaluations have become possible. Average latency data $\langle \tau^*(f) \rangle$ are shown in Fig. 7, for the two populations, computed with the above mentioned selection criterion. The application of the selection criterion has significantly modified the average latency of the "fail" population, by cutting out roughly one-third of the latency data, and the latency estimates are consistent with those obtained with the alternative definition of latency, shown in Fig. 1.

As shown in Table I, the average latency difference between the two populations becomes very small after the selection, with a significant difference in the 1.5 and 2.5 kHz
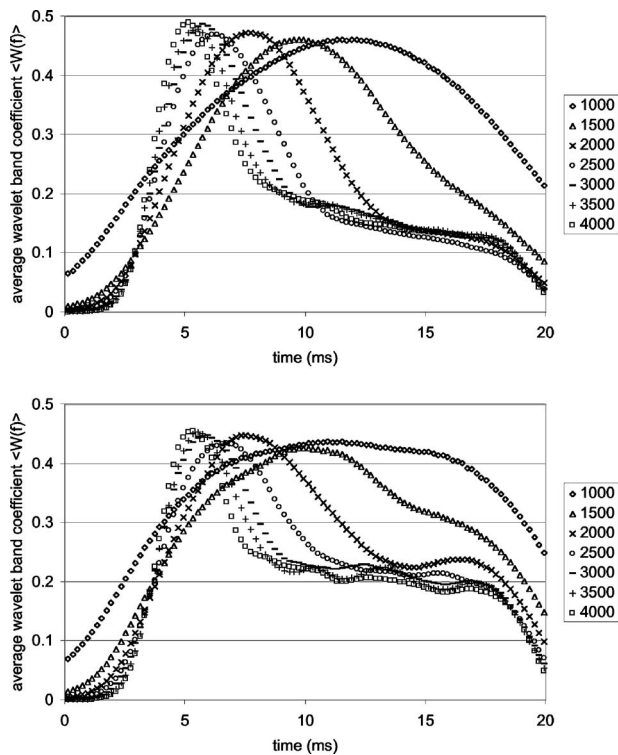
FIG. 4. Average wavelet band coefficients of the "pass" (top) and "fail" (bottom) population. The plateau on the long latency side of high-frequency components is associated with the systematic error in the determination of the average latency τ^*, but does not affect the latency τ, given by the time of the peak of these curves.

bands only. A systematic latency excess is, however, observed in the "fail sel" population over the whole frequency range. We would like to stress that the large number of ears makes the standard errors very small, as shown in Table I and Fig. 7, and that our conservative data selection choice ensures that the statistical distributions can be meaningfully compared, as shown by the shape of the selected data histograms (Fig. 6).

The selected latency data were best fitted to a power law of the frequency, Eq. (5). We recall that τ_{nc} was assumed to
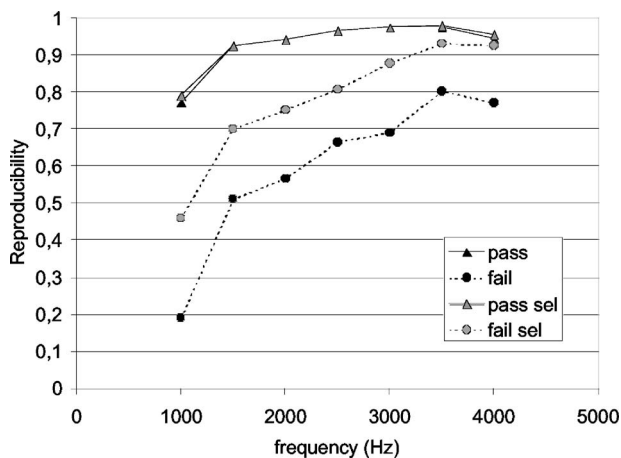
FIG. 5. A Comparison between the median of the band reproducibility distributions of the "pass" and "fail" populations, and of the "pass sel" and "fail sel" subsamples obtained with the selection criterion based on band reproducibility, as described in the text.
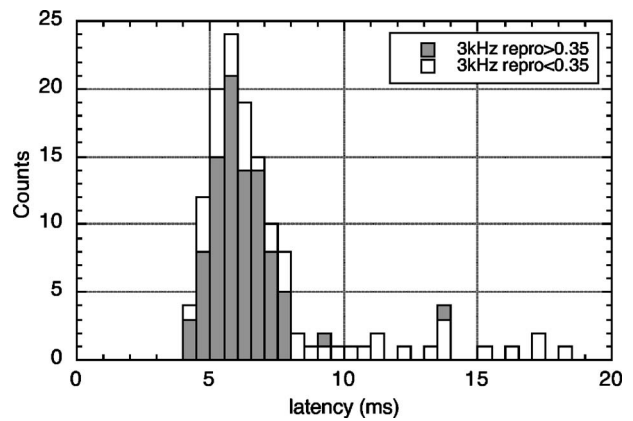
FIG. 6. Experimental histograms of the 3 kHz latency for the "fail" population, before and after the data selection. The dark region of the histogram represents the selected latency subset, "fail sel," surviving the selection procedure based on band reproducibility. A quasi-Gaussian shape is recovered.

be 0.5 ms, according to Sisto and Moleti (2002) and Moleti and Sisto (2003). With this assumption, the power law best fit computed on the average selected latency data $\langle \tau^*(f) \rangle$, yielded $a = 11.2$ ms, $b = -0.61$ for the "pass" ears and $a = 11.5$ ms, $b = -0.59$ for the "fail sel" ears. Using the alternative estimate of the average latency $\langle \tau(f) \rangle$, one gets, respectively, $a = 11.4$ ms, $b = -0.68$ for the "pass" ears and $a = 11.5$ ms, $b = -0.64$ for the "fail" ears.

The estimate of the tuning curve $Q(f)$ obtained by applying Eq. (7) to the experimental latency data is shown in Fig. 8 for the two populations of ears. In the case of $Q(\tau^*(f))$, standard errors are also plotted to show that the difference between the two populations is marginally significant over the whole frequency range. In the same figure are also shown the values of $Q(f)$ obtained from the alternative definition of average latency $\langle \tau(f) \rangle$, which are slightly lower, and follow the same trend.

IV. DISCUSSION

The large number of ears examined in this study provides a rather precise estimate of the average spectral latency
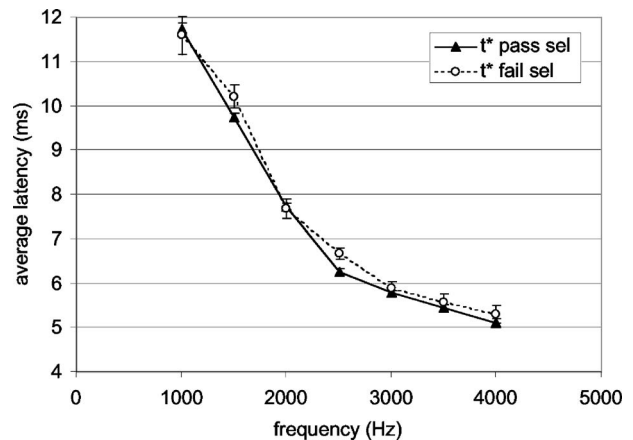
FIG. 7. A comparison between the average latency $\langle \tau^*(f) \rangle$ of the "pass sel" (triangles) and of the "fail sel" (circles) neonate ears, selected according to the procedure based on band reproducibility.

TABLE I. Average latencies $\tau^*(f)$ [see Eq. (3)] of the "pass sel" and "fail sel" populations and p values associated with the difference between them. The latency values obtained without data selection (τ^*_{pass} and τ^*_{fail}) and using the alternative definition of average latency (τ_{pass} and τ_{fail}) [see Eq. (4)] are also shown for comparison. The threshold value $p<0.05$ has been used as a conventional criterion for statistical significance.

f(Hz)	1000	1500	2000	2500	3000	3500	4000
$\tau^*_{\text{pass sel}}$	11.75±0.15	9.51±0.10	7.65±0.08	6.51±0.07	6.03±0.06	5.83±0.09	5.59±0.10
$\tau^*_{\text{fail sel}}$	11.80±0.44	10.28±0.27	7.99±0.23	6.79±0.12	6.12±0.13	5.95±0.19	5.80±0.19
$p_{\text{value sel}}$	n.s.	0.008	n.s.	0.041	n.s.	n.s.	n.s.
τ^*_{pass}	11.63	9.52	7.65	6.51	6.03	5.83	5.57
τ^*_{fail}	11.28	10.67	8.87	7.92	6.99	7.30	6.96
τ_{pass}	11.72	9.74	7.72	6.26	5.78	5.44	5.1
τ_{fail}	11.6	10.22	7.68	6.68	5.9	5.58	5.3

of normal hearing neonates, with a very small experimental uncertainty, in regard to random errors. Unfortunately, the latency estimate is subject to systematic errors, which have been discussed in the previous sections. These systematic errors tend to affect the latency distributions, producing a significant number of unreasonably long latency values at high frequencies.

As described in the previous sections, two alternative definitions of average latency were used, which, in previous studies (Sisto and Moleti, 2002; Lucertini *et al.*, 2002) on the OAE latency of normal-hearing and impaired adults, had given almost coincident results. The large difference observed in the present work between the results obtained for the "fail" population with the two alternative definitions of average latency is a clear indication of the mentioned systematic error, which seems to be characteristic of the neonate OAEs. Indeed, the observed large latency difference between "pass" and "fail" ears using no data selection criterion turned out to be mostly due to a small set of these unreasonably long high-frequency latency values in the "fail" population. This systematic error source was eliminated by restricting the analysis to the bands in which the "fail" ears showed a relatively high reproducibility. This conservative choice made the difference between the average band reproducibility of the two populations very small, thus it is not surprising that the residual difference between the average latency of the two populations has become correspondingly small. In

FIG. 8. An OAE-based estimate of the average cochlear tuning curves $Q(f)$ of the two populations, based on the two definitions of average latency $\langle\tau^*(f)\rangle$ (computed on the selected subsample) and $\langle\tau(f)\rangle$.

Sisto and Moleti (2002), a much larger latency difference was observed, but in that study the two adult populations were discriminated by actual audiometric tests, with an average hearing threshold difference between 10 and 40 dB between the two populations. In this work, the discrimination is based on a screening parameter (reproducibility), which is only statistically correlated to cochlear functionality, so the neonates of our "fail" population, are actually "normal hearing" subjects with a slightly lower cochlear response. Moreover, the "fail sel" data have a relatively good reproducibility (see Fig. 5). Yet a significant difference has still been observed, between two neonate populations that were not very different from each other.

We remark that we are discussing the relation between two observable quantities: the TEOAE reproducibility, correlated to response level, which is what discriminates between our two populations, and the latency. They are dependent, respectively, on the gain and on the quality factor of the nonlinear cochlear amplifier, which are not trivially related to each other. Our finding that the latency slightly increases in "fail" ears is due to the correlation between these two cochlear amplifier parameters. According to transmission line cochlear models (e.g., Moleti and Sisto, 2003, and references therein), this effect is associated with a difference in cochlear tuning between the two samples. Any further interpretation depends on the particular nonlinear model assumed for describing the cochlear response. Most nonlinear oscillator models (e.g., Tubis and Talmadge, 1998; Sisto and Moleti, 1999) involve a nonlinear damping term, which implies that Q (at sound levels well above the hearing threshold) is higher when the excitation level is lower. At a given stimulus level, this means higher Q for lower gain, in agreement with the present results. In the case of substantial hearing loss, as in Sisto and Moleti (2002) and in Lucertini *et al.* (2002), other less trivial mechanisms could be involved: the reduction of saturation of the response associated with the nonlinear damping term could also be due to a reaction of the auditory system to the threshold increase, which modifies the parameters of the feedback mechanisms to try to compensate for hearing loss. These rather speculative hypotheses should be tested by further more complex experiments on hearing impaired subjects, also involving other OAE techniques, and variable stimulus levels.

From a clinical diagnostic point of view, this study shows that the application of latency measurements to the
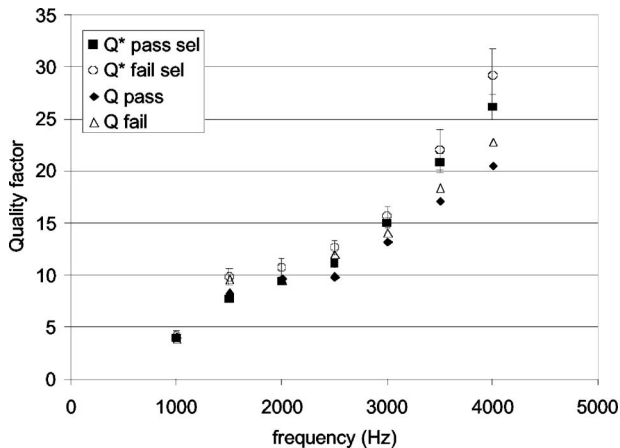
early detection of cochlear damage in neonates is a difficult task, partly because the interpretation is not straightforward, partly because the measured difference is rather small, and becomes significant only after averaging over a large population, and partly because neonatal OAEs seem to be intrinsically more sensitive to the systematic errors described in this work, leading to unreasonably long latency estimates at high frequency. This may also be a consequence of the larger high-frequency content and SOAE prevalence in the neonatal OAE waveforms and of the generally higher quality factor of neonatal cochlear response, which makes it more probable the occurrence of high-frequency multiply reflected OAE signals.

The observed dependence of latency on frequency is in agreement with a power law with a negative exponent, significantly smaller than unity (−0.61 and −0.59 for "pass" and "fail sel" ears, respectively). This slope, which is a clear indication of cochlear scale-invariance symmetry breaking (Sisto and Moleti, 2002; Moleti and Sisto, 2003), is in agreement with experimental latency estimates coming from previous TEOAE, DPOAE, SFOAE, and ABR studies.

From a normative point of view, it should be stressed that the estimate of the "pass" average latency was obtained on a large and homogeneous sample of full-term neonates. Thus it could be used as a reference, e.g., for comparing the spectral latency of premature neonates, using the same data analysis technique.

From a technical point of view, we have evidenced that, particularly in the case of neonates, much care must be given to the detection of possible systematic errors associated with the individual latency estimate. These errors partly arise from the necessity of using an acquisition window to cut off the linear ringing artifact. The use of the nonlinear acquisition paradigm strongly limits the linear ringing artifact, yet its residual fraction requires the use of data windowing. This limits the applicability of the method to 4 kHz, because the higher-frequency (and shorter latency) cochlear response is significantly affected by the window onset. In the case of long-lasting emissions, which are quite common in neonates, the reduction of the early high-frequency response due to the window can make the second peak of the wavelet coefficient, associated with double reflection along the cochlea, higher than the first one. This leads to mistakenly attributing the peak time (our definition of latency) to the second peak, whose delay is roughly double the correct one. The presence of strong spontaneous emissions, which is quite common in neonates, should also be considered, in a dedicated study, as a possible source of systematic error.

In this work we have controlled the rate of systematic errors in the determination of individual latencies by selecting relatively high reproducibility data, but it would be interesting to check if the application of different acquisition or stimulation paradigms could limit the effect of the linear ringing perturbation. This would permit anticipation of the window start time, extending the latency measurements to a higher-frequency range. This possibility would be very important, because the neonate OAE spectrum has a large high-frequency component, and, more generally, because sensorineural hearing loss is typically first observed at frequencies of 4 kHz and above. An improvement in the acquisition techniques would allow more reliable comparisons between populations with greater different average values of reproducibility, in order to confirm or disclaim the present results.

The estimates of the cochlear tuning curve shown in Fig. 8 imply an increase of tuning with frequency in the 1–4 kHz range. This result is similar to that obtained by Moleti and Sisto (2003), using a much smaller independent TEOAE dataset, and confirms the increase of the quality factor $Q(f)$ with frequency that had been observed in other studies, based on DPOAE suppression tuning curve data (Kummer *et al.*, 1995; Gorga *et al.*, 2003) and on Stimulus Frequency OAE measurements (Shera *et al.*, 2002; Shera and Guinan, 2003).

V. CONCLUSIONS

TEOAE latency data of a large statistical sample of "pass" (333) and "fail" (133) neonate ears are presented in this study. Latency was measured with an optimized time-frequency wavelet analysis of TEOAE nonlinear data, with a frequency resolution of 500 Hz, between 1 and 4 kHz. Despite the fact that the two populations did not differ very much in their average reproducibility, a small yet significant difference between the average latencies of the two populations has been found, which could be associated with better tuning at high stimulus levels in ears with poorer cochlear response. The results of this study resemble the TEOAE latency increase observed in hearing impaired adults (Sisto and Moleti, 2002), and also in audiometrically "normal" ears of subjects exposed to noise (Lucertini *et al.*, 2002), but the interpretation of all these findings is not model independent, and needs more complex experiments to be clarified. In particular, experiments using different stimulus levels would be very useful to test the predictions of any given nonlinear cochlear model. An improved data acquisition and/or analysis technique would also be needed to get reliable latency information in the high-frequency range (>4 kHz). The latency values of the "pass" population are in agreement with previous estimates of the neonatal OAE latency (Moleti and Sisto, 2003), and may be effectively used for statistical comparisons in future studies involving different populations (infants, adults, impaired neonates, etc.). In this study we also provide new estimates of the cochlear tuning curve in neonates.

Abdala, C. (**2001**). "DPOAE suppression tuning: cochlear immaturity in premature neonates or auditory aging in normal-hearing adults?," J. Acoust. Soc. Am. **110**, 3155–3162.

Attias, J., Furst, M., Furman, V., Reshef, I., Horowitz, G., and Bresloff, I. (**1995**). "Noise-induced otoacoustic emission loss with or without hearing loss," Ear Hear. **16**, 612–618.

Bray, P., and Kemp, D. T. (**1987**). "An advanced cochlear echo technique suitable for infant screening," Br. J. Audiol. **21**, 191–204.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. F. (**2003**). "Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **114**, 263–278.

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

Hall, A. J., and Lutman, M. E., (**1999**). "Methods for early identification of noise-induced hearing loss," Audiology **38**, 277–280.

Hussain, D. M., Gorga, M. P., Neely, S. T., Keefe, D. H., and Peters, J., (**1998**). "Transient evoked otoacoustic emissions in patients with normal hearing and in patients with hearing loss," Ear Hear. **19**, 434–449.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**) "Identification of otoacoustic emissions components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). "Acoustic emissions cochleography: practical aspects," in *Cochlear Mechanics and Otoacoustic Emissions*, Scandinavian Audiology (Suppl. 25), edited by G. Cianfrone and F. Grandori, pp. 71–95.

Kowalska, S., and Sulkowski, W. (**1997**). "Measurements of click-evoked otoacoustic emissions in industrial workers with noise-induced hearing loss," Int. J. Occup. Med. Environ. Health **10**, 441–459.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). "Suppression tuning characteristics of the 2f1-f2 distortion product otoacoustic emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 972–978.

Mallat, S. G. (**1989**). "A theory for multiresolution signal decomposition: the wavelet representation," IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693.

Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.

Moulin, A., and Kemp, D. T. (**1996**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans," I. General characteristics, J. Acoust. Soc. Am. **100**, 1617–1639.

Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.

Prijs, V. F., Schneider, S., and Schoonhoven, R. (**2000**). "Group delays of distortion product otoacoustic emissions: relating delays measured with f1- and f2-sweep paradigms," J. Acoust. Soc. Am. **107**, 3298–3307.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.

Shera, C. A., and Guinan, Jr., J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.

Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**(5), 3318–3323.

Shera, C. A., Talmadge, C. L., and Tubis, A. (**2000**). "Interrelations among distortion-product phase-gradient delays: their connection to scaling symmetry and its breaking," J. Acoust. Soc. Am. **108**, 2933–2948.

Sisto, R., and Moleti, A. (**1999**). "Modeling otoacoustic emissions by active nonlinear oscillators," J. Acoust. Soc. Am. **106**, 1893–1906.

Sisto, R., and Moleti, A. (**2002**). "On the frequency dependence of the otoacoustic emission latency in hypoacoustic and normal ears," J. Acoust. Soc. Am. **111**, 297–308.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Tognola, G., Grandori, F., Avan, P., Ravazzani, P., and Bonfils, P. (**1999**). "Frequency-specific information from click evoked otoacoustic emissions in noise-induced hearing loss," Audiology **38**, 243–250.

Tognola, G., Grandori, F., and Ravazzani, P. (**1998**). "Wavelet analysis of click-evoked otoacoustic emissions," IEEE Trans. Biomed. Eng. **45**, 686–697.

Tognola, G., Ravazzani, P., and Grandori, F. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.

Tubis, A., and Talmadge, C. L. (**1998**). "Ear canal reflectance in the presence of spontaneous otoacoustic emissions. I. Limit-cycle oscillator model," J. Acoust. Soc. Am. **103**, 454–461.

Vohr, B. R., and Maxon, A. B. (**1996**). "Screening infants for hearing impairment," J. Pediatr. **128**, 710–714.

Wit, H. P., van Dijk, P., and Avan, P. (**1994**). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141–147.

Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Two modes of motion of the alligator lizard cochlea: Measurements and model predictions

A. J. Aranyosi
Speech and Hearing Biosciences and Technology Program, Harvard-MIT Division of Health Sciences and Technology and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Dennis M. Freeman[a]
Speech and Hearing Biosciences and Technology Program, Harvard-MIT Division of Health Sciences and Technology, Research Laboratory of Electronics, and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114

Measurements of motion of an *in vitro* preparation of the alligator lizard basilar papilla in response to sound demonstrate elliptical trajectories. These trajectories are consistent with the presence of both a translational and rotational mode of motion. The translational mode is independent of frequency, and the rotational mode has a displacement peak near 5 kHz. These measurements can be explained by a simple mechanical system in which the basilar papilla is supported asymmetrically on the basilar membrane. In a quantitative model, the translational admittance is compliant while the rotational admittance is second order. Best-fit model parameters are consistent with estimates based on anatomy and predict that fluid flow across hair bundles is a primary source of viscous damping. The model predicts that the rotational mode contributes to the high-frequency slopes of auditory nerve fiber tuning curves, providing a physical explanation for a low-pass filter required in models of this cochlea. The combination of modes makes the sensitivity of hair bundles more uniform with radial position than that which would result from pure rotation. A mechanical analogy with the organ of Corti suggests that these two modes of motion may also be present in the mammalian cochlea. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1993147]

I. INTRODUCTION

The transformation from transverse motion of the basilar membrane (BM) to shearing displacement of the reticular lamina is one of the largest unresolved issues in cochlear mechanics. Most cochlear models assume that the cochlea has multiple modes of motion, where each mode represents a natural vibration of the system. Recent experiments show evidence of multiple modes of motion at the BM (Nilsen and Russell 1999, but see Cooper 2000), reticular lamina (Ulfendahl *et al.* 1995), tectorial membrane (Hemmert *et al.* 2000), outer hair cells (Karavitaki and Mountain 1997), inner hair cells (Mountain and Cody 1999), and organ of Corti (Richter and Dallos 2003; Scherer and Gummer 2004), as well as in neural responses (Lin and Guinan 2000, 2004), and otoacoustic emissions (Lukashkin and Russell 2003) of mammalian cochleae. The complexity of the mammalian cochlea and the major contribution of active nonlinear processes to cochlear mechanics allow for a potentially rich set of modes, making modal analysis difficult. However, we can gain some insight into cochlear modes by analyzing the motion of a simpler cochlea.

In contrast to the mammalian cochlea, the alligator lizard basilar papilla is relatively simple (Mulroy 1974). In response to sound, the BM moves in phase along its entire length, with tuning similar to that of the stapes (Peake and Ling 1980; Manley *et al.* 1988). A mechanical model of the alligator lizard cochlea assumes that the basilar membrane drives rotation of the basilar papilla (Weiss and Leong 1985). Two experimental studies qualitatively confirmed this assumption (Frishkopf and DeRosier 1983; Holton and Hudspeth 1983). However, both of those studies noted an increase in the motion of the basal end of the papilla above 3 kHz, suggesting that the motion of the papilla is more complex than pure rotation.

The present studies were undertaken to characterize the number and nature of modes of motion of the alligator lizard basilar papilla. These modes were characterized by measuring the motion of this structure in response to sound stimulation. The measurements demonstrate that the basilar papilla undergoes both a translational and a rotational mode of motion simultaneously, and that both modes can contribute to cochlear sensitivity. An analogy with the organ of Corti suggests that similar modes may be present in the mammalian cochlea as well.

II. METHODS

The methods used to collect and analyze data have been described in a previous publication (Aranyosi and Freeman

[a]Electronic mail: freeman@mit.edu
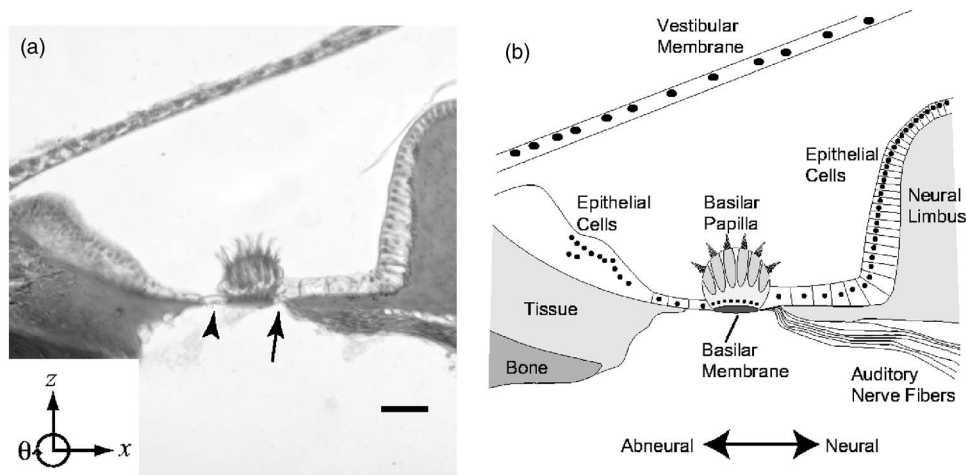
FIG. 1. Histological cross-section (**a**) and labeled drawing (**b**) of the alligator lizard basilar papilla. The basilar papilla is circular in cross section and rests on a stiff basilar membrane, which appears dark in the figure and is coupled to surrounding tissue by epithelial cells. The attachment is marked with an arrow on the neural side, and with an arrowhead on the abneural side. The methylene blue stain highlights the difference between the basilar membrane and the epithelial cells. The entire basilar papilla is about 400 μm long in the direction orthogonal to this image, so that the flat area surrounding the basilar papilla is oval-shaped when viewed in the z direction. The footplate of the columella is located above the vestibular membrane, is parallel to the BM, and moves in the z direction. The axis in the lower left defines the directions of x, z, and Θ which are used throughout this paper. The x axis is positive in the neural direction, and the z axis is positive in the direction pointing from scala tympani to scala media. Scale bar=50 μm.

2004). We present a brief summary here. The alligator lizard cochlea was carefully excised and placed in an experiment chamber. This chamber clamped the cochlea over a hole separating an endolymphlike apical fluid from a perilymph-like basolateral fluid. The experiment chamber was designed to allow high-resolution imaging of the clamped cochlea. A piezoelectric disk coupled to the basal fluid generated pressure differences across the BM to drive cochlear motion. Sound stimuli were delivered at 120 dB sound pressure level (SPL) in fluid, corresponding to 85–100 dB SPL at the tympanic membrane, depending on frequency (Rosowski *et al.* 1985). All animal procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care.

To image cochlear motions, a green light-emitting diode was strobed with a 1/8 duty cycle at the frequency of the sound stimulus. The resulting image, which shows the position of the cochlea at a chosen phase of the sound stimulus, was captured with a charge coupled device camera. Eight such images were taken at eight evenly spaced stimulus phases. This process was repeated at 60–100 focal planes separated by 1 μm. The resulting data set at each stimulus frequency was a sequence of three-dimensional images taken at different stimulus phases, where each voxel in the images has dimensions of $0.19 \times 0.19 \times 1.0$ μm^3. Computer vision algorithms (Horn and Weldon 1988; Davis and Freeman 1998; Timoner and Freeman 2001) were used to measure the three-dimensional motion of structures from the images. Noise in the measurements was due to ambient vibrations, drift of the tissue, and noise inherent in the imaging system. This noise was aliased by the measurement process to the stimulus frequency and its harmonics. Most measurements showed primarily sinusoidal motion. In addition, when the light source was strobed to alias motion to 2–3 Hz, the motions seen by eye in the microscope were primarily sinusoidal. Consequently, to eliminate noisy measurements, we included measurements only when higher harmonics of mo-

tion were at least 10 dB smaller than the fundamental component. Papilla motion was examined qualitatively in twelve preparations, and quantitatively in five of these.

III. MEASUREMENTS OF BASILAR PAPILLA MOTION

A. Anatomy of the basilar papilla

Although in many ways similar to the mammalian cochlea, the basilar papilla of the alligator lizard has a relatively simple structure. Figure 1 shows a cross-sectional view of the basilar papilla, with several structures labeled. The basilar papilla itself is roughly cylindrical, and contains hair cells and supporting cells. The hair bundles of hair cells project into scala media, with their sensitive axis parallel to the curved apical surface of the basilar papilla (Shotwell *et al.* 1981). The BM contains a rigid plate at the base of the basilar papilla, but this plate does not extend across the entire BM. Instead, epithelial cells attach the basilar papilla to surrounding tissue on both sides. The footplate of the columella, analogous to the mammalian stapes, sits above the plane of the image in Fig. 1 and lies parallel to the basilar membrane (Mulroy 1974). Thus unlike in the mammalian cochlea, motions of the footplate and basilar membrane are oriented in the same direction. Figure 1 illustrates the convention for directions used in this paper.

B. Displacement versus lateral position

Figure 2 shows trajectories of displacement for four locations on the body of the basilar papilla in one preparation. The image in the background shows a radial (xz) cross section of the papilla, with the neural side on the right. The circles show displacement at each location at each of eight phases of the 5 kHz sound stimulus; the filled circles indicate the displacement at the peak of the electrical stimulus driving the piezoelectric disk. At each location, the papilla moves in
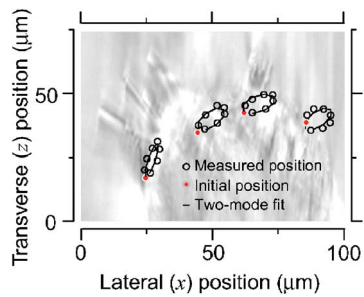
FIG. 2. (Color online) Two-dimensional trajectories of displacement of the basilar papilla. The background of this plot shows a cross-sectional view of the basal end of the basilar papilla, with the neural side to the right. The BM is just below the bottom of this image. Circles show the displacement of four points on the papilla as a function of stimulus phase in response to a 5 kHz tone. The displacements are plotted at the location at which they were measured, but the motions are magnified by a factor of 5 to improve visibility. The filled circle indicates the measured position at the peak of the electrical drive to the piezoelectric bimorph. Lines represent the least-squares fit of two-mode motion to the data as described in Sec. IV. The fit had 0.79 μm z translation, 0.85 μm x translation, and 1.1° rotation about the center of the papilla. Rotation and x translation were in phase with each other, and lagged z translation by 58°. The rms error of the fit was 0.130 μm.

an elliptical pattern. Displacements in each direction are nearly sinusoidal; higher harmonics are more than 40 dB below the fundamental component. The solid lines in the figure denote the best fit of a rigid-body model of papilla displacement to the measurements. This fit is discussed further in Sec. IV.

Figure 3 shows displacement trajectories of the basal end of the papilla from a total of 142 nonoverlapping $5 \times 5 \times 5 \ \mu m^3$ regions in five preparations at frequencies between 3 and 4 kHz. In each preparation, motion at each point is elliptical, although the shapes of these ellipses vary across preparations. For a given preparation, the shapes of the elliptical motions vary with position in a systematic manner. The major axes of the ellipses are largely in the z direction near the abneural edge of the papilla, with an x component that becomes larger toward the neural edge. In addition, the ellipses are more circular near the neural edge than near the abneural edge. When viewed with the neural side of the papilla on the right, the elliptical motion is typically in a clockwise direction; that is, x displacement lags z displacement. Motion in the y direction, which corresponds to the longitudinal axis of the cochlea, was typically near the noise floor of the measurement system.

C. Phase of x relative to z displacement

The measurements of Fig. 3 can be described as the sum of sinusoidal displacement in the x and z directions. Similar measurements were made for about 1000 nonoverlapping regions with volumes of $(5 \ \mu m)^3$ throughout each basilar papilla at each frequency studied. A total of 72 such data sets were analyzed across the five papillae studied, yielding 75 594 total measurements. Figure 4 shows the phase lag of x relative to z displacement for each of the 72 data sets used in this study, as well as a combined average of all measurements. For 75.4% of the measurements, x displacement lags z displacement. For only 4.6% of the measurements, x displacement leads z displacement by more than 22.5°. There is a trend for the phase lag of x displacement to increase with frequency.
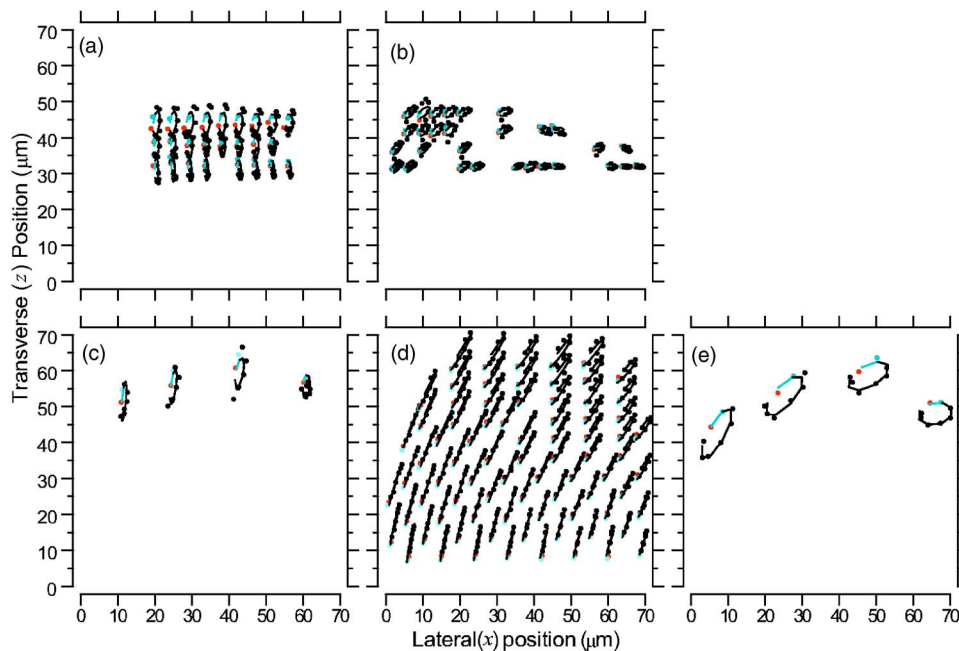
FIG. 3. (Color online) Elliptical motion of the basilar papilla in five preparations. These plots show the measured displacement of multiple locations at the basal end of the basilar papilla. In each plot, the neural edge of the papilla is to the right. The motion at each location is primarily elliptical. The tallest ellipses are typically near the abneural edge, and the widest are typically near the top center of the papilla. Symbols are as in Fig. 2. Dark gray (red) dots are the position at the peak of the electrical drive to the piezoelectric bimorph; light gray (cyan) dots are the position at the next measured phase. Motions are magnified by a factor of 10 relative to positions, except in **d**, for which the magnification is a factor of 5. Stimulus frequencies are 3000 Hz (**a**), 3464 Hz (**b**), 3746 Hz (**c**), 3834 Hz (**d**), and 3981 Hz (**e**). Solid lines show the best fit of a rigid-body model, described in Sec. IV, to the entire set of measurements for each papilla. The light gray (cyan) portion of the line connects the first two phases. Measurements at some locations were omitted for clarity.
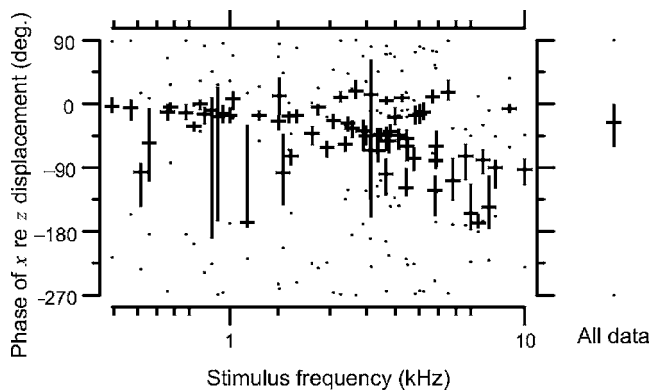
FIG. 4. Phase of *x* relative to *z* displacement as a function of frequency. For each papilla at each frequency measured, displacements in the *x* and *z* directions were measured for many (roughly 1000) nonoverlapping regions, and the phase of *x* relative to *z* displacement was determined. The median, interquartile range, and full range of these phase values are plotted for 72 such data sets obtained at frequencies from 500 Hz to 10 kHz from five papillae. The horizontal bar for each histogram indicates the median value, the vertical bar indicates the interquartile range, and the dots represent the full range of measurements. Phases above 90° or below −270° are wrapped into that range. The median values and interquartile ranges are almost all at or below 0°, indicating that *x* displacement typically lags *z* displacement. The bar to the right of the plot pools measurements across all papillae and frequencies. A total of 75 594 phase measurements are included in this bar, of which 56 976 have a relative phase less than zero.

IV. MODAL DECOMPOSITION OF BASILAR PAPILLA MOTION

In Fig. 3, the motion at any location can be approximated by sinusoidal motion in the *x* and *z* directions. However, these motions vary systematically with position, suggesting that the motion of the entire papilla can be described as the sum of a translational and a rotational mode. To perform this modal decomposition, the measured displacements at all locations in a single longitudinal cross section were fit in a least-squares sense by rigid-body motion of the basilar papilla. This rigid-body fit had three degrees of freedom: translation in the *x* and *z* directions and rotation in the Θ direction, where the center of rotation was chosen arbitrarily to be the center of the basilar papilla. These fits were computed separately for each frequency at which measurements were made. These three degrees of freedom were reduced to two by shifting the center of rotation for all frequencies based on two criteria. First, since translation in the *x* direction was typically in phase with rotation, it was assumed that these two components of motion represented a single mode. Accordingly, the center of rotation was shifted in the *z* direction to minimize the sum-squared magnitude of the *x* translational component across frequency. Second, since measurements of basilar membrane motion *in vivo* show little frequency dependence (Peake and Ling 1980), the center of rotation was shifted in the *x* direction to minimize the variance in magnitude of the *z* component across frequency.

Figure 5 shows the result of this modal decomposition process for the five preparations of Fig. 3. Although there is considerable variability across preparations, this variability consists largely of vertical offsets in the magnitude; that is, some preparations moved more than others. Aside from these changes in magnitude, no systematic differences between
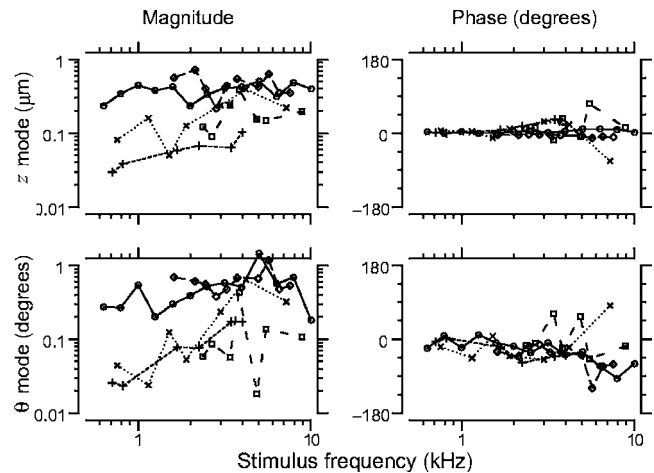
FIG. 5. Modal decomposition of basilar papilla motion. These plots show the magnitude (left) and phase (right) of the *z* (top) and Θ (bottom) modes of motion of five preparations as a function of frequency. The *z* mode magnitude is largely independent of frequency, and the phase is near 0°. The Θ mode magnitude peaks between 4 and 5 kHz, and the phase increasingly lags with frequency. All phases for each papilla are plotted relative to the phase of the *z* mode at the lowest frequency measured.

preparations were seen. The origin of these magnitude differences is unknown. For this reason, with the exception of Secs. V B and VI C, the remainder of this paper is focused on the relative contribution of each mode as a function of frequency rather than the absolute motions.

The magnitude of the *z* translational mode is largely independent of frequency for most papillae. The phase of the *z* translational mode is near 0° at all frequencies for all preparations, relative to the phase of *z* translation at the lowest frequency measured. The magnitude of the Θ mode consistently has a local maximum near 4–5 kHz. In addition, Θ is in phase with *z* at low frequencies and increasingly lags with increasing frequency. The center of rotation is typically on the medial side of the basilar papilla, above the level of the BM. The differing frequency dependence of the *z* and Θ components of motion indicates that these motions constitute two separate modes rather than a single complex mode.

V. A MECHANICAL MODEL OF THE BASILAR PAPILLA

A. Deriving the model from anatomy and measurements

The anatomy of the alligator lizard basilar papilla suggests a simple mechanical model which is consistent with the results shown in Secs. III and IV. The attachment of the BM and basilar papilla to the surrounding tissue is asymmetric; the neural side is about half as wide and twice as thick as the abneural side (see Fig. 1). Consequently, pressure across the BM can induce both translation and rotation of the basilar papilla. Figure 6 shows a simplified mechanical representation of the basilar papilla which demonstrates how two modes of motion can arise. The stiffness of the attachment on the neural side of the basilar papilla is larger than that on the abneural side. The applied pressure is equal on both sides, causing the abneural side to move more than the neural side. Consequently, the basilar papilla undergoes both translation
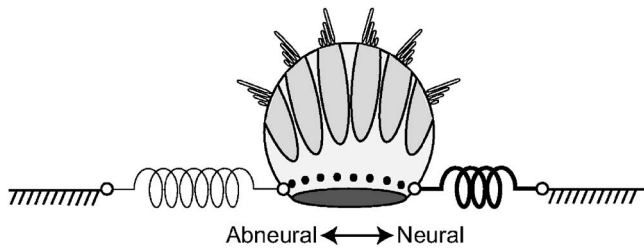
FIG. 6. Mechanical representation of basilar papilla attachment. The basilar papilla and basilar membrane are attached to surrounding tissue in an asymmetric fashion (Fig. 1). The tissue on the neural side is thicker and shorter than on the abneural side. In this figure, each side is represented by a spring that can pivot at both attachments. The spring on the neural side is stiffer and shorter than on the abneural side, as represented by the thicker line. Force exerted on the basilar membrane will cause the two springs to be stretched by different amounts, allowing the basilar papilla to both translate and rotate. Although the cochlear fluids are expected to contribute viscous and inertial impedance, such contributions are not shown in this drawing.

and rotation. According to this model, both translation and rotation are second-order resonant systems, but the frequency dependence of the two modes may differ.

The frequency dependence of the two modes of motion can be represented by an electrical circuit as shown in Fig. 7. In this representation, the translational and rotational modes have been explicitly separated. Both the current measurements and those of previous studies (Peake and Ling 1980; Rosowski et al. 1985) suggest that the translational impedance of the BM is compliant, so the contributions of mass and damping to impedance in the translational domain have been ignored. The compliance C_m of the BM in the translational domain is the proportionality constant relating the applied force $F = PA$ at the basilar membrane to the resulting velocity V_{BM}; that is, C_m represents the net compliance of the BM. The applied pressure also acts in the rotational domain to exert a torque $\tau = PAL$ on the basilar papilla, where the transformer ratio L is an effective lever arm that determines the scaling from force to torque. Because the rotational mode is mechanically resonant, this portion of the circuit includes a moment of inertia J_r, a rotational compliance C_r, and a rotational damping R_r.

B. Estimating model parameter values from measurements

The model in Fig. 7 has six parameters: C_m, C_r, J_r, R_r, A, and L. By estimating the two transformer ratios A and L,
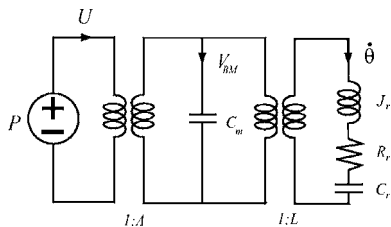
FIG. 7. Circuit model of the basilar papilla. The circuit model contains three domains. In the acoustic domain, sound pressure P generates a volume velocity U. U acts on the BM area A to generate a velocity V_{BM} across the BM compliance C_m in the translational domain. Because of the asymmetry of the BM attachments, represented by an effective lever arm L, the pressure P also acts in the rotational domain to generate a rotational velocity $\dot{\theta}$. This rotational velocity acts on a moment of inertia J_r, a rotational compliance C_r, and a rotational damping R_r.

the other parameter values can be determined from measurements. Because of the variability in motion magnitude across preparations, a range of parameter values is estimated here. The area A of the BM has been measured to be 0.03 mm^2 (Rosowski et al. 1985). The effective lever arm L can be estimated by representing the applied pressure P with an equivalent force at the center of the basilar membrane, and measuring the distance from this center to the center of rotation of the basilar papilla. Based on the center of rotation determined by the modal fits, $L \approx 30~\mu$m.

Given the input pressure $P = 120$ dB SPL, the compliance $C_m = |z|/P \cdot A$ can be estimated from the average magnitude of the z component of motion, which is between 0.1 and 0.4 μm for most preparations. The resulting compliance is $0.12 \leq C_m \leq 0.48$ m/N. Similarly, $C_r = |\Theta_{LF}|/P \cdot A \cdot L$. The low-frequency asymptote of rotation $|\Theta_{LF}|$ is between 0.05° and 0.5°, so $3.5 \times 10^7 \leq C_r \leq 3.5 \times 10^8$ rad/N m. Assuming a second-order resonance, $J_r = 1/C_r(2\pi f_p)^2$, which for a resonant frequency of 5 kHz yields $4 \times 10^{-18} \leq J_r \leq 4 \times 10^{-17}$ kg m^2. At f_p, the net impedance of C_r and J_r is zero, so $R_r = P \cdot A \cdot L/2\pi f_p |\Theta_p|$. With $0.1° < \Theta_p < 1°$, the resulting estimates of R_r range from 5×10^{-14} to 5×10^{-13} N m s/rad. The damping R_r can also be estimated from the quality of tuning $Q \approx 1/C_r R_r \omega$. The shallow peak in magnitude of Θ and slow rolloff of phase with frequency seen in Fig. 5 suggest that $Q < 2$. Using $C_r = 10^8$ rad/N m, $R_r > 10^{-13}$ N m s/rad.

VI. DISCUSSION

A. Gross motions are comparable to those measured *in vivo*

Peake and Ling (1980) measured the velocity of the BM in the transverse (z) direction using the Mössbauer technique and found it to be a linear function of sound pressure. In response to a 75 dB SPL tone at 1.4 kHz, they measured a velocity of −25 dB re 1 cm/s, which corresponds to a peak displacement of 90 nm. With an estimated middle-ear pressure gain of 35 dB at this frequency (Rosowski et al. 1985), we expect our stimuli to correspond to 85 dB SPL at the eardrum, which would result in a peak BM z displacement of 280 nm based on the Peake and Ling (1980) data. Figure 3 shows peak z displacements of 250–600 nm near the abneural edge of the basilar papilla. These measurements were made near 5 kHz, where the rotational mode contributes significantly to z displacement, so the displacements are expected to be smaller near 1.4 kHz. Thus the z motions measured in this study are roughly comparable to those measured *in vivo* using a different technique. However, there are many caveats to comparing these studies directly: Our comparison uses a rough estimate of the middle ear pressure gain; in some preparations, displacements in the transverse direction were considerably smaller than in others (Fig. 3); the efficiency of coupling sound pressure to the papilla in our experiment chamber may differ from that *in vivo*; Peake and Ling (1980) measured motion on the perilymphatic surface of the BM, but we measured motion on the endolymphatic surface of the basilar papilla; The Mössbauer source used by Peake and Ling (1980) was $\sim 40 \times 40~\mu$m, much larger than
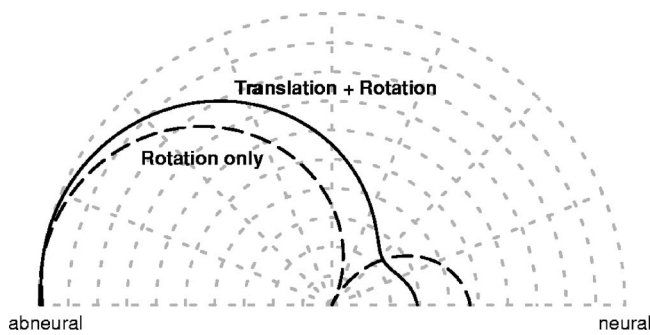
FIG. 8. Relative sensitivity of hair bundles to sound as a function of position. The plot shows the shearing displacement of the basilar papilla (r axis) as a function of position along the surface of the basilar papilla (θ axis), normalized by the maximum value of each curve. The dashed line shows the predicted shearing displacement if the basilar papilla underwent pure rotation about a point near the neural edge of the BM. The solid line shows the normalized shearing displacement predicted from the two-mode motion used to fit the measurements of Fig. 1. Except near the neural edge, where there are no hair bundles, the shearing displacement due to the sum of translation and rotation is more uniform with position than that due to rotation alone.

our measurement regions—the motion of the $(5 \ \mu\text{m})^3$ regions measured here may differ from the average motion of a larger region. Nonetheless, the reasonable agreement between the magnitude of motion measured *in vivo* and *in vitro* supports the idea that the mechanical properties measured here are relevant to normal hearing.

B. The relative sensitivity of different hair bundles is made more uniform by two-mode motion

Hair bundle deflection in this cochlea is believed to be driven by shearing motion at the reticular lamina (Freeman and Weiss 1990) in the direction of maximum sensitivity of hair bundles (Shotwell *et al.* 1981). Because of the circular shape of the papilla in cross section, the direction of shearing displacement at the surface of the papilla changes with lateral position on the papilla (Mulroy 1974). Thus, the drive for hair bundle deflection depends not only on the magnitudes and phases of the two modes of papilla motion, but on the lateral position of each hair bundle.

Rotation of the basilar papilla generates a shearing displacement that is a strong function of lateral position. Since the center of rotation is on the neural side of the basilar papilla, in response to a given rotation the motion of the abneural side is larger. In addition, this motion is parallel to the surface of the reticular lamina on the abneural side, but perpendicular to the surface near the neural side. As a result, the sensitivity of hair cells to the rotational mode of motion of the basilar papilla varies dramatically with radial position (Fig. 8). When both modes of motion are included, the sensitivity of hair cells is less dependent on position. The translational mode generates shearing forces for hair bundles not excited by the rotational mode and vice versa. Consequently, at nearly all locations which have hair bundles, the relative shearing displacement is more uniform with position for two-mode motion than for pure rotation. Thus, the combination of two modes causes the hair bundles to have a more uniform sensitivity to sound across the width of the basilar papilla.

C. Model parameters are physiologically reasonable

Rosowski *et al.* (1985) describe measurements of the acoustic impedance at the tympanic membrane before and after destroying the cochlear partition. Those measurements assign the BM an acoustic compliance of $5.7 \times 10^{-16} \ \text{m}^3/\text{Pa}$. When reflected into the mechanical domain using their measured $A=0.03 \ \text{mm}^2$, the mechanical compliance is $0.63 \ \text{m/N}$; using their "effective BM area" of $0.07 \ \text{mm}^2$ gives a mechanical compliance of $0.12 \ \text{m/N}$. When the rotational compliance C_r is reflected into the mechanical domain and added to the translational compliance C_m, the net compliance ranges from 0.15 to 0.80 m/N. Thus, the range of compliances estimated in the current study is roughly consistent with those seen in measurements at the tympanic membrane.

Since the basilar papilla is roughly cylindrical, the moment of inertia J_r can be compared to that of a rotating cylinder, where the center of rotation is a distance L from the center of mass; namely, $J_r = \rho l[(\pi r^4/2) + \pi r^2 L^2]$, where ρ is the mass density of the cylinder, l is the length, and r is the radius. Assuming that the basilar papilla length is about $400 \ \mu\text{m}$, the density is that of water ($\rho=1000 \ \text{kg/m}^3$), and the center of rotation is near the neural edge ($L=30 \ \mu\text{m}$), our estimate of J_r yields an effective basilar papilla radius of $35 \leq r \leq 60 \ \mu\text{m}$. This effective radius is somewhat larger than the anatomical radius of the basilar papilla, which is roughly $30-35 \ \mu\text{m}$. This difference is most likely due largely to the entrained motion of fluid. Since the center of rotation is not the center of the papilla, rotation of the papilla induces bulk displacement of fluid. The hair bundles that project from the basilar papilla also entrain fluid. Moreover, at 5 kHz (the frequency at which J_r was estimated), fluid viscosity creates a boundary layer about $8 \ \mu\text{m}$ thick that moves with the papilla (Freeman and Weiss 1988). This entrained fluid increases both the effective radius r and the effective distance L from the center of rotation to the center of mass; both effects would cause an increase in our estimate of r.

The damping of the rotational mode, however, is larger than the damping expected for a cylinder moving through fluid. The fluid force on a cylinder oscillating in fluid is a function of the cylinder diameter and length and the oscillation frequency (Freeman and Weiss 1985). For a cylinder of $30 \ \mu\text{m}$ diameter with a length of $400 \ \mu\text{m}$ oscillating at 5 kHz, the fluid impedance is roughly $\mu \times 4 \times 10^{-2} \ \text{N s/m}$, where μ is the dynamic viscosity of the fluid, assumed to be that of water. However, more than half of this impedance is inertial and contributes to the effective mass of the cylinder, as described above. The viscous component of this impedance, about 40% of the total, is $1.8 \times 10^{-5} \ \text{N s/m}$. The minimum estimate of the rotational damping of the basilar papilla was $5 \times 10^{-14} \ \text{N m s/rad}$; reflecting this value into the translational domain gives $5.5 \times 10^{-5} \ \text{N s/m}$. Thus the total damping of basilar papilla rotation is at least three times larger than that predicted from the motion of a cylinder. This difference suggests that fluid drag due to motion of the basilar papilla is not the primary contributor to viscous damping in this cochlea.

A. Aranyosi and D. M. Freeman: Two modes of lizard cochlear motion

One additional source of viscous damping is in the motion of hair bundles relative to cochlear fluids. Based on a hydrodynamic model of such motion (Freeman and Weiss 1990), which predicts the sharpness of mechanical tuning of hair bundles in this cochlea (Aranyosi and Freeman 2003), the viscous impedance to hair bundle rotation is $4.56 \times 10^{-3} \, h^2/w$ N s/m, where h is the height of the hair bundle and w is its width. The free-standing region of the basilar papilla contains about 150 hair bundles with heights ranging from about 12 to more than 30 μm and widths of about 8 μm (Mulroy 1974). The net viscous impedance of these hair bundles as predicted by the hydrodynamic model is on the order of $4-8 \times 10^{-5}$ N s/m. This value suggests that much of the damping seen in the rotational mode of motion in the current study results from the rotation of hair bundles through a viscous fluid. This finding is significant, since the motion of hair bundles relative to cochlear fluids determines the sensitivity and frequency selectivity of hearing in much of this cochlea (Aranyosi and Freeman 2004). If a principal source of energy loss in this cochlea is viscous damping at the level of hair bundles, then the force generated by hair bundle motility (Martin and Hudspeth 1999; Ricci *et al.* 2000) is ideally situated to counteract this viscous damping and sharpen tuning in this cochlea (Manley *et al.* 2001; Aranyosi and Freeman 2005). Additionally, although the reduction in viscous damping due to hair bundle motility is likely to cause only a small change in the impedance seen at the basilar membrane, such a change may be large enough to explain the presence of otoacoustic emissions in lizard species (Manley 1997; Manley *et al.* 2001).

D. Phase lag between translational and rotational modes—the "missing low-pass filter"

The resonant frequency of the rotational mode of basilar papilla motion is typically near 4–5 kHz and is largest at the basal end of the papilla. Hair cells in this region have best frequencies between 3 and 4 kHz. Thus, the papilla resonance at the basal end is just beyond the range of frequencies to which hair cells in this cochlea are most sensitive. At 5 kHz, hair cells with best frequencies of 3–4 kHz are 20–40 dB less sensitive than at the best frequency (Weiss *et al.* 1976). Thus, the mechanical resonance of the papilla is ideally positioned to increase the high-frequency slopes of neural tuning curves without altering frequency selectivity in the range of best frequencies.

Previous models of the alligator lizard cochlea required a first-order mechanical low-pass filter to explain the high-frequency slopes of tuning curves from neurons innervating the free-standing region (Weiss and Leong 1985; Freeman and Weiss 1990; Shatz 2000). However, these models did not specify the physical basis of this filter. The measurements and model reported here show that at high frequencies, the rotational mode of motion is attenuated relative to the translational mode. This attenuation provides a physical explanation for the required low-pass filter. However, since at least some hair bundles are sensitive to the translational mode of motion, variations of as much as 6 dB/octave should be present in the high-frequency slopes of tuning curves of au-

ditory nerve fibers. Such variations have been seen in auditory nerve fiber recordings (Weiss *et al.* 1976).

E. Implications for mammalian cochlear mechanics

Although more complex than that of the alligator lizard, the mammalian BM also has two regions of differing stiffnesses, separated by the base of the outer pillar cell—the arcuate zone under the tunnel of Corti and the pectinate zone which extends from the outer pillar cell to the spiral ligament. However, the nature of the difference between zones is not entirely clear. Stiffness measurements made in the basal region of the guinea pig cochlea showed that the arcuate zone was about five times stiffer than the pectinate zone (Miller 1985). Measurements in the gerbil cochlea showed the opposite trend (Olson and Mountain 1994; Naidu and Mountain 1998). In the gerbil hemicochlea, the arcuate zone is less stiff in the basal turn, but stiffer in the middle turn (Emadi *et al.* 2004). Measurements of basilar membrane motion in the guinea pig show that these two zones can move in antiphase (Nilsen and Russell 1999, 2000). Other measurements in guinea pigs and gerbils do not show antiphase motion, but do show an asymmetric vibration pattern (Cooper 1999, 2000; Rhode and Recio 2000; Ren 2002). Such a pattern is roughly consistent with the motion of a beam that is clamped at the outer spiral lamina and simply supported at the inner spiral lamina (Cooper 1999; Homer *et al.* 2004). The fit is improved if the beam is made less stiff in the arcuate zone. Thus although many details of the differences have not yet been clearly established, it is clear that both the stiffness and attachment of the basilar membrane differ in the two zones. These differences suggest that the organ of Corti is likely to undergo both translation and rotation. Such an interaction of multiple modes of motion appears to be necessary to explain responses of auditory nerve fibers (Lin and Guinan 2004). In addition, force generation by outer hair cells may preferentially excite one mode. For these reasons, characterizing these modes more precisely is critical to understanding cochlear function.

Aranyosi, A. J., and Freeman, D. M. (**2003**). "Measured and modeled motion of free-standing hair bundles in response to sound stimulation," *Biophysics of the Cochlea: From Molecules to Models*, A. W. Gummer, ed. (World Scientific, Singapore), pp. 81–88.

Aranyosi, A. J., and Freeman, D. M. (**2004**). "Sound-induced motions of individual cochlear hair bundles," Biophys. J. **87**, 3536–3546.

Aranyosi, A. J., and Freeman, D. M. (**2005**). "Negative damping by hair bundles can sharpen tuning in the alligator lizard cochlea," *Abstracts of the Twenty-Eighth Midwinter Meeting of the Association for Research in Otolaryngology*, New Orleans, LA.

Cooper, N. P. (**1999**). "Radial variation in the vibrations of the cochlear

partition," *Recent Developments in Auditory Mechanics*, H. Wada, T. Takasaka, K. Ikeda, K. Ohyama, and T. Koike, eds. (World Scientific, Singapore), pp. 109–115.

Cooper, N. P. (**2000**). "Radial variation in the sound-evoked vibrations of the cochlear partition," *Abstracts of the Twenty-Third Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, FL.

Davis, C. Q., and Freeman, D. M. (**1998**). "Statistics of subpixel registration algorithms based on spatio-temporal gradients or block matching," Opt. Eng. (Bellingham) **37**, 1290–1298.

Emadi, G., Richter, C.-P., and Dallos, P. (**2004**). "Stiffness of the gerbil basilar membrane: Radial and longitudinal variations," J. Neurophysiol. **91**, 474–488.

Freeman, D. M., and Weiss, T. F. (**1985**). "On the role of fluid inertia and viscosity in stereociliary tuft motion: Analysis of isolated bodies of regular geometry," *Lecture Notes in Biomathematics 64: Peripheral Auditory Mechanisms*, J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, and A. Tubis, eds. (Springer, Berlin), 147–154.

Freeman, D. M., and Weiss, T. F. (**1988**). "The role of fluid inertia in mechanical stimulation of hair cells," Hear. Res. **35**, 201–208.

Freeman, D. M., and Weiss, T. F. (**1990**). "Hydrodynamic analysis of a two-dimensional model for micromechanical resonance of free-standing hair bundles," Hear. Res. **48**, 37–68.

Frishkopf, L. S., and DeRosier, D. J., (**1983**). "Mechanical tuning of free-standing stereociliary bundles and frequency analysis in the alligator lizard cochlea," Hear. Res. **12**, 393–404.

Hemmert, W., Zenner, H. P., and Gummer, A. W. (**2000**). "Three-dimensional motion of the organ of Corti," Biophys. J. **78**, 2285–2297.

Holton, T., and Hudspeth, A. J. (**1983**). "A micromechanical contribution to cochlear tuning and tonotopic organization," Science **222**, 508–510.

Homer, M., Champneys, A., Hunt, G., and Cooper, N. (**2004**). "Mathematical modeling of the radial profile of basilar membrane vibrations in the inner ear," J. Acoust. Soc. Am. **116**, 1025–1034.

Horn, B. K. P., and Weldon, Jr., E. J. (**1988**). "Direct methods for recovering motion," Int. J. Comput. Vis. **2**, 51–76.

Karavitaki, K. D., and Mountain, D. C. (**1997**). "Multiple modes of vibration in the apical turn of the gerbil cochlea," *Abstracts of the Twentieth Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg Beach, FL.

Lin, T., and Guinan, Jr., J. J., (**2000**). "Auditory-nerve-fiber responses to high-level clicks: Interference patterns indicate that excitation is due to the combination of multiple drives," J. Acoust. Soc. Am. **107**, 2615–2630.

Lin, T., and Guinan, Jr., J. J. (**2004**). "Time-frequency analysis of auditory-nerve-fiber and basilar-membrane click responses reveal glide irregularities and non-characteristic-frequency skirts," J. Acoust. Soc. Am. **116**, 405–416.

Lukashkin, A. N., and Russell, I. J. (**2003**). "A second low-frequency mode of vibration in the intact mammalian cochlea," J. Acoust. Soc. Am. **113**, 1544–1550.

Martin, P., and Hudspeth, A. J. (**1999**). "Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli," Proc. Natl. Acad. Sci. U.S.A. **96**, 14306–14311.

Manley, G. A. (**1997**). 'Diversity in hearing-organ structure and the characteristics of spontaneous otoacoustic emissions in lizards,' *Diversity in Auditory Mechanics*, E. R. Lewis, G. R. Long, R. F. Lyon, P. M. Narins, and C. R. Steele, eds. (World Scientific, Singapore), 32–38.

Manley, G. A., Yates, G., and Köppl, C. (**1988**). "Auditory peripheral tuning: Evidence for a simple resonance phenomenon in the lizard *Tiliqua*," Hear. Res. **33**, 181–190.

Manley, G. A., Kirk, D., Köppl, C., and Yates, G. K. (**2001**). "In-vivo evidence for a cochlear amplifier in the hair-cell bundle of lizards," Proc. Natl. Acad. Sci. U.S.A. **98**, 2826–2831.

Miller, C. E. (**1985**). "Structural implications of basilar membrane compliance measurements," J. Acoust. Soc. Am. **77**, 1465–1474.

Mountain, D. C., and Cody, A. R. (**1999**). "Multiple modes of inner hair cell stimulation," Hear. Res. **132**, 1–14.

Mulroy, M. J. (**1974**). "Cochlear anatomy of the alligator lizard," Brain Behav. Evol. **10**, 69–87.

Naidu, R. C., and Mountain, D. C. (**1998**). "Measurements of the stiffness map challenge a basic tenet of cochlear theories," Hear. Res. **124**, 124–131.

Nilsen, K. E., and Russell, I. J. (**1999**). "Timing of cochlear feedback: Spatial and temporal representation of a tone across the basilar membrane," Nat. Neurosci. **2**, 642–648.

Nilsen, K. E., and Russell, I. J. (**2000**). "The spatial and temporal representation of a tone on the guinea-pig basilar membrane," Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758.

Olson, E. S., and Mountain, D. C. (**1994**). "Mapping the cochlear partition's stiffness to its cellular architecture," J. Acoust. Soc. Am. **95**, 395–400.

Peake, W. T., and Ling, Jr., A. L., (**1980**). "Basilar-membrane motion in the alligator lizard: Its relation to tonotopic organization and frequency selectivity," J. Acoust. Soc. Am. **67**, 1736–1745.

Ren, T. (**2002**). "Longitudinal pattern of basilar membrane vibration in the sensitive cochlea," Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106.

Rhode, W. S., and Recio, A. (**2000**). "Study of mechanical motions in the basal region of the chinchilla cochlea," J. Acoust. Soc. Am. **107**, 3317–3332.

Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2000**). "Active hair bundle motion linked to fast transducer adaptation in auditory hair cells," J. Neurosci. **20**, 7131–7142.

Richter, C.-P., and Dallos, P. (**2003**). 'Micromechanics in the gerbil hemicochlea, *Biophysics of the Cochlea: From Molecules to Models*, A. W. Gummer, ed. (World Scientific, Singapore), pp. 271–277.

Rosowski, J. J., Peake, W. T., Lynch, T. J., Leong, R., and Weiss, T. F. (**1985**). "A model for signal transmission in an ear having hair cells with free-standing stereocilia: II. Macromechanical stage," Hear. Res. **20**, 139–155.

Scherer, M. P., and Gummer, A. W. (**2004**). "Vibration pattern of the organ of Corti up to 50 kHz: Evidence for resonant electromechanical force," Proc. Natl. Acad. Sci. U.S.A. **101**, 17652–17657.

Shatz, L. F. (**2000**). "The effect of hair bundle shape on hair bundle hydrodynamics of inner ear hair cells at low and high frequencies," Hear. Res. **141**, 39–50.

Shotwell, S. L., Jacobs, R., and Hudspeth, A. J. (**1981**). "Directional sensitivity of individual vertebrate hair cells to controlled deflection of their hair bundles," Ann. N.Y. Acad. Sci. **374**, 1–10.

Timoner, S. J., and Freeman, D. M. (**2001**). "Multi-image gradient-based algorithms for motion estimation," Opt. Eng. (Bellingham) **40**, 2003–2016.

Ulfendahl, M., Khanna, S. M., and Heneghan, C. (**1995**). "Shearing motion in the hearing organ measured by confocal laser heterodyne interferometry," NeuroReport **6**, 1157–1160.

Weiss, T. F., Mulroy, M. J., Turner, R. G., and Pike, C. L. (**1976**). "Tuning of single fibers in the cochlear nerve of the alligator lizard: Relation to receptor morphology," Brain Res. **115**, 71–90.

Weiss, T. F., and Leong, R. (**1985**). "A model for signal transmission in an ear having cells with free-standing stereocilia. III: Micromechanical stage," Hear. Res. **20**, 157–174.

Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials

Kelly L. McDonald
Rotman Research Institute, Baycrest Centre for Geriatric Care, Ontario, Canada

Claude Alain[a]
Rotman Research Institute, Baycrest Centre for Geriatric Care, Ontario, Canada
and Department of Psychology, University of Toronto, St-George Campus, Ontario, Canada

The contribution of location and harmonicity cues in sound segregation was investigated using behavioral reports and source waveforms derived from the scalp-recorded evoked potentials. Participants were presented with sounds composed of multiple harmonics in a free-field environment. The third harmonic was either tuned or mistuned and could be presented from the same or different location from the remaining harmonics. Presenting the third harmonic at a different location than the remaining harmonics increased the likelihood of hearing the tuned or slightly (i.e., 2%) mistuned harmonic as a separate object. Partials mistuned by 16% of their original value "pop out" of the complex and were paralleled by an object-related negativity (ORN) that superimposed the N1 and P2 components. For the 2% mistuned stimuli, the ORN was present only when the mistuned harmonic was presented at a different location than the remaining harmonics. Presenting the tuned harmonic at a different location also yielded changes in neural activity between 150 and 250 ms after sound onset. The behavioral and electrophysiological results indicate that listeners can segregate sounds based on harmonicity or location alone. The results also indicate that a conjunction of harmonicity and location cues contribute to sound segregation primarily when harmonicity is ambiguous. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000747]

I. INTRODUCTION

One of the primary functions of the auditory system is to assign incoming acoustic elements from different physical sound sources into perceptual auditory objects. This process, known as auditory scene analysis, requires parsing sounds that originate from different physical objects and grouping together sounds that emanate from the same physical sound source. Bregman (1990) proposed two classes of mechanisms that aid in determining which components belong to a particular sound source. "Primitive grouping mechanisms" use basic stimulus properties to segregate the incoming sounds. Listeners may be innately tuned to these properties and thus they do not depend on specific experiences. These grouping mechanisms most likely utilize general properties of sound sources, such as frequency, intensity, phase, harmonicity, and temporal coherence. "Schema-based mechanisms" use knowledge from past experience to group the acoustic input into meaningful mental representations. These schemata are thought to be learned and therefore are dependent on a listener's specific experiences.

The general problem of auditory scene analysis, and the more specific task of revealing those acoustic cues that are critical for the perception of distinct auditory objects, has received considerable attention over the last two decades (for reviews, see Alain and Arnott, 2000; Bregman, 1990; Car-lyon, 2004; Kubovy and Van Valkenburg, 2001). For example, complex sounds consisting of multiple tonal elements can be grouped and segregated on the basis of frequency periodicity (i.e., the harmonic series). Previous research has shown that the likelihood of hearing more than one concurrent auditory object increases as the amount of mistuning increases for lower tonal components of a harmonic series, resulting in the "pop out" of the lower mistuned partials from the complex (e.g., Chalikia and Bregman, 1989; Hartmann *et al.*, 1990; Lin and Hartmann, 1998; Moore *et al.*, 1986). Another factor that may contribute to concurrent sound segregation is the spatial distance between the two sound sources. For example, performance in identifying speech sounds improves with increasing spatial separation of concurrent streams of sounds (e.g., Arbogast *et al.*, 2002; Bronkhorst and Plomp, 1988, 1992; Drennan *et al.*, 2003). The perceived azimuth of a particular source is influenced by either interaural time differences (ITDs) or interaural level differences (ILDs). The former plays an important role for localizing lower frequency sounds whereas the latter appears to be more important in localizing high frequency sounds.

Few studies have examined the contribution of both harmonicity and location cues in parsing concurrent sounds and the findings that have been reported provide conflicting evidence about the relative importance of these two cues in segregating simultaneously present auditory events. For example, Gockel and Carlyon (1998) showed that the ability to detect mistuning was better when both the target harmonic

[a]Electronic mail: calain@rotman-baycrest.on.ca

and the remaining components were presented to the same ear than when they were presented to opposite ears (see also, Lee and Green, 1994). Interestingly, performance in detecting mistuning was comparable in a dichotic situation in which the perceived location of the remaining harmonics was contralateral to the target harmonic. However, using a short sound duration (i.e., 110 ms), Lee and Green (1994) reported similar thresholds in detecting mistuning for lower harmonics when the mistuned harmonic was presented ipsilateral or contralateral to the remaining harmonics. Buell and Hafter (1991) also found that the segregation of inharmonic components was little affected by ITD and that listeners were unable to segregate harmonic components even when the components were presented with different ITD. They concluded that tones that are harmonically related are combined into a single auditory object regardless of their spatial location and that harmonicity is a more compelling cue for object perception than common spatial location (for different results and conclusion see Stellmack and Dye, 1993). The listeners' task, however, was to identify the stimulus containing an ITD and participants were not asked to detect the number of sound sources that had been presented.

The current study aims to further investigate the relative contribution of harmonicity and location cues in parsing concurrent sounds using event-related brain potentials (ERPs). Recording ERPs is a powerful measure for examining the time course and level of processing at which concurrent sound segregation takes place. This technique allows for the examination of neural activity within hundreds of milliseconds following sounds. Because ERPs can be recorded to the same stimuli when they are either task-relevant or task-irrelevant, they provide a means to evaluate the effects of variables such as selective attention on concurrent sound segregation. Recent studies have identified a new ERP component referred to as the object-related negativity (ORN), whose amplitude correlates with listeners' likelihood of reporting two simultaneous auditory objects (Alain *et al.*, 2001a, 2002; Alain and Izenberg, 2003). The ORN is thought to index a mismatch process between the perceived frequency of the mistuned harmonic, and the frequency expected on the basis of a harmonic template extrapolated from the components of the incoming stimulus. The perception of the mistuned harmonic as a separate event is also associated with a late positive wave (P400). Like the ORN, its amplitude was correlated with perceptual judgment, being larger when participants perceived the mistuned harmonic as a separate tone. However, in contrast with the ORN, this component was present only when participants were required to respond whether they heard one or two auditory stimuli. Thus, whereas the ORN appears to be more associated with automatic processing, the P400 seems to be more related to controlled processes.

In the present study, ERPs were recorded during both active and passive listening conditions to investigate the degree to which harmonicity and location cues were processed automatically. Based on previous studies of concurrent sound segregation (Alain *et al.*, 2001a, 2002, 2003; Alain and Izenberg, 2003), it is expected that an ORN will be present in both the active and passive listening conditions and that the

P400 will be elicited only when participants are involved in the active detection of more than one simultaneous auditory event. Furthermore, it is expected that coincident harmonicity and location cues will increase the amplitude of both components as well as improve accuracy and reaction time in detecting concurrent sounds.

II. EXPERIMENT 1

A. Methods

1. Participants

Twelve right-handed adults (between 22 and 36 years of age, mean age$=28.58\pm4.14$ years; 6 males) participated in the experiment. All participants had pure-tone thresholds less than or equal to 30 dB HL at frequencies in the range 250–8000 Hz in both ears. Each participant gave informed consent in accordance with the University of Toronto and Baycrest Centre for Geriatric Care guidelines. Participants included volunteers from the University of Toronto and laboratory staff. All individuals were paid for their participation.

2. Stimuli and task

The stimuli consisted of complex sounds comprised of 10 pure tones of equal intensity. The fundamental frequency (f_0) was 200 Hz and stimulus duration was 400 ms, including a 10 ms rise-fall time. Stimuli were digitally generated at a sampling rate of 25 kHz using a System 3 Real-Time Processor from Tucker Davis Technology and presented over two EMI loudspeakers via a GSI 61 Clinical Audiometer. The loudspeakers were placed at a 45° angle from listeners' heads and the distance from each loudspeaker to the estimated center of participants' heads was 1.58 m while the distance between loudspeakers was 2.23 m. Stimuli were either "harmonic," where all of the harmonics were integer multiples of the f_0, or "mistuned," in which the third harmonic was shifted upwards or downwards by either 2% or 16% of its original value. It was expected that the harmonic with 16% mistuning would be easily segregated from the remaining complex while the 2% mistuning would represent a more ambiguous cue (Alain *et al.*, 2001b; Moore *et al.*, 1986). The third harmonic was played from either the same loudspeaker or a different loudspeaker than the other nine harmonics. The location of the target harmonic changed from trial to trial, eliminating the opportunity for participants to focus on the output of one loudspeaker to make their judgments. The output of each speaker was matched for intensity and the stimuli were presented at 65 dB SPL. Sound calibration measurements were made using a free-field microphone placed at the estimated center of a participant's head (Damilar SPL meter model 824).

Participants were tested individually in a double-walled sound attenuated chamber (IAC model 1204A). Each participant completed an active and passive listening session, which took place on two separate days (within a two week period). In the active listening session, a single complex tone was presented on each trial. Participants indicated whether they heard one sound (i.e., a buzz) or two sounds (i.e., a buzz plus another sound with a pure tone quality) following the presentation of each stimulus. Participants were instructed to

respond as quickly and accurately as possible by pressing one of two buttons on a response box using the index and middle fingers of their right hand. Participants did not receive any feedback on their performance. Within a block of trials, six different and equiprobable stimuli were played at random (three possible levels of mistuning −0%, ±2%, and ±16%; third harmonic presented from either the same or different loudspeaker as the remaining harmonics). A chin rest was used throughout the experiment to ensure that participants' heads remained in a constant position relative to the loudspeakers. The intertrial interval, i.e., the length of time between a participant's response and the next trial, was 1500 ms. Participants were presented with a small sample of trials prior to data collection to familiarize them with the task and response box. A total of 864 trials were presented to each participant (144 per stimulus type; 2 locations × 3 levels of mistuning).

In the passive listening session, the same stimuli were presented but participants were asked to ignore the stimuli and watch a subtitled movie of their choice. The interstimulus interval varied randomly between 800 and 1200 ms. A total of 1560 trials were presented to each participant (260 per stimulus type).[1]

The order of presentation of the active and passive listening sessions was counter-balanced across participants, with six participants completing the active listening condition first and the other six completing the passive session first.

3. Electrophysiological recording

Electrophysiological signals were digitized continuously (bandpass: 0.05–100 Hz; 500 Hz sampling rate) from an array of 64 electrodes, including those from the standard 10–20 placement, using NeuroScan Synamps and stored for offline analysis. During the recording, all electrodes were referenced to the Cz electrode; for data analysis they were re-referenced to an average reference. Electrodes placed at the outer canthi and at the superior and inferior orbit monitored vertical and horizontal eye movements.

The analysis epoch included 200 ms of prestimulus activity and 800 ms of poststimulus activity. Trials contaminated by excessive peak-to-peak deflection (±150 μV), excluding those channels adjacent to the eyes, were automatically rejected before averaging. ERPs were then averaged separately for each electrode site, stimulus type, and listening condition before being digitally low-pass filtered to attenuate frequencies above 15 Hz. For each individual average, ocular artifacts, such as blinks, saccades, and lateral movements, were corrected by ocular source components using Brain Electrical Source Analysis (BESA) software (Picton et al., 2000). ERPs to stimuli presented from the left loudspeaker (either all components of the stimulus or all harmonics except the third) were averaged separately from ERPs elicited by stimuli presented from the right loudspeaker. To improve the signal-to-noise ratio and to reduce the number of comparisons, the electrophysiological data were averaged over side of presentation of the harmonic series with electrode positions transposed for harmonic series presented from the right loudspeaker so that electrodes over the right

hemisphere were contralateral to the location of the harmonic series (e.g., electrode F4 was renamed electrode F3).[2] Then, trials in which the harmonic series were presented from the left and right loudspeakers were averaged together. For each participant, we used a weighted average procedure, which takes into account the number of trials in each individual average. Additionally, in order to facilitate the comparison between the active and passive listening conditions, the ERPs were averaged irrespective of the participant's responses in the active listening condition.

4. Data analysis

For the behavioral data, an analysis of variance (ANOVA) for repeated measures was performed on the proportion of trials where participants reported hearing two concurrent auditory objects with percentage mistuning (0%, 2%, and 16%) and location (same or different from the remaining harmonics) as factors. A similar analysis was performed on the response time data.

In most electrophysiological research, ERP analyses are performed on a subset of electrodes where the components of interest show the maximum amplitude. Although these analyses are acceptable, they do not take into account information available from the whole montage. This is particularly evident in situations where large arrays of electrodes are used to measure brain activity (e.g., 64 or 128 electrodes). Furthermore, it implies either an *a priori* decision regarding the electrodes that are entered in the analyses or an *a posteriori* decision based on visual inspection of the data. In the present study, we minimized such problems by comparing the source waveforms obtained for each participant and stimulus condition. Such a strategy has been used extensively to analyze data from neuromagnetic recordings of auditory stimuli and has the advantage of reducing a large number of channels to a smaller and more manageable number of sources (e.g., 2 dipole sources, one in the left and one in the right hemisphere).

The spatio-temporal source analyses were performed with the BESA 5.1 software package (MEGIS Software, GmbH, Munich). To enhance the signal-to-noise ratio, the fit procedure was based on the group mean ERPs collapsed across stimulus type. The analysis assumed a four-shell ellipsoidal head model with relative conductivities of 0.33, 0.33, 0.0042, and 1 for the head, scalp, bone, and cerebrospinal fluid, respectively, and sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness). As an initial step, two regional sources were seeded in left and right Heschl's gyrus (Talairach coordinates: X-Loc = ±47, Y-Loc = −26, Z-Loc = 11) to account for the neural activity between 0 and 260 ms after sound onset. Each source contained three orthogonal dipoles to account for all directions of current flow at the source location (tangential, radial, and anterior/posterior). Maintaining the symmetrical and orthogonal constraints, the orientations of the tangential sources were then aligned with the maximum direction of activity and the three sources were converted to single dipoles. Two additional regional sources were seeded in the left and right inferior parietal lobule (IPL) to account for the sustained potential between 300 and 800 ms post-stimulus. This area was chosen

TABLE I. Group mean responses (and standard error of the mean) as a function of harmonicity and location.

Percentage mistuning	Location	Percentage of stimuli heard as two sounds	Confidence intervals lower–upper bound	Response time (ms)	Confidence intervals lower–upper bound
0%	Same	20.31 (4.46)		737.88 (36.32)	
	Different	32.81 (4.57)		752.39 (39.26)	
	Difference	12.50 (4.18)	3.30–21.70	14.52 (14.86)	−18.19–47.22
2%	Same	69.44 (8.06)		752.85 (52.77)	
	Different	79.80 (7.04)		714.24 (56.70)	
	Difference	10.36 (3.76)	2.08–18.64	−38.62 (11.84)	−64.69−−12.55
16%	Same	96.01 (1.80)		607.72 (43.93)	
	Different	96.41 (1.70)		601.80 (40.49)	
	Difference	0.41 (0.50)	−0.70–1.51	−5.92 (7.20)	−21.77−9.94

because it has been found to be active during auditory localization tasks (Arnott *et al.*, 2004). The regional sources were then converted to single dipoles and the anterior-posterior component of the regional source in Heschl's gyrus was removed since it accounted for less than 1% of the total variance. As a final step, the orientation of the tangential and radial sources in Heschl's gyrus and the IPL sources were allowed to change. For the optimization of the radial and tangential sources in Heschl's gyrus, the IPL sources were turned off. The resulting model was held fixed and used as a spatial filter to derive the source waveforms (three in each hemisphere) in each participant and for all stimulus types and listening conditions, separately. The ANOVAs were performed on mean amplitude (nAm) over selected latency regions based on previous studies using similar stimuli (Alain *et al.*, 2001a; Alain and Izenberg, 2003; Alain *et al.*, 2002). The factors included in the ANOVA were percentage of mistuning (0%, 2%, 16%), location (same, different), and hemisphere (ipsi- or contralateral to the harmonic series).

B. Results

1. Behavioral data

Participants were more likely to report hearing two concurrent sounds when the third harmonic was mistuned, $F(2,22)=102.48$, $p<0.001$, all pairwise comparisons $p<0.01$ (Table I). Participants were also more likely to report hearing two separate auditory objects when the third and remaining harmonics were presented from different loudspeakers $F(1,11)=15.06$, $p<0.01$. There was a significant interaction between both harmonicity and location, $F(2,22)=4.19$, $p<0.05$. When the third harmonic was in tune or mistuned by only 2%, participants were more likely to indicate hearing two sounds when it was presented at a different location than the remaining harmonics, $F(1,11)>7.58$, $p<0.05$ in both cases. In comparison, the likelihood of reporting hearing two concurrent sounds was little affected by the location of the third harmonic when it was mistuned by 16% of its original value, $F(1,11)=0.65$, $p=0.44$. This is unsurprising given the proficiency of the participants at identifying two sounds when the third harmonic was mistuned by 16%.

We also examined the effects of location and harmonicity on response time for trials in which harmonic and inharmonic stimuli were identified as one or two objects, respec-

tively. Trials that were identified as two sounds when there was no mistuning and trials identified as one sound when mistuning was present were excluded from the analysis. A significant main effect of harmonicity was found for participants' mean response times, $F(2,22)=11.58$, $p<0.001$. Planned comparisons revealed that participants responded more quickly to the 16% mistuned stimuli than the 2% mistuned or tuned stimuli, $p<0.01$ for both. There was no significant difference in response times between tuned and 2% mistuned stimuli. Response times were also significantly faster when the third harmonic was presented at a different location than the remaining harmonics, $F(1,11)=11.76$, $p<0.01$. There was only a trend towards an interaction between harmonicity and location cues, $F(2,22)=3.71$, $p<0.06$. When the third harmonic was presented at a different location than the remaining harmonics, participants tended to respond more quickly when the third harmonic was mistuned but more slowly when it was in tune.

2. Electrophysiological data

a. Active listening. Figure 1(A) shows the grand average ERPs for 65 channels. The ERPs were comprised of N1 and P2 waves peaking at about 110 and 200 ms after sound onset, respectively. The N1-P2 was maximal over the fronto-
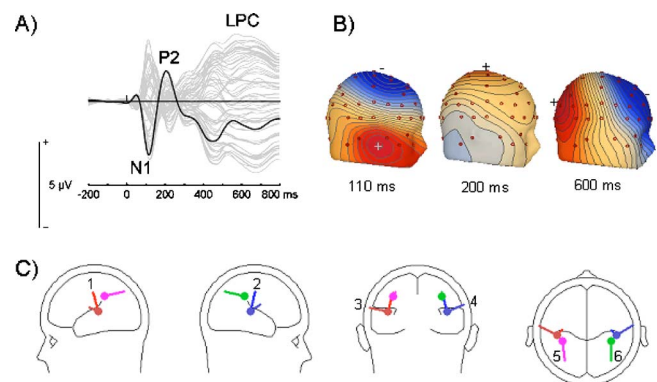
FIG. 1. (Color online) (A) Group mean auditory evoked potentials (for 65 channels) elicited by complex sounds averaged over stimulus type during active listening. The Cz electrode is shown in black whereas the other channels are shown in gray. (B) Iso contour maps are shown for the N1, P2, and LPC (late positive complex). The contour spacing was set at 0.30 μV. (C) Dipole source model of the grand average ERPs recorded during listening conditions.

A) Active Listening: 2%

B) Active Listening: 16%
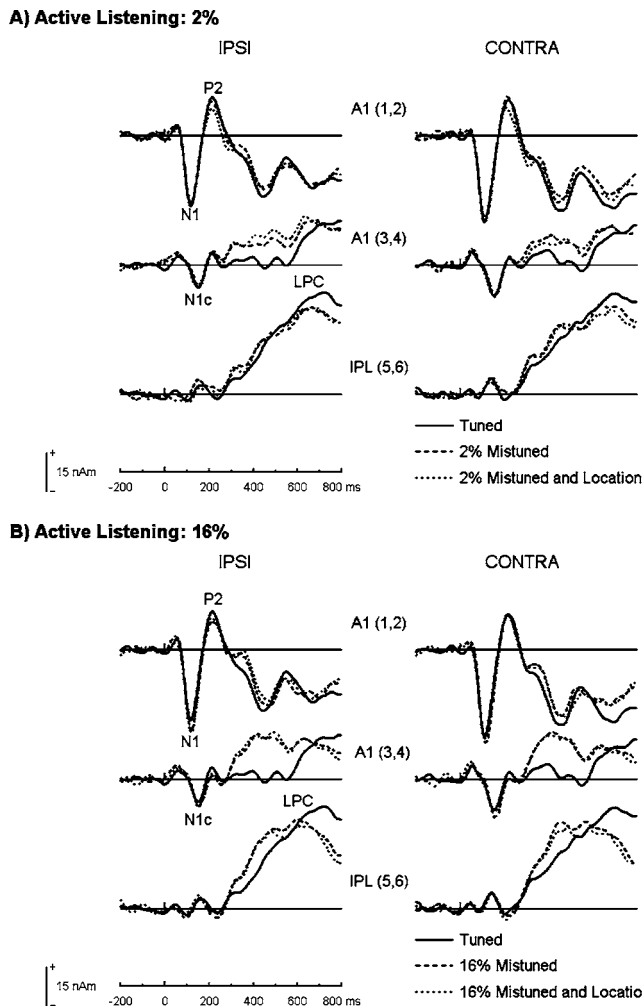
FIG. 2. Active listening. Source waveforms of the group mean ERPs recorded during active listening. (A) Source waveforms for the tuned and the 2% mistuned harmonic presented either at the same or at a different location than the remaining harmonics. (B) Source waveforms for the tuned and the 16% mistuned harmonic presented either at the same or at a different location than the remaining harmonics.

central scalp region and was followed by a sustained potential that was positive and maximal over the parietal regions [Fig. 1(B)]. This neural event is referred to as the late positive complex (LPC). The dipole model for the scalp-recorded data is illustrated in Fig. 1(C). The residual variance for the whole epoch was 4.08% for ERPs recorded during active listening and 3.70% for those obtained during passive listening, suggesting that this model provides a good fit to the scalp-recorded data.

Figure 2 shows group mean source waveforms when the mistuned harmonic was presented from the same or at a different location than the remaining harmonics during active listening. The N1 and P2 waves of the scalp-recorded data were best represented on the tangential sources whereas the N1c was expressed as radial sources. The LPC was accounted for by sources in IPL.

The N1 wave peaked earlier and its amplitude was larger in the hemisphere contra- (119 ms; -36.19 nA m) than ipsi-lateral (124 ms; -31.43 nA m) to the harmonic series [$F(1,11)=13.69$ and 15.57, respectively, $p<0.005$ in both cases]. The N1 peak latency and amplitude increased with

mistuning, $F(2,22)=4.36$ and 8.57, $p<0.05$ in both cases. This effect of mistuning on N1 latency and amplitude may be caused by the superimposition of the N1 and P2 wave by the ORN (see the following). A similar analysis on the N1c revealed a longer latency but larger amplitude in the hemisphere contra- (152 ms; -13.94 nA m) than ipsi-lateral to the harmonic series (146 ms; -10.47 nA m), $F(1,11)=5.62$ and 40.36, $p<0.05$ in both cases. There was also a tendency, of borderline significance, for the N1c to be earlier when the third harmonic was presented at a different location than the remaining harmonics [147 ms vs 150 ms for same location; $F(1,11)=4.24$, $p=0.06$]. The effect of mistuning on the N1c latency was not significant, but its amplitude increased with mistuning, $F(2,22)=4.57$, $p<0.05$. The effects of location on the N1 and N1c amplitude or latency were not significant.

The P2 peak latency and amplitude measured at the tangential source was quantified between 140 and 280 ms after sound onset. The main effect of mistuning or location was not significant. However, there was a significant interaction between mistuning and location on the P2 latency, $F(1,11)=9.53$, $p<0.01$. For the tuned stimuli, the P2 latency was earlier when the third harmonic was presented at a different location than the remaining harmonics (207 vs 222 ms, $t(10)=3.31$, $p<0.007$). For the 2% mistuned stimuli, the P2 latency was not affected by location whereas for the 16% mistuned stimuli the P2 was delayed when the third harmonic was presented at a different location than the remaining harmonics [228 vs 217 ms, $t(11)=2.30$, $p<0.05$]. For the P2 amplitude, the main effect of mistuning or location was not significant nor was the interaction between these variables.

The neural activity associated with distinguishing concurrent auditory objects is best illustrated by subtracting source waveforms to harmonic stimuli originating from the same location from those obtained for mistuned stimuli presented either at the same or different location than the remaining harmonics (Fig. 3). In the active listening condition, the difference wave revealed a biphasic response similar in latency with the scalp-recorded ORN and P400 waves.

The ORN and P400 latency and amplitude varied with the amount of mistuning and therefore the effect of location on these neural events was quantified over different latency windows. For the 2% mistuned stimuli, the first reliable difference in neural activity was noticed between 200 and 300 ms post-stimulus. An ANOVA on the mean amplitude from the tangential sources yielded a significant interaction between mistuning and location, $F(2,22)=5.68$, $p<0.02$. That is, the conjunction of harmonicity and location lead to significant enhancement in negativity but only for the 2% mistuned harmonic stimuli, $p<0.001$. There was no significant effect of location on the tuned or the 16% mistuned harmonic.

For the 16% mistuned stimuli, the first reliable difference in neural activity was noticed between 100 and 200 ms post-stimulus. An ANOVA on the mean amplitude from the tangential sources yielded a main effect of mistuning, $F(2,22)=12.56$, $p<0.001$. Planned comparisons revealed a significant ORN only when the third harmonic was mistuned by 16%, $p<0.001$. There was no significant difference in the
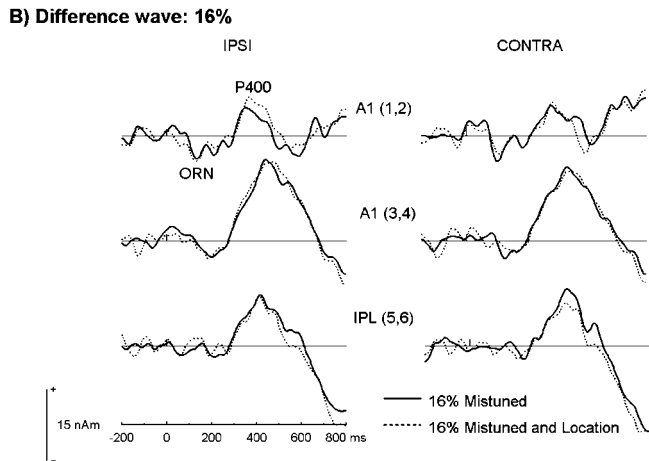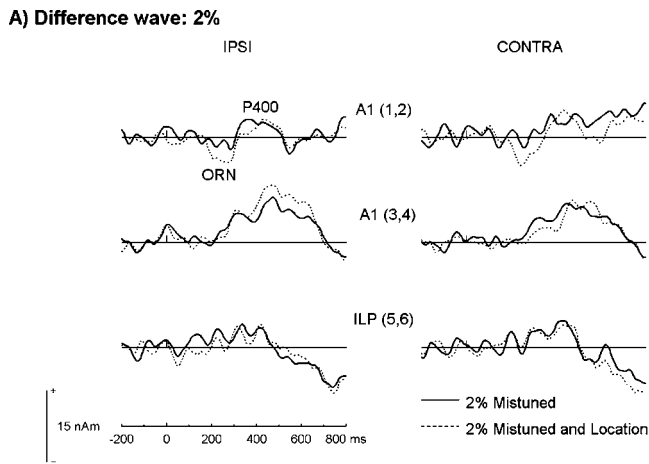
FIG. 3. Active listening. Object-related negativity isolated by subtracting source waveforms from the tuned harmonic presented at the same location from those obtained for mistuned stimuli presented either at the same or different location as the harmonic series. (A) ORN for the 2% mistuned harmonic. (B) ORN for the 16% mistuned harmonic.

source waveforms for the tuned and 2% mistuned stimuli. The main effect of location was not significant nor did location interact with other factors.

The ORN was followed by a P400 response peaking at about 450 ms after sound onset. An ANOVA on mean amplitude between 400 and 500 ms yielded a main effect of mistuning, $F(2,22)=15.54$, $p<0.001$. The P400 was larger for the 16% mistuned stimuli, intermediate for the 2% and smallest for the tuned stimuli, all pairwise comparisons significant at $p<0.02$. The main effect of location was not significant nor was the interaction between location and harmonicity. However, the interaction between location and hemisphere tended toward significance, $F(1,11)=3.66$, $p=0.08$.

b. Passive listening. Figure 4 shows group mean source waveforms when the mistuned harmonic was presented from the same or at a different location than the remaining harmonics during passive listening. As for the active listening condition, the N1 and P2 waves recorded during passive listening were best represented at the tangential sources. However, the N1c and the late positive complex were smaller and more difficult to identify than during active listening. Consequently, the analyses of the source waveforms were limited to the tangential component.
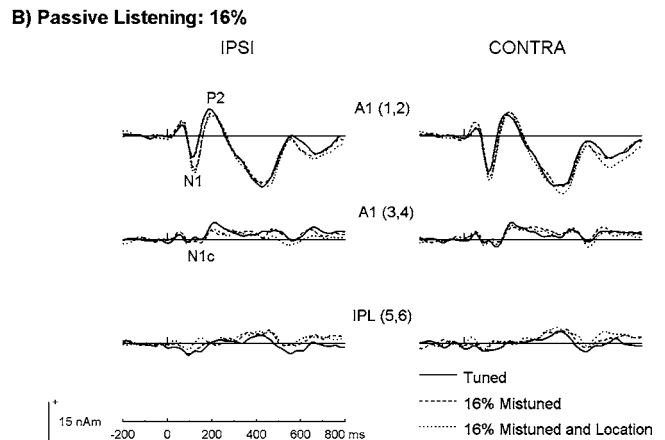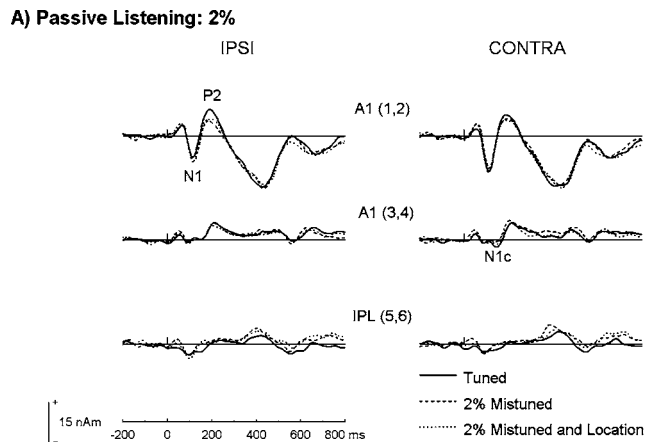
FIG. 4. Passive listening. Source waveforms of the group mean ERPs during passive listening. (A) Source waveforms for the tuned and the 2% mistuned harmonic presented either at the same or at a different location than the remaining harmonics. (B) Source waveforms for the tuned and the 16% mistuned harmonic presented either at the same or at a different location than the remaining harmonics.

As in active listening, the N1 wave peaked earlier and its amplitude was larger in the hemisphere contra- (117 ms; -14.94 nA m) than ipsi-lateral (122 ms; -11.72 nA m) to the harmonic series [$F(1,11)=31.37$ and 11.70, respectively, $p<0.006$ in both cases]. The N1 peak latency and amplitude increased with mistuning, $F(2,22)=5.44$ and 12.07, $p<0.05$ in both cases. The main effect of location was not significant nor did location interact with any other factors. The P2 peak latency and amplitude measured at the tangential source was quantified between 140 and 280 ms after sound onset. The main effect of mistuning, location or hemisphere on P2 latency and amplitude was not significant nor was the interaction between mistuning and location.

Figure 5 shows ORN elicited by mistuned stimuli presented at either the same or different location than the remaining harmonics. An ANOVA on the mean amplitude for the 100–200 ms interval from the tangential sources yielded a main effect of mistuning, $F(2,22)=18.47$, $p<0.001$, all pairwise comparisons at $p<0.05$. The main effect of location was not significant nor was the interaction between location and any other factor.

C. Discussion

Participants were more likely to report hearing two distinct sounds when the stimulus contained a mistuned har-

A) Difference Wave: 2%

IPSI   CONTRA

A1 (1,2)

ORN

A1 (3,4)

IPL (5,6)

—— mistuning only

········ mistuning and location

15 nAm   -200   0   200   400   600   800 ms

B) Passive Listening: 16%

IPSI   CONTRA

A1 (1,2)

ORN

A1 (3,4)

IPL (5,6)

—— mistuning only

----- mistuning and location

15 nAm   -200   0   200   400   600   800 ms
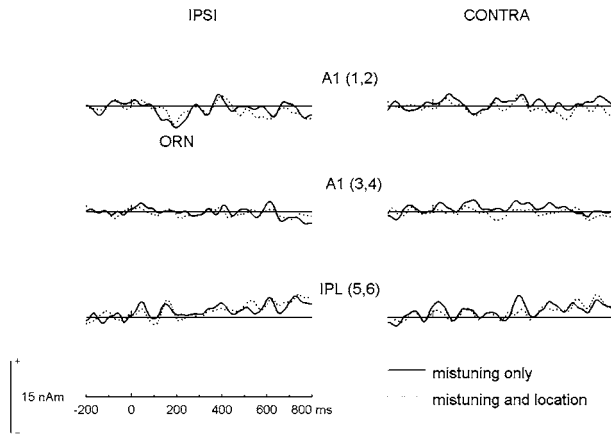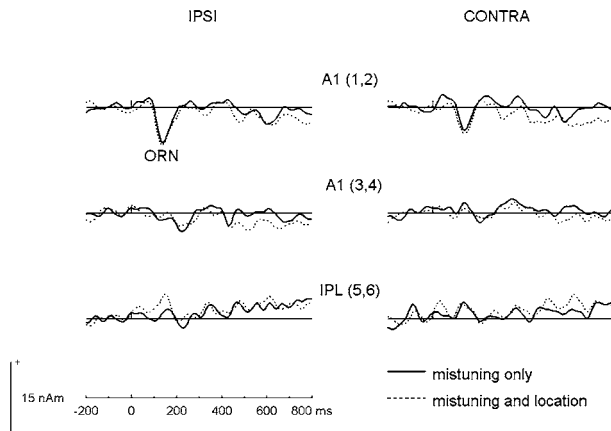
FIG. 5. Passive listening. Object-related negativity (ORN) isolated by subtracting source waveforms for the tuned harmonic presented at the same location from those obtained for mistuned stimuli presented either at the same or different location as the harmonic series. (A) ORN for the 2% mistuned harmonic. (B) ORN for the 16% mistuned harmonic. Note that ORN is little affected by attention whereas the P400 is absent.

monic. The greater the degree of inharmonicity within the stimulus, the more quickly and reliably participants reported the presence of two sounds. This is consistent with previous research (e.g., Alain *et al.*, 2001a, 2002; Hartmann *et al.*, 1990; Moore *et al.*, 1986) and provides further confirmation that frequency periodicity provides an important cue in distinguishing co-occurring auditory objects.

Presenting the third harmonic at a different location than the remaining harmonics also modulated participants' responses, increasing the likelihood of reporting that two sounds were present when the stimulus contained a tuned or a mistuned harmonic at near-threshold levels, and decreasing the response time needed to indicate this. The fact that participants reported hearing two concurrent objects significantly more often when the third harmonic was presented from a separate loudspeaker than when it came from the same speaker even when the third harmonic was in tune with the rest of the partials, indicates that sounds can be segregated based on location alone. When the third harmonic was mistuned by 16%, however, the additional location cue did not further influence listeners' responses. This is not surprising given that the likelihood of reporting hearing two con-

current sounds was already at 96% when the third harmonic was presented at the same location as the remaining harmonics.

The effects of harmonicity and location on perception were associated with distinct patterns of neural activity. As in previous studies, the 16% mistuned harmonic "popped out" of the complex sound and was accompanied by an ORN and P400 (Alain *et al.*, 2001a, 2002; Alain and Izenberg, 2003). The latter was present only when listeners were asked to make a response whereas the ORN was observed in both active and passive listening conditions, consistent with the proposal that concurrent sound segregation based on inharmonicity occurs automatically (Alain and Izenberg, 2003; Dyson and Alain, 2004). Although the two listening conditions differed in terms of stimulation rate, it is unlikely that this factor played an important role since the ORN recorded during passive listening is not affected by stimulus onset asynchrony (Alain *et al.*, 2002).

Like participants' phenomenological experience, the source waveforms for the 16% mistuned stimuli were little affected by presenting the third harmonic at a different location than the remaining harmonics. A different pattern of results emerged for the tuned and near threshold-mistuned stimuli. For the tuned stimuli, the P2 wave peaked earlier when the third harmonic was presented at a different location, providing further evidence that sound can be segregated based on location alone. In the present study, the location-based segregation of the tuned harmonic appears to be indexed by a different neural event than mistuning. It is possible that the simultaneous activity of harmonically related, yet spatially distinct physical sources, may have contributed to a "wider" acoustic image with a perceived location somewhere between the two loudspeakers. The sum of acoustic energy from both speakers could also account in part for the early P2 wave. Although the effect of location on the P2 latency was reversed with increasing mistuning, this could be due to the superimposition of the ORN on the N1 and P2 wave.

Paralleling the behavioral data for the 2% mistuned stimuli, there was a late negative potential that resembled the ORN, which was greater in amplitude when the mistuned harmonic was presented at a different location than the remaining harmonics. However, this effect of location appeared relatively late in comparison to the ORN elicited by 16% mistuning and was present only when the stimuli were attended. Thus, it is possible that the P2 amplitude decrement was caused by the superimposition of the P2 wave by attention-related negativity waves or a late ORN. This late negativity associated with the conjunction of mistuning and location could be related to the N2b wave, which reflects stimulus classification and decision processes that control behavioral responses in discrimination tasks (Ritter *et al.*, 1979, 1982). For instance, participants may have adopted a strategy in which they equally divided their attention between the two speakers thereby facilitating the detection of the third harmonic when it was presented at a different location than the remaining harmonics. This would seem a reasonable strategy given that the probability of sounds emanating from the left or right speaker was equiprobable.

When all of the harmonics were in tune and presented from one location, participants indicated the presence of two sounds on 20% of the trials. This may reflect some response bias in responding that two objects were present. It is also possible that participants heard the third harmonic as a separate object because it was the only harmonic that was ever mistuned. For example, participants may have realized that the changes in the frequency component were always in the same frequency region and therefore listened more carefully for sounds at that particular frequency. It has been shown that individuals are able to identify a single harmonic in a complex sound if they have previously listened to that harmonic presented alone (Plomp, 1964; Plomp and Mimpen, 1968). This directed attention might have been facilitated by those trials in which the third harmonic was presented from a different location because it may have increased the participants' focus on this frequency region and further primed them to hear the frequency of the third harmonic.

In this first experiment, participants were never asked to indicate where they perceived the sound with a pure tone quality as coming from, making it difficult to determine whether participants could localize the third harmonic when they reported hearing it as a separate object. This is particularly important for trials in which the third harmonic was tuned or slightly mistuned since they were likely to benefit the most from the addition of a location cue. In Experiment 2, listeners were asked to indicate the location of the sound with the pure tone quality. This experimental design allowed us to assess whether participants could segregate and localize the third harmonic. In addition, the f_0 varied randomly to decrease the likelihood that participants would focus on one frequency region to complete the task and a larger spatial separation was used to promote segregation based on location alone.

III. EXPERIMENT 2

A. Methods

1. Participants

Ten right-handed adults (between 20 and 38 years of age, mean age=27.9 ± 5.65 years; 4 males) participated in the experiment. All participants had pure-tone thresholds less than or equal to 30 dB HL at frequencies in the range 250–8000 Hz in both ears. Each participant gave informed consent in accordance with the University of Toronto and Baycrest Centre for Geriatric Care guidelines. Participants were volunteers from the University of Toronto and laboratory staff, and were paid for their participation.

2. Stimuli and task

Stimuli were complex sounds comprised of 10 harmonics of equal intensity with an f_0 of 175, 200, or 225 Hz. Three different f_0's were chosen to decrease the likelihood that participants would focus on one frequency region to complete the task. For the mistuned stimuli, the third harmonic was shifted upwards or downwards by 4% of its original value. This magnitude of mistuning was chosen because it is above threshold (Alain *et al.*, 2001b) and therefore should be detected by most listeners, but not so overwhelm-

ing a frequency cue as the 16% mistuning used in the first experiment. We also used only one level of mistuning to increase the number of observations per stimulus and to reduce response bias since 50% of stimuli did not include a mistuned harmonic as opposed to only 33% in the first experiment. As in the first experiment, the third harmonic was played from either the same loudspeaker or a different loudspeaker than the other nine harmonics.

On each trial, participants first indicated whether they heard one sound or two sounds following the presentation of each stimulus. If they indicated that two sounds were present, then participants specified the location (left or right of center) of the sound with the pure tone quality (i.e., the mistuned harmonic). Participants did not receive any feedback on their performance. Within a block of trials, twelve different and equiprobable stimuli were played at random (three levels of f_0; two possible levels of mistuning −0% and ±4%; third harmonic presented at either the same or different loudspeaker as the remaining harmonics). A chin rest was used throughout the experiment to ensure that participants' heads remained in a constant position relative to the loudspeakers (at an angle of 60° to each speaker). The angle was increased from the first experiment to emphasize the location cue. The distance from each loudspeaker to the estimated center of participants' heads was 1.32 m while the distance between loudspeakers was 2.23 m. A total of 960 trials were presented to each participant (80 per stimulus type).

3. Electrophysiological recording and data analysis

Recording and analysis of ERPs were similar to that of Experiment 1. ERPs were averaged across the three different f_0 to increase the signal-to-noise ratio. As in Experiment 1, two regional sources were seeded in left and right Heschl's gyrus to account for the neural activity between 0 and 260 ms after sound onset. Maintaining the symmetrical and orthogonal constraints, the orientations of the tangential sources were then aligned with the maximum direction of activity and the three sources were converted to single dipoles. Two dipoles were seeded in the left and right IPL to account for the sustained potential between 300 and 800 ms post-stimulus. The anterior-posterior component of the regional source in primary auditory cortex was removed since it accounted for less than 1% of the total variance and the orientation of the tangential, radial, and parietal sources was allowed to change. The resulting model was held fixed and used as a spatial filter to derive the source waveforms (three in each hemisphere) in each participant and for all stimulus types, separately. The residual variance was 4.84%.

B. Results

1. Behavioral data

Participants were more likely to report hearing two concurrent sounds when the third harmonic was mistuned, $F(1,9)=267.78$, $p<0.001$, and when the third and remaining harmonics were presented from different loudspeakers, $F(1,9)=30.05$, $p<0.001$ (Table II). There was also a significant interaction between harmonicity and location, $F(1,9) =16.83$, $p<0.005$. When the third harmonic was in tune,

TABLE II. Group mean responses (and standard error of the mean) as a function of harmonicity and location.

Percentage mistuning	Location	Percentage of stimuli heard as two sounds	Confidence intervals lower–upper bound	Response time (ms)	Confidence intervals lower–upper bound
0%	Same	18.81 (4.35)		546.04 (42.86)	
	Different	32.12 (5.29)		522.17 (32.23)	
	Difference	13.32 (2.50)	7.67–18.96	−23.87 (16.64)	−61.52–13.78
4%	Same	87.88 (2.56)		622.82 (49.36)	
	Different	91.29 (1.86)		620.33 (53.75)	
	Difference	3.42 (1.16)	0.80–6.04	−2.50 (20.66)	−49.24–44.25

participants reported hearing two sounds more often when it was presented at a different location than the remaining harmonics, $t(9)=7.58$, $p<0.05$, whereas listeners' perception was little affected by differences in spatial location when the third harmonic was mistuned by 4%, $t(9)=0.65$, $p=0.44$. It is possible that participants were responding at ceiling before the location cue was introduced. There was also no significant difference in response time among the various stimulus types.

The overall proportion of trials in which participants indicated hearing two concurrent auditory events was 58%, and participants could accurately identify the location of the third harmonic on 34% of those trials. Performance was comparable for tuned and mistuned stimuli. A close examination of the individual data revealed a bimodal distribution, with five participants almost always reporting hearing the third harmonic at the same location as the remaining harmonics (mean=99%). Thus, although participants reported hearing two auditory events on many occasions, they were unable to accurately indicate the location of the target signal (i.e., the third harmonic) on these trials. Four of the remaining five participants were able to accurately localize the third harmonic more than 50% of the time (mean=70%) in which they reported hearing two concurrent sounds. The last participant was able to accurately localize the third harmonic on only 35% of trials previously identified as containing two sounds.

2. Electrophysiological data

Figure 6 shows group mean source waveforms when the mistuned harmonic was presented from the same or at a different location than the remaining harmonics. As in the first experiment, the N1 wave peaked earlier and its amplitude
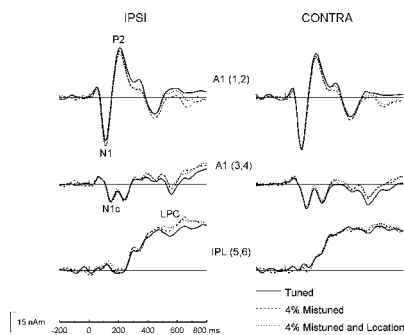
was larger in the hemisphere contra- (111 ms; −40.56 nA m) than ipsi-lateral (114 ms; −37.29 nA m) to the harmonic series $[F(1,11)=11.84$ and 14.07, respectively, $p<0.01$ in both cases]. The effects of mistuning or location on N1 peak latency and amplitude were not significant. The N1c peaked at about 144 ms after sound onset and tended to be larger in the hemisphere contralateral to the harmonic series, $F(1,9)=4.70$, $p=0.058$. The effects of mistuning and location on the N1c peak latency and amplitude were not significant.

For the P2 wave, the main effect of mistuning or location was not significant. However, there was a main effect of hemisphere $[F(1,9)=8.38$, $p<0.02]$, and a significant interaction between hemisphere and location on the P2 peak amplitude, $F(1,11)=6.12$, $p<0.05$. There was a greater P2 amplitude decrement in the hemisphere ipsi- than contra-lateral to the harmonic series when the third harmonic was presented at a different location than the remaining harmonics. There was also a main effect of mistuning on the tangential component during the 130–230 ms interval, $F(1,9)=5.61$, $p<0.05$. This is best illustrated in the difference wave between source waveforms elicited by tuned and mistuned stimuli presented either at the same or different location than the remaining harmonics (Fig. 7). The ORN was little affected by the conjunction of location and harmonicity. However, there was a significant interaction between mistuning and location for the P400 (400–500 ms), $F(1,9)=10.56$, $p<0.01$. For the tuned stimuli, the P400 showed larger amplitude when the third harmonic was presented at a different location whereas for the mistuned stimuli, adding a location cue generated smaller amplitude relative to the mistuned stimuli presented at the same location as the remaining harmonics.

Half of our participants could localize the third harmonic when they reported hearing two concurrent auditory events. The other half was unable to localize the third har-

FIG. 6. Source waveforms for the group mean ERPs recorded in Experiment 2.
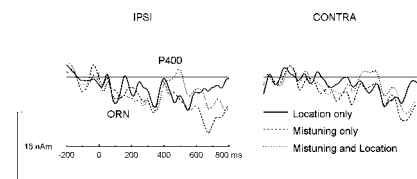
FIG. 7. Object-related negativity isolated by subtracting source waveforms for the tuned harmonic presented at the same location from those obtained for tuned and mistuned stimuli presented either at the same or different location as the harmonic series. Note that only source waveforms from the tangential source are displayed.

ORN
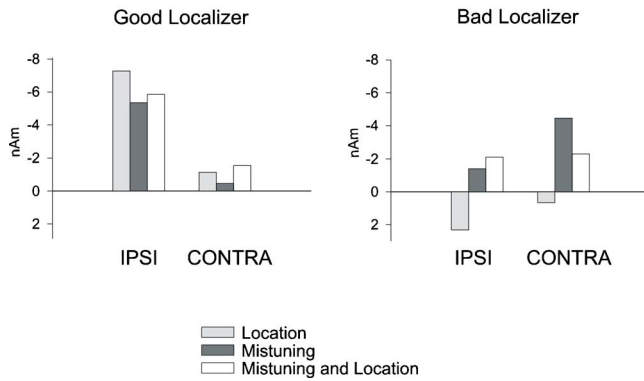
FIG. 8. Bar graph showing the ORN mean amplitude between 120 and 230 ms after sound onset for subsets of participants who could (good localizer) or could not (bad localizer) accurately localize the third harmonic when heard as a separate sound.

monic regardless of mistuning and/or location. We further examined the effect of mistuning and location on those listeners ($N=5$) that could localize the third harmonic when reported as standing out of the complex sound (Fig. 8). In this subset of participants, we found a significant interaction among mistuning, location, and hemisphere, $F(1,4)=16.43$, $p<0.02$. There was a significant ORN when the tuned harmonic was presented from a different location than the remaining harmonics and the amplitude was larger in the hemisphere contra-lateral to the location of the single harmonic.[3] This ORN for location only was comparable to the one recorded to a mistuned harmonic presented at either the same or different location than the remaining harmonics. In the subset of participants that could not localize the third harmonic, there was also a significant interaction among mistuning, location, and hemisphere, $F(1,4)=8.36$, $p<0.05$. However, the ORN was significant only when mistuning was present and was larger in the hemisphere contra-lateral to the harmonic series. Lastly, an ANOVA with group as an additional between-subject factor yielded a significant group $\times$ mistuning $\times$ location $\times$ hemisphere interaction, $F(1,8)=21.32$, $p<0.002$, indicating that location effects on brain activity were different in individuals that could accurately localize the third harmonic presented at a different location than the remaining harmonics.

C. Discussion

The likelihood of hearing two concurrent sounds was greater for mistuned than tuned stimuli and for tuned stimuli when the third harmonic was presented at a different location than the remaining harmonics. The participants' likelihood of hearing the mistuned harmonic as a separate object was little affected by differences in spatial location. Although 4% mistuning may be less salient than 16%, it appears that it was sufficient in itself to guide listeners' responses.

As in the first experiment, the perception of concurrent auditory objects was paralleled by an ORN. The ORN was present for both location- and harmonicity-based segrega-

tion. Coupled with the behavioral data, the ORN for segregation based on location alone is consistent with recent studies showing an ORN to ITD leading to the perceptual segregation of a dichotically embedded pitch (Hautus and Johnson, 2005; Johnson *et al.*, 2003). The fact that a significant ORN was recorded in cases where sounds were segregated based on location alone indicates that this component is more related to the perception of concurrent events than the level of mistuning per se.

When the third harmonic was mistuned by 4%, participants reported hearing two concurrent objects on almost 90% of the trials. This is higher than in previous studies (Alain *et al.*, 2001a; Moore *et al.*, 1986) and may indicate a tendency for participants to report hearing two sounds when they perceive mistuning, even if they did not clearly hear the third harmonic as a separate entity. The fact that half of the participants could not accurately locate the position of the spatially or harmonically segregated sound when they reported hearing two auditory objects is consistent with a bias toward responding hearing two simultaneous auditory events.

As mentioned earlier, some participants who reported hearing two concurrent auditory objects could not localize the source of the third harmonic on those trials. There are several possible explanations for such findings. One possibility is that sound object segregation based on spatial cues and sound localization do not rely on the same mechanisms. For example, Thiran and Clarke (2003) report a case study in which a patient with a right temporo-parieto-frontal ischemic lesion was able to segregate sounds on the basis of a location cue but was unable to accurately localize auditory events. Another possibility is that those individuals that were not able to localize the third harmonic did not hear it as a separate object but rather reported hearing two sounds on those trials in which they perceived the mistuning. The level of mistuning could have been ambiguous for less experienced listeners. It is interesting to note that those five individuals that had difficulty localizing the third harmonic also had less experience with the sort of task used in the current study than those that could accurately localize the third harmonic when it was heard as a separate object.

As in the first experiment, participants often reported hearing two concurrent sounds when the third harmonic was in tune and presented at the same location as the remaining harmonics. The level of performance for the tuned stimuli cannot easily be accounted for by attention to a particular frequency because the f_0 varied randomly from trial to trial. However, it remains possible that schema-guided processes may have played a role in the perception of the tuned harmonic as a separate sound. For instance, listeners may have learned to quickly shift their attention to a particular frequency region according to the f_0 of the incoming stimulus.

IV. GENERAL DISCUSSION

The aim of the present study was to evaluate the relative contribution of harmonicity and location cues during auditory scene analysis. More specifically, we were interested in assessing the role of location cues in concurrent sound seg-

regation. In both experiments, we found comparable effects of spatial separation on listeners' likelihood of reporting two simultaneous auditory objects.

The perception of concurrent auditory objects was paralleled by an ORN that superimposed the N1 and P2 waves. A clear ORN was recorded when the sounds were segregated based on harmonicity. In Experiment 1, the location cues modulated ERPs approximately 250 ms after sound onset whereas in Experiment 2 location cues modulated ERPs as early as 150 ms after sound onset. Differences between the two experiments may be related to the greater spatial separation in Experiment 2.

The harmonic sieve has often been proposed to account for the perception of the mistuned harmonic as a separate event. That is, a template would be formed on the basis of the fundamental frequency and then the harmonics would "pass" through some sort of filter (i.e., sieve). When a harmonic is sufficiently mistuned, it can no longer "pass" through the sieve and therefore would generate a mismatch signal indicating that there is likely more than one sound source active. Such detection of inharmonicity based on a harmonic sieve could occur as early as 30 ms after sound onset (Dyson and Alain, 2004). Although a harmonic sieve or template could account for the effects of mistuning, such a model has difficulty explaining the effects of location as well as the interactions between inharmonicity and location cues on electrophysiological and behavioral data. One alternative possibility to the harmonic sieve would be a model in which the template extrapolated from the incoming stimulus includes both harmonicity and location information.

A secondary question was whether a difference in spatial location would impair listeners' ability to perceptually fuse the stimuli containing exclusively harmonic components. When the harmonic stimuli were presented from two locations, the harmonicity and location cues were working in direct opposition to one another. Previous studies have found that a discrepancy in location may be ignored when other cues favor the fusion of components (Deutsch, 1975). The likelihood of indicating the perception of two sounds significantly increased when the location cue was introduced, suggesting that a contradictory location cue can influence listeners' perception of a harmonic stimulus. A previous study investigating perceptual grouping of sequential sounds found that increasing the spatial separation between a target tone and background tones by 32° reduced the perceptual capture of the target tone (O'Connor and Sutter, 2000). It is possible that the same thing occurred here—the perceptual capture of the third harmonic was impeded by its spatial separation from the remaining harmonics. The loudspeakers in the current studies were separated by angles of 90° and 120° relative to the center of participants' heads (each loudspeaker was 45° or 60° left/right of center), well past the 32° used in the study of sequential grouping. Evidence from previous research suggests that our auditory space may be divided into two broadly tuned perceptual channels, occupying the left and right auditory hemifields, respectively, each extending 30° across the midline (Boehnke and Phillips, 1999; Phillips,

1999). From this perspective, the spatial separation used in both experiments falls well outside these perceptual channels.

In the present study, both ITD and ILD may have contributed to segregation based on spatial location. Given the frequency of the third harmonic ($\sim$600 Hz), it is likely that segregation was based primarily on ITD, which plays an important role in localizing low frequency sounds. This interpretation differs from that of previous work showing that ITD provides little help in parsing concurrent vowels whereas ILD does help the listener somewhat (Darwin and Hukin, 1999). This discrepancy may be related to the fact that our stimuli were presented in free field as opposed to over headphones. The effects of spatial separation between the target harmonic and the remaining harmonics on both behavioral and electrophysiological data were relatively small, albeit significant. Previous works have shown that reverberation reduces the advantage of spatial separation between sound sources (Culling *et al.*, 1994; Plomp, 1976). Hence, in the present study room reverberation could have contributed to the small effects of spatial separation on ERPs.

In summary, the behavioral and electrophysiological results indicated that sounds are segregated based on harmonicity and/or location. In particular, when harmonicity provides a somewhat ambiguous cue, a difference in the spatial location of one sound can aid the listener in segregating it from other sounds. The effects of mistuning on brain activity were present during both active and passive listening, consistent with the proposal that processing harmonicity occurs automatically. In comparison, the effects of location on brain activity were observed only during active listening. This suggests that sound segregation based on sound location may rely on top-down attentional processes to a greater extent than segregation based on frequency periodicity. Additional studies are needed to confirm this hypothesis as well as to further assess where in the brain these effects may be mediated.

[1]We recorded more trials in the passive listening condition because the ISI was shorter and thus allowed us to record more observations within the same amount of recording time.
[2]Note that the midline electrodes remained unchanged during the transposition.
[3]An ANOVA based on a subset of four subjects (excluding the one that could only occasionally localize the third harmonic) generated the same pattern, although the three-way interaction failed to reach significance, $F(1,3)=9.45$, $p=0.054$.

Alain, C., and Arnott, S. R. (**2000**). "Selectively attending to auditory objects," Front. Biosci. **5**, D202–212.
Alain, C., Arnott, S. R., and Picton, T. W. (**2001**). "Bottom-up and top-down influences on auditory scene analysis: Evidence from event-related brain potentials," J. Exp. Psychol. Hum. Percept. Perform. **27**, 1072–1089.
Alain, C., and Izenberg, A. (**2003**). "Effects of attentional load on auditory scene analysis," J. Cogn Neurosci. **15**, 1063–1073.
Alain, C., McDonald, K. L., Ostroff, J. M., and Schneider, B. (**2001b**). "Age-related changes in detecting a mistuned harmonic," J. Acoust. Soc. Am. **109**, 2211–2216.
Alain, C., Schuler, B. M., and McDonald, K. L. (**2002**). "Neural activity associated with distinguishing concurrent auditory objects," J. Acoust. Soc. Am. **111**, 990–995.
Alain, C., Theunissen, E. L., Chevalier, H., Batty, M., and Taylor, M. J.

(**2003**). "Developmental changes in distinguishing concurrent auditory objects," Brain Res. Cognit. Brain Res. **16**, 210–218.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (**2004**). "Assessing the auditory dual-pathway model in humans," Neuroimage **22**, 401–408.

Boehnke, S. E., and Phillips, D. P. (**1999**). "Azimuthal tuning of human perceptual channels for sound location," J. Acoust. Soc. Am. **106**, 1948–1955.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sounds* (MIT, London).

Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Bronkhorst, A. W., and Plomp, R. (**1992**). "Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing," J. Acoust. Soc. Am. **92**, 3132–3139.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Carlyon, R. P. (**2004**). "How the brain separates sounds," Trends in Cognitive Science **8**, 465–471.

Chalikia, M. H., and Bregman, A. S. (**1989**). "The perceptual segregation of simultaneous auditory signals: Pulse train segregation and vowel segregation," Percept. Psychophys. **46**, 487–496.

Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1994**). "Effects of simulated reverberation on the use of binaural cues and fundamental-frequency differences for separating concurrent vowels," Speech Commun. **14**, 71–95.

Darwin, C. J., and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol. Hum. Percept. Perform. **25**, 617–629.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am. **57**, 1156–1160.

Drennan, W. R., Gatehouse, S., and Lever, C. (**2003**). "Perceptual segregation of competing speech sounds: The role of spatial location," J. Acoust. Soc. Am. **114**, 2178–2189.

Dyson, B., and Alain, C. (**2004**). "Representation of concurrent acoustic objects in primary auditory cortex," J. Acoust. Soc. Am. **115**, 280–288.

Gockel, H., and Carlyon, R. P. (**1998**). "Effects of ear of entry and perceived location of synchronous and asynchronous components on mistuning detection," J. Acoust. Soc. Am. **104**, 3534–3545.

Hartmann, W. M., McAdams, S., and Smith, B. K. (**1990**). "Hearing a mistuned harmonic in an otherwise periodic complex tone," J. Acoust. Soc. Am. **88**, 1712–1724.

Hautus, M. J., and Johnson, B. W. (**2005**). "Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch," J. Acoust. Soc. Am. **117**, 275–280.

Johnson, B. W., Hautus, M., and Clapp, W. C. (**2003**). "Neural activity associated with binaural processes for the perceptual segregation of pitch," Clin. Neurophysiol. **114**, 2245–2250.

Kubovy, M., and Van Valkenburg, D. (**2001**). "Auditory and visual objects," Cognition **80**, 97–126.

Lee, J., and Green, D. M. (**1994**). "Detection of a mistuned component in a harmonic complex," J. Acoust. Soc. Am. **96**, 716–725.

Lin, J. Y., and Hartmann, W. M. (**1998**). "The pitch of a mistuned harmonic: Evidence for a template model," J. Acoust. Soc. Am. **103**, 2608–2617.

Moore, B. C., Glasberg, B. R., and Peters, R. W. (**1986**). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," J. Acoust. Soc. Am. **80**, 479–483.

O'Connor, K. N., and Sutter, M. L. (**2000**). "Global spectral and location effects in auditory perceptual grouping," J. Cogn Neurosci. **12**, 342–354.

Phillips, D. P. (**1999**). "Auditory gap detection, perceptual channels, and temporal resolution in speech perception," J. Am. Acad. Audiol **10**, 343–354.

Picton, T. W., van Roon, P., Armilio, M. L., Berg, P., Ille, N., and Scherg, M. (**2000**). "The correction of ocular artifacts: A topographic perspective," Clin. Neurophysiol. **111**, 53–65.

Plomp, R. (**1964**). "The ear as frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Plomp, R. (**1976**). "Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of a single competing sound source (speech or noise)," Acustica **34**, 200–211.

Plomp, R., and Mimpen, A. M. (**1968**). "The ear as a frequency analyzer. II," J. Acoust. Soc. Am. **43**, 764–767.

Ritter, W., Simson, R., Vaughan, H. G., Jr., and Friedman, D. (**1979**). "A brain event related to the making of a sensory discrimination," Science **203**, 1358–1361.

Ritter, W., Simson, R., Vaughan, H. G., Jr., and Macht, M. (**1982**). "Manipulation of event-related potential manifestations of information processing stages," Science **218**, 909–911.

Stellmack, M. A., and Dye, R. H., Jr. (**1993**). "The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity," J. Acoust. Soc. Am. **93**, 2933–2947.

Thiran, A. B., and Clarke, S. (**2003**). "Preserved use of spatial cues for sound segregation in a case of spatial deafness," Neuropsychologia **41**, 1254–1261.

Informational masking release in children and adults

Joseph W. Hall III,[a] Emily Buss, and John H. Grose
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599

This study assessed informational masking and utilization of cues to reduce that masking in children aged 4–9 years and in adults. The signal was a train of eight consecutive tone bursts, each at 1 kHz and 60 ms in duration. Maskers were comprised of a pair of synchronous tone-burst trains, with randomly chosen frequencies spanning 200–5000 Hz, with a protected region 851–1175 Hz. In the reference condition, maskers were eight bursts in duration, with a fixed frequency within intervals. Experiment 1 tested two monotic masking release conditions: within-interval randomization of masker burst frequency and the introduction of leading masker bursts. Experiment 2 examined masking release in which the signal was presented to one ear and masking components were presented to both ears (masker components in the contralateral ear were 10 dB higher than those in the ipsilateral ear). Both adults and children demonstrated a significant informational masking effect, with children showing a larger effect on average. Both groups also showed significant release from masking in the two monotic conditions, although children received somewhat less benefit from the masking release cues. The binaural condition supported a moderate release from informational masking in adults, but resulted in *increased* informational masking in children. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992675]

I. INTRODUCTION

Previous investigations have shown that stimulus components well separated in frequency from a sinusoidal target signal can cause large masking effects when the spectral composition of the remote masker varies randomly on a trial-by-trial basis (e.g., Neff and Callaghan, 1987; Neff and Green, 1987; e.g., Watson, 1987; Neff and Callaghan, 1988; Lutfi, 1990; Leek *et al.*, 1991; Kidd *et al.*, 1994; 2002; Richards *et al.*, 2002; Kidd *et al.*, 2003). Such effects are usually referred to as informational rather than energetic masking because they do not appear to depend upon the excitation of peripheral sensori-neural elements associated with the signal frequency. Informational masking is of particular interest because of its potential to provide insights about auditory processing of complex and variable sounds that occur in natural settings.

There is a small body of research suggesting that informational masking is particularly large in young listeners. For example, Werner and Bargones (1991) showed that infants experienced appreciable masking arising from spectral components two octaves above the signal frequency. This result was interpreted in terms of nonsensory masking in the infants. Informational masking is also relatively large in young children. When tested using maskers having trial-by-trial frequency variability, preschool and school-aged children often showed more informational masking than adults (Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003).

The present study is concerned primarily with release from informational masking. Studies by Neff (1995) and by Kidd and colleagues (Kidd *et al.*, 1994; Arbogast *et al.*,

2002) have shown that informational masking in adults can be reduced substantially by the introduction of stimulus cues thought to promote the perceptual segregation of the signal from the masking stimulus. For example, informational masking can be reduced by providing cues that influence perceptual segregation via auditory streaming, spectro-temporal coherence, and binaural spatial separation (Kidd *et al.*, 1994; Arbogast *et al.*, 2002). In general, informational masking can be reduced by the introduction of stimulus features that make the signal sound different or separate from the masker (Durlach *et al.*, 2003a; 2003b). The study of release from informational masking is important because it may provide clues about auditory processes that enable listeners to separate attended signals from complex, dynamic acoustical maskers.

Whereas psychoacoustical experiments indicate that children are especially prone to informational masking, the ramifications that this may have for hearing are not clear. One reason for this is that little is currently known about the extent to which children experience release from informational masking with the introduction of segregation cues. If such release is robust in children, then informational masking effects in real environments, where multiple sound source segregation cues often exist (e.g., Bregman, 1990; e.g., Yost, 1991), may not be of particular significance. There is some evidence that release from informational masking in children may not be particularly robust. A recent study by Wightman *et al.* (2003) indicated that children in the age range of approximately 4–13 years showed less release from masking than adults when the masker was presented to the nonsignal ear; in this case, the segregation cue was presentation of the signal and masker to opposite earphones, a manipulation that could result in masking release via differential

lateralization of the signal and masker. Whereas this stimulus manipulation resulted in essentially total release from masking in adult listeners, it resulted in essentially no release for the younger children and only partial release for many of the older children.

The purpose of the present investigation was to study the development of release from informational masking further. Experiment 1 explored developmental effects for release from informational masking using monaural sound segregation cues related to onset asynchrony and spectro-temporal coherence. Previous studies on adult listeners (Kidd *et al.*, 1994) have shown that such manipulations can result in significant release from informational masking. At the outset, it was hypothesized that children would show relatively poor benefit from monotic cues to informational masking. This hypothesis is consistent with the findings of reduced efficacy of binaural cues to masking release reported by Wightman *et al.* (2003).

II. EXPERIMENT 1

A. Methods

1. Stimuli

The conditions used here were based on the multiburst paradigm of Kidd *et al.* (1994). In this paradigm, the signal was a train of eight bursts of a 1000-Hz pure tone, each burst 60 ms in duration including 10-ms $\cos^2$ ramps. These bursts were concatenated, with no interburst gaps, resulting in a total signal duration of 480 ms. The condition associated with the largest informational masking in this paradigm is referred to as a *multiburst-same* (MBS) stimulus. Here, the masker was comprised of two trains of eight, 60-ms tone bursts that were synchronous with the signal bursts. For each observation interval, the frequency for the bursts in each of the trains was randomly selected from a range 200–5000 Hz. Within this range, there was a "protected" frequency region, extending from 851–1175 Hz, from which random masker picks were excluded. The purpose of the protected region was to limit the influence of energetic masking. Once selected, the frequency of each masker train remained constant over the observation interval. An example of a signal-present stimulus in the MBS condition is illustrated in Fig. 1 [panel (A)], with the tone bursts represented by dashed lines and gray regions indicating the range of possible masker frequencies.

There were also two masking conditions incorporating cues for perceptual segregation of the signal and masker components. After Kidd *et al.* (1994), one of these employed a perceptual segregation cue related to spectro-temporal coherence, and is referred to as the *multiburst-different* (MBD) condition. Here, the masker was again composed of two trains of stimuli, each comprised of eight 60-ms bursts. However, in this MBD stimulus, the frequency within each of the two trains varied randomly from burst to burst (over the ranges 200–851 and 1175–5000 Hz). Thus, as illustrated in Fig. 1 [panel (B)], the frequency components within each of the two masker trains varied randomly over the course of the eight bursts, while the frequency of the signal remained fixed over the course of the eight bursts. Kidd *et al.* (1994) specu-
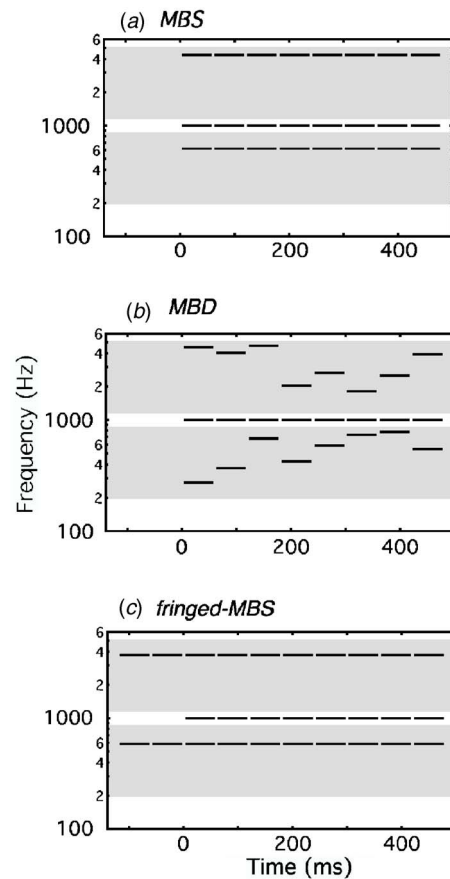
FIG. 1. Stimulus schematic for experiment 1.

lated that masking release in the MBD condition resulted from a perceptual process in which the spectro-temporal constancy of the signal allowed its segregation from the set of randomly varying masker components.

The second masking release condition employed a temporal onset asynchrony cue. Here, the signal and masker were the same as for the MBS masker except that the masker began two bursts prior to the signal onset. In this case, the masker was comprised of ten 60-ms bursts, with the frequency within each train remaining constant across the ten bursts. An example of a signal-present stimulus in the *fringed*-MBS condition is shown in Fig. 1 [panel (C)]. The assumption was that, if spectral components possess temporal asynchrony, they should be easier to separate perceptually (Bregman and Pinker, 1978; Dannenbring and Bregman, 1978; Darwin, 1984). Note that, when the signal bursts were present, they were synchronous with the masker bursts, and that the asynchrony in this paradigm was with respect to the masker burst train beginning earlier than the stimulus burst train. This stimulus bears some resemblance to a masking release condition used by Kidd *et al.* (1994), where signal bursts were presented synchronously with alternate masker bursts. It also shares similarities with the "masker first" technique used by Richards *et al.* (2004), where a pretrial cue of masker alone resulted in release from informational masking.

Sound generation was performed digitally, in the time domain, using an array processor (TDT, APOS), with the signal and masker played out via separate channels of a

digital-to-analog converters (TDT, DD1) at a rate of 20 kHz. The level of each channel was adjusted via a programmable attenuator (TDT, PA4). Each channel was antialias filtered at 7 kHz (Kemo, VBF8), summed (TDT, SM3), and played to the left channel of a Sony MDR V6 earphone.

2. Procedures

Stimuli were presented in blocks of three-alternative, forced-choice trials, and responses were tracked adaptively using a 3-up, 1-down paradigm (Levitt, 1971). An initial step size of 8 dB was reduced to 4 dB after the first reversal, and to 2 dB after the second reversal. The adaptive track included six total reversals, with the track result estimated as the average level at the final four reversals. An adaptive track typically consisted of 25–30 trials. Three such track estimates were obtained, and a fourth was performed if the first three estimates varied by more than 3 dB. Data reported below were the average of all track estimates obtained for a particular condition. Thresholds were obtained first for the eight-burst signal presented in quiet. This threshold in quiet was then used to set the signal level for the remaining conditions of the experiment. In the masking conditions, the signal level was maintained at 30 dB SL (with regard to the threshold in quiet), and the level of the masker was changed adaptively to estimate the masked threshold. The maximum masker level allowed by the software was 83 dB SPL, but very few tracks reached this maximum, so this constraint had minimal impact on the results presented here. Each listener first completed the MBS condition and then completed one of the two masking release conditions.

There were two versions of experimental procedure. In the version used with the younger children (under 7.5 years of age), pictures on a video monitor signaled the listening intervals, and the children responded by pointing to the picture associated with the signal. The experimenter entered the response and visual feedback was provided via the action of an animated figure, for example, a frog turning into a prince (e.g., Veloso *et al.*, 1990). In a second version, older children and adults received visual cues of listening intervals, entered responses, and received visual feedback via a handheld response box. Whereas it is possible that use of different methods with the younger children could have introduced differences into the data, our own pilot work using a wide range of experimental paradigms has suggested that these two methods most effectively maintain attention of these very different populations.

3. Listeners

All listeners had pure-tone detection thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996). None had a history of chronic ear disease, and none had a known history of otitis media within a 3-year period preceding testing. There were 30 child listeners (11 females and 19 males), ranging in age from 4.5 to 9.0 years, with a mean age of 6.6 years. Because previous research suggested a developmental effect in the preschool years, effort was made to recruit children in this age range. The 12 children under 6.0 years of age will be re-

ferred to as the *preschool* group. There were 20 adult listeners (13 females and 7 males), ranging in age from 19 to 33 years, with a mean age of 24.5 years. All listeners were paid for participation and provided data in 1–2 sessions lasting no more than 1 h each.

Each listener was assigned to one of the two masking release conditions. In the conditions examining masking release for the MBD manipulation, there were 11 adults and 13 children. In the conditions examining the temporal fringe manipulation, there were 9 adults and 17 children. The reason for restricting each listener to only one of the masking release conditions was to limit the possibility that susceptibility to informational masking in one set of conditions might become unstable due to exposure to other condition sets. Pilot data on a small number of adult listeners suggested that such instability was a possibility. Although we wished to reduce the possibility of such instability in performance, we wanted to be able to measure any instability in susceptibility to informational masking that might occur within the conditions completed. For that reason, each listener completed a final set of thresholds in which the MBS condition was repeated. This allowed us to determine whether, during the course of testing, the listener might have learned strategies that reduced informational masking in the baseline condition.

B. Results and discussion

Thresholds for detection in quiet were significantly lower for the adults than for the children by 5.3 dB ($t_{46.6}=5.3; p<0.001$). For the children, the threshold in quiet ranged from 1.8 to 22.8 dB SPL, and for adults the threshold in quiet ranged from −1.0 to 8.3 dB SPL. Because the signal level at threshold in quiet varied, the signal level used in the masking conditions, set at 30 dB *re*: detection in quiet, also varied.

The top two panels of Fig. 2 plot results for experiment 1, with listeners in the MBD group shown in the top panel and listeners in the fringed-MBS group shown in the middle panel. The bottom panel shows results for experiment 2 (see below). In each panel, the masker levels at threshold for the initial MBS condition and the masking release condition are plotted in dB SPL as a function of age. In examining the masked thresholds, bear in mind that, because the signal level was fixed at 30 dB SL and the masker level was varied, a relatively lower masker level at threshold reflects relatively greater masking. A low masker level at threshold is expected in the MBS condition, consistent with a large informational masking effect, and higher masker level at threshold is expected in the MBD and fringed-MBS conditions, consistent with a release from masking. One striking feature of the present data is the enormous range of individual differences. Such variance is often noted in studies of informational masking (Neff and Dethlefs, 1995). In the MBS condition of the present study, the masker level at threshold spanned approximately 45 dB within the adult group, whereas the variation in this condition was less among the children (approximately 28 dB). In contrast, variability was very high among the children in the MBD and fringed-MBS conditions, whereas variability was relatively low for the adults in these
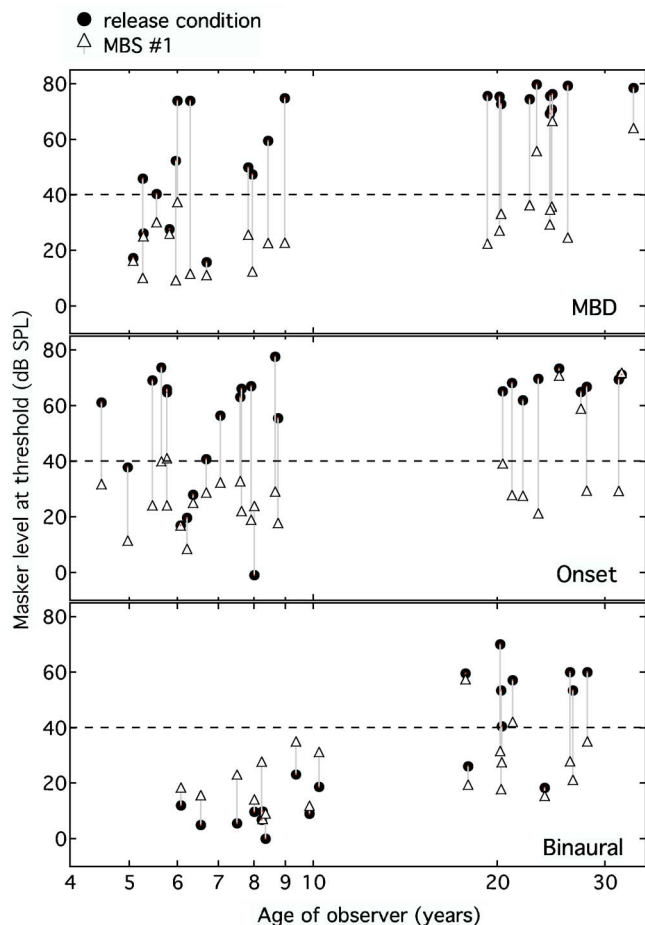
FIG. 2. Masker level at threshold is plotted in dB SPL as a function of age for individual listeners. The dashed line indicates the (approximate) maximum sensitivity achieved by child listeners in the *multiburst-same* (MBS) condition. The top, middle, and bottom panels are associated with the MBD (13 children, 11 adults), fringed-MBS (17 children, 9 adults), and binaural (10 children, 10 adults) conditions, respectively.

TABLE I. Masker levels at threshold (dB SPL) and masking release (dB) for experiments 1 and 2. Standard deviations are in parentheses. The "matched adult" data are for the subset of adults who obtained a masker level at threshold of 40 dB or lower in the initial MBS condition (comparable to children). Values connected by dashed lines are instances where child performance was significantly worse than for adults (comparisons are between children and all adults for the masker levels at threshold and between children and matched adults for masking release). A single asterisk denotes masking release that was statistically significant. The double asterisk denotes significant *increase* in masking.

		Adult		Child		Matched adult
MBD cue, Expt. 1	n	11		13		8
MBS (1st est)		39.1	-------------	20.1		30.5
		(15.7)		(8.9)		(5.3)
MBD		75.3	-------------	46.7		74.2
		(3.3)		(20.5)		(3.1)
MBS (2nd est)		44.9	-------------	20.8		35.0
		(19.7)		(12.0)		(11.7)
Masking release				26.6[*]	-------------	43.7[*]
				(20.7)		(7.3)
Fringe cue, Expt. 1	n	9		17		6
MBS (1st est)		41.8	-------------	25.3		29.1
		(19.9)		(9.0)		(5.8)
Fringed-MBS		67.9	-------------	50.5		66.9
		(3.5)		(22.7)		(2.9)
MBS (2nd est)		55.6	-------------	20.5		46.7
		(17.7)		(9.5)		(11.7)
Masking release				25.3[*]	-------------	37.7[*]
				(20.7)		(7.4)
Binaural cue, Expt. 2	n	10		10		8
MBS (1st est)		29.6		19.4		24.6
		(12.9)		(9.6)		(7.1)
Binaural		49.9	-------------	10.0		47.8
		(16.4)		(6.7)		(17.9)
MBS (2nd est)		41.4		22.0		36.4
		(20.7)		(9.9)		(20.3)
Masking release				−9.4[**]	-------------	23.2[*]
				(7.0)		(13.8)

conditions. Because of the gross differences in variability between groups, nonparametric methods were used to examine group differences and some within-group differences in the statistical tests reported below. Nonparametric tests were used in cases where the Levene's test of equality of variances (Levene, 1960) indicated unequal variances.

1. Developmental results for informational masking

Among the children, the masker levels at threshold in the MBS condition were uniformly low, with most points falling between 10–40 dB SPL, indicating appreciable informational masking. This is somewhat in contrast to the adults, where some listeners showed low masker levels at threshold (10–40 dB SPL), but others tolerated considerably higher masker levels (up to 72 dB SPL). Thus, compared to adults, children showed informational masking that was larger and more consistent (see Fig. 2). Masker levels at threshold in the initial MBS baseline condition were significantly better for adults than children (see Table I). In the MBD group, the adult MBS threshold was 19 dB better (Mann-Whitney U =20.0; p=0.002), and in the fringed-MBS group, the adult MBS threshold was 16.5 dB better (Mann-Whitney U=38.0; p=0.03). This developmental effect is consistent with previ-

ous work showing more informational masking in pre-school and young school-aged children than adults (Allen and Wightman, 1995; Oh *et al.*, 2001; Wightman *et al.*, 2003). The fact that this difference is modest when contrasted with the individual variability within listener group, which was on the order of 50 dB for adults and 30 dB for the children, should be taken as evidence that the children were only somewhat more susceptible to informational masking than adults.

The wide range of adult results in the MBS conditions allows identification of adults with baseline informational masking comparable to the child listeners: a subset of 70% of the adults had masker levels at threshold below 40 dB SPL in the initial MBS condition, comparable to those of the typical child listener (8 of 11 adult listeners in the MBD group and 6 of 9 adult listeners in the fringed-MBS group). This division is indicated by the dashed horizontal lines in Fig. 2, and will be referred to again below in the statistical analyses assessing amount of masking release.

2. Developmental results for release from informational masking

In the adults, both monaural release conditions consistently improved performance in all listeners who had appreciable informational masking in the MBS baseline condition. Adults showing very little informational masking release were presumably already performing near ceiling in the baseline condition. Children, on the other hand, did not benefit uniformly from the release manipulations. In one extreme case, performance of an 8-year-old child deteriorated substantially in the fringed-MBS condition (see Fig. 2), a result that replicated with additional data collection several days after the first session. Despite this nonuniformity, the performance of most children tended to improve in both of the release conditions.

Of primary interest here was whether children differed from adults in terms of the amount of release from informational masking. One coarse way in which this was assessed was simply in terms of the masker level at threshold in the masking release conditions. The Mann-Whitney U nonparametric test indicated that children had significantly lower (poorer) masker levels at threshold than adults in both of the monaural informational masking release conditions. In the MBD condition, this difference was 28.6 dB ($U=22; p=0.03$), and in the fringed-MBS condition it was 17.4 dB ($U=31; p=0.02$) (see Table I). Note that in this comparison, all of the adults (even those who showed very little informational masking in the baseline condition) were included. The comparison of informational masking release between adults and children is complicated somewhat by this fact, as the adults who did not show much informational masking in the baseline condition could not show much release from masking (ceiling effect). One general observation is that, for the adult listeners who did demonstrate low masker levels at threshold in the baseline condition, the masker levels at threshold increased substantially for the release condition in every case. In contrast, there were some cases among the children where the masker levels at threshold remained low in the release condition. We therefore performed statistical assessment of group differences in release from informational masking comparing children and the subset of 70% of adults who showed comparable baseline informational masking. Masking release was defined as the difference in the masker level at threshold between the first block of MBS and the monaural release condition, either MBD or fringed-MBS. Mean masking release is shown in Table I. In the MBD condition, the average masking release was 17.1 dB greater for the adults than for the children. This difference was significant ($U=22; p=0.03$). In the fringed-MBS condition, the average masking release was 12.4 dB greater for the adults than for the children. This difference was not significant ($U=31; p=0.18$). In order to confirm that children actually showed a significant masking release in both masking release conditions, Wilcoxon signed ranks tests were performed in which the initial masker level at threshold in the MBS condition was compared to the masker level at threshold in the masking release condition (either MBD or fringed-MBS).

These tests indicated significant masking release for the children in both the MBD condition ($z=2.5; p=0.01$) and the fringed-MBS condition ($z=3.2; p=0.001$).

Three brief comments regarding the above statistical outcomes are in order. For each, recall that the statistical outcomes pertain to the subset of adults who showed substantial informational masking in the initial MBS baseline. The first comment pertains to the fact that the children achieved a masking release that was not statistically different from that of the adults in the MBS/fringed-MBS comparison, but achieved significantly less masking release than the adults in the MBS/MBD comparison. A feature of the data that is in line with this outcome is that, whereas many children performed like the adults in the fringed-MBS condition, fewer did so in the MBD condition (cf. Fig. 2). For example, in the fringed-MBS condition, the worst adult masker level at threshold was approximately 60 dB SPL, and over half of the children achieved a masker level at threshold that was as good or better than this value. In the MBD condition, the worst adult masker level at threshold was approximately 70 dB SPL, and only 3 of 13 children achieved a masker level at threshold that was as good or better than this value. Overall, these results support a tentative interpretation that children in the age range investigated here are more likely to show adult-like levels of masking release for cues related to temporal asynchrony than for cues related to spectro-temporal coherence. The second comment pertains to the finding that adults and children did not differ significantly in terms of masking release in the MBS/fringed-MBS comparison. It is nevertheless clear that, whereas all adults showed masking release here, some of the children did not. The third comment pertains to the finding that children showed a significant release from masking for both of the monaural cues investigated here. Examination of Fig. 2 reveals that this result must be interpreted with care, as there were instances, for both cues, where individual children showed little or no masking release.

Comparisons of *preschool* and *young school-aged* children yielded mixed evidence regarding a difference in informational masking release between these groups. For the children who listened to the MBD release condition, 3 of 6 *preschool* children had the smallest masking release obtained in this condition (just 1–2-dB improvement), but 2 of 6 had a masking release above the mean for the child group (36- and 43-dB improvement). For the children who listened to the fringed-MBS release condition, 5 of 6 of the *preschool* children had masking release above the mean for the child group, and the remaining child had a release just 1 dB less than the mean. These results, while inconclusive, suggest that any difference between *preschool* and *young school-aged* children in ability to make use of the MBD and fringed-MBS cues for informational masking release is highly listener dependent, and may be more pronounced in, or restricted to, the MBD manipulation.

3. Developmental results related to learning

Because the MBS baseline condition was run both before and after the masking release condition, these data can be used to estimate changes in performance with listening

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Hall, Buss, and Grose: Informational masking release   1609
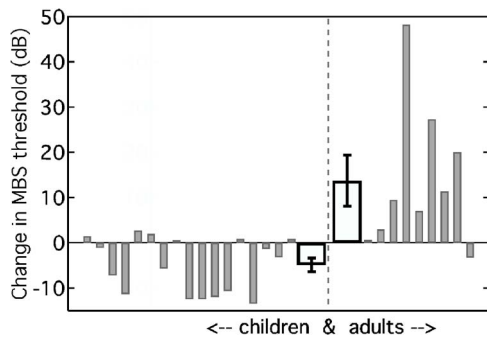
FIG. 3. Individual (narrow, filled bars) and average (wide, unfilled bars) results for the change in the masker level at threshold between the first and second estimates for the MBS condition. The data are from the listeners in the fringed-MBS group of experiment 1. The error bars show plus and minus 1 standard error of the mean.

experience. Such a change could result from exposure to the masking release condition, additional exposure to the baseline condition itself, or central factors unrelated to the particular stimuli encountered (e.g., fatigue or general familiarity with testing procedures). The mean MBS results and standard deviations appear in Table I. In statistical tests of change in the MBS result for the adults, only those listeners with a masker level at threshold below 40 dB for the first estimate were included. Children showed relatively consistent intersubject variability across the estimates for the first and second MBS conditions, so change in masker level at threshold was assessed using a t-test. In contrast, adults showed high intersubject variability in the second MBS condition, so the nonparametric Wilcoxon test was used to assess change in performance here. For the MBD release group, there was no significant change in performance over the two MBS conditions for either group ($t_{11}=0.2$; $p=0.84$ for children and $Z=1.4$; $p=0.16$ for adults). For the fringed-MBS group, both groups showed significant change. The adults improved in the second MBS estimate ($Z=2.0$; $p=0.04$), while the children showed a small but significant decrease in performance ($t_{16}=3.3$; $p=0.005$). These differences are shown in Fig. 3, with individual data in filled, narrow bars (ordered by age) and mean data in unfilled, wide bars. As with the initial MBS condition, adults had better thresholds than children for the second MBS condition (see Table I) both in the MBD group (Mann-Whitney $U=20.0$; $p=0.004$), and in the fringed-MBS group (Mann-Whitney $U=5.5$; $p<0.001$).

There is no evidence that sensitivity of the children in the baseline condition improved over the course of testing. For adults, however, exposure to the fringed-MBS release condition improved thresholds. It is not obvious why the performance in the MBS condition improved for the adults in the context of the fringed-MBS condition but not in the context of the MBD condition. It is possible that the improvement within the fringed-MBS context is related to the fact that stimuli are quite similar between the MBS and the fringed-MBS conditions. These conditions are identical (except for the leading masker bursts in the fringed-MBS condition), and it is possible that this facilitated the transfer of strategies learned in the fringed-MBS condition to the final

MBS condition. A very different strategy may be utilized in the MBD condition, perhaps involving spectral variations that are absent from the MBS condition. A caveat to this interpretation is that the observed differences may have arisen from individual differences between the randomly assigned listener samples comprising the two groups of adults.

III. EXPERIMENT 2

Whereas experiment 1 examined the ability of children to make use of monaural cues to reduce informational masking, Wightman *et al.* (2003) studied the ability of children to make use of binaural cues related to sound lateralization. In that study, the informational masking condition involved signal and masker presentation to a single ear, and the informational masking release condition involved signal presentation to one ear and masker presentation to the opposite ear. This manipulation eliminated informational masking in adults, reduced or eliminated informational masking in some school-age children, but had little or no effect on performance of the preschool children. In some contrast, the present experiment 1 indicated little evidence of a consistent difference between the preschool and school-aged children. Overall, experiment 1 (where monaural cues for masking release were employed) showed more evidence for release from informational masking in younger children than in the study of Wightman *et al.* (2003) (where cues for masking release were binaural in nature). Several aspects of the methods of the Wightman *et al.* study differed substantially from those of experiment 1, making comparisons across the monaural and binaural conditions problematic. The most obvious was that there was no masking energy in the signal ear in the masking release condition of the Wightman *et al.* study. Furthermore, stimuli in the Wightman *et al.* study were presented at a steady level for the 360-ms presentation interval, rather than bursting on and off, and the signal level was adjusted to obtained threshold, rather than the masker level adjustment used here. Finally, each listening interval was preceded by a signal cue in the Wightman *et al.* study. Because of the large differences in method between studies, we decided to run a binaural condition that shared features with both the current experiment 1 and the study of Wightman *et al.* The intent was to examine whether developmental effects for release from informational masking may be relatively pronounced for binaural spatial hearing cues related to interaural intensity differences.

A. Methods

1. Stimuli

In an initial implementation of this experiment, stimuli were identical to those in the MBS condition of experiment 1, with the exception that, whereas the signal was presented to the left ear only, masker components were presented diotically. This was similar to a previous approach by Kidd *et al.* (1994), where a release from informational masking was presumably due to the fact that the signal was lateralized to one ear, and the masker was lateralized to the midline. Our initial data using this method showed masking release that was relatively unreliable, so we modified the approach such that the masker components in the contralateral ear were pre-

sented at a level 10 dB higher than the masker components presented in the ipsilateral ear (masker frequencies and phases were identical). The rationale for this modification was to enhance the difference between the lateralized images of the signal versus the masker. This will be referred to as the *binaural* condition. As pointed out above, this masking release condition was like the MBS condition of experiment 1, except that masker components were presented to both ears. The condition was intended to be akin to the release condition in the Wightman *et al.* (2003) study, in that the masker stimuli were manipulated in such a way as to promote lateralization of the masker to the ear contralateral to the signal.

2. Procedures

The procedures for experiment 2 were identical to those of experiment 1, with the exception that the maximum level of the ipsilateral masker was reduced from 83 dB SPL (experiment 1) to 73 dB SPL.

3. Listeners

Listeners participating in this experiment were demographically quite similar to those participating in experiment 1. All had pure-tone detection thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 1996), and none had a history of chronic ear disease. There were ten child listeners, ranging in age from 6.1 to 10.2 years, with a mean age of 8.2 years. There were ten adult listeners, ranging in age from 18 to 28 years, with a mean age of 22.3 years. All listeners were paid for their participation and provided data in 1–2 sessions lasting no more than 1 h each. None of the listeners in experiment 2 had previously participated in any psychophysical studies of informational masking, including those performed in experiment 1.

B. Results and discussion

Thresholds for detection in quiet were significantly lower for the adults than for the children by 5.5 dB (t_{18} =2.3; p=0.03). The bottom panel of Fig. 2 plots the masker level at threshold (ipsilateral ear) for the first block of MBS and the binaural block as a function of listener age.

1. Developmental results for informational masking

As in experiment 1, the initial MBS condition was associated with uniformly poor performance for the children (large informational masking), and a range of results among the adult listeners. Masker levels at threshold in the initial MBS baseline condition (see Table I) were higher (better) for the adults than for the children, but the probability value (U=25.5; p=0.06) did not reach statistical significance (α =0.05). As in experiment 1, there was a wide range of adult performance in the MBS condition, allowing identification of adults with baseline informational masking comparable to the child listeners: a subset of 80% of the adults had masker levels at threshold below 40 dB SPL, comparable to those of the typical child listener.

2. Developmental results for release from informational masking

Among the adults, the binaural manipulation improved performance in some of the listeners who had appreciable informational masking in the MBS baseline condition (see the bottom panel of Fig. 2). Children, on the other hand, showed a consistent lack of masking release, with most even showing additional masking in the binaural condition. Overall, the adults showed significantly higher (better) masker levels at threshold in the binaural condition than did the children (see Table I). The Mann-Whitney U nonparametric test indicated that children had significantly lower (poorer) masker levels at threshold than adults in the binaural condition (U=2.0; p<0.001). Paired t-tests were performed to compare the initial MBS and binaural conditions. For the adults, this was performed on the 80% of the listeners who showed baseline informational masking comparable to that for the children (see Fig. 2). The tests showed a significant improvement in performance for the adults (t_7=4.7; p =0.002) and a significant decline in performance for the children (t_9=4.3; p=0.002). We note that, if some of the children had demonstrated a masking release in the binaural condition, we would have tested younger children to explore differential effects of age (the youngest child in this experiment was 6.1 years of age). Among the children tested, there was no significant correlation between age and the amount of increased masking in the binaural condition (r=0.03; df =8; p=0.93).

The present results, where children actually showed a masking increase in the condition designed to result in masking release, contrast with the results of experiment 1 where children typically showed a decrease in masking in the masking release condition. This indicates that the ability of children to take advantage of masking release cues depends strongly upon the type of cue available. The present results indicate that a lateralization cue based upon interaural intensity differences was not utilized well by the children tested here.

In comparing the present binaural results to those of Wightman *et al.* (2003), it might seem incongruous that, whereas some of the children showed masking release in the Wightman *et al.* study, none of the children did so in the present study. It is possible that this difference is related to the overall salience of the cue for masking release in the different paradigms. In the Wightman *et al.* study, it is clear that the masking release cue in the binaural condition was highly salient among the adults, as there was essentially no informational masking among adults in that condition. In the present study, it can be inferred that the masking release cue in the binaural condition was less salient, as some of the adults continued to show substantial masking here. Thus, the failure of any of the children in the present study to show masking release in the binaural condition may be related to the relatively modest salience of the cue for masking release. The results of the present study are in general agreement with the conclusions of Wightman *et al.*, in that children were relatively poor in achieving benefit from the spatial hearing cue.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Hall, Buss, and Grose: Informational masking release   1611

3. Developmental results related to learning

As in experiment 1, the change in performance in the MBS condition was quantified as the difference in the masker level at threshold between the first and second estimates for MBS condition, and only the subset of adults with thresholds below 40 dB for the first estimate was included. The mean results appear in Table I. Although the change in the MBS threshold was greater for the adults than for the children, the change was not significant for either group ($t_7 = 0.99$; $p = 0.35$ for the children and $Z = -1.54$; $p = 0.12$ for the adults). As with the initial MBS condition, adults had better thresholds than children for the second MBS condition (see Table I), but the difference was not statistically significant (Mann-Whitney $U = 25.0$; $p = 0.06$).

4. Relation between informational masking and threshold in quiet, assessed across experiments 1 and 2

In both of the experiments of this study, it was found that the children had higher thresholds in quiet than the adults. Although not directly related to the main aims of the study, a question of interest is whether there was a relationship between the threshold in quiet and the masker level at threshold in the MBS condition. A *negative* relationship (high threshold in quiet associated with a low masker level at threshold) might be expected on the grounds that a high threshold in quiet will occur for poor listeners, and poor listeners may not be able to tolerate high masker levels in the MBS condition. On the other hand, a *positive* relationship (high threshold in quiet associated with a high masker level at threshold) might be expected on the grounds that a high threshold in quiet will result in the signal being presented at a relatively high SPL in the MBS condition (30 dB SL), therefore requiring a relatively high SPL masker level to mask it. A correlation performed on the data of the children of both experiments (40 children total) supported the latter expectation ($r = 0.61$; df $= 38$; $p < 0.001$). Because children had higher thresholds in quiet than adults, this correlation would suggest a bias toward lower informational masking in the MBS condition for children than adults (the higher thresholds in quiet for the children being associated with higher masker levels at threshold in the MBS condition). Nevertheless, the results of this study showed significantly *more* informational masking in the children than the adults in the MBS condition. This is likely due to the fact that, among adults, interlistener variability in informational masking overshadowed factors related to the threshold in quiet. This is supported by the observation that the correlation between threshold in quiet and the masker level at threshold in the MBS condition was not significant ($r = 0.24$; df $= 28$; $p = 0.20$) among the 30 adults tested in experiments 1 and 2.

IV. GENERAL DISCUSSION

A key strength of the present approach was that we were able to identify subsets of adults who showed informational masking that was comparable to that demonstrated by the children, thereby allowing reasonable tests of potential developmental differences in release from informational mask-

ing. The present results indicated no significant difference between adults and children for informational masking release related to temporal asynchrony (although there were several individual cases in which children did not show masking release here), and indicated significantly smaller masking release in children for cues related to spectro-temporal coherence (although most children achieved substantial masking release here). In general, there were remarkable individual differences in the ability of children to achieve masking release in the monaural conditions examined. These individual differences appeared to be related only poorly to age across the children tested here. The results of the present study indicate that the ability to achieve informational masking release from lateralization cues related to interaural intensity differences may develop relatively late. This result is consistent with previous results from Wightman *et al.* (2003), where children often continued to show considerable informational masking even when the masker was moved to the ear contralateral to the signal. Overall, these results suggest that children have a relatively poor ability to benefit from informational masking release cues that hinge upon the differential lateralizations of the signal and masker. In experiment 2, the binaural cue did not result in release from informational masking in any of the children, but instead resulted in a significant increase in masking in this group. Although the methodologies differed between experiment 2 and the Wightman *et al.* study, both studies used manipulations that would tend to isolate the percepts of the signal and maskers to opposite ears. The manipulation in the present study was less radical than in Wightman *et al.*, where the masker was physically absent from the signal ear in the masking release condition. Taken as a whole, the available results suggest that the development of the use of cues related to informational masking release is highly dependent upon the type of cue under investigation.

As Wightman *et al.* (2003) noted, caution should be used in generalizing results obtained in highly constrained laboratory conditions to more natural listening situations. For example, even though the results obtained to date show poor performance by children on masking release tasks that are associated with sound lateralization (listening via earphones), it is possible that such deficits would not occur in tasks involving sound localization in real environments, where the available cues are more natural than those occurring in highly controlled but unnatural earphone studies. In this regard, it should be kept in mind that a developmental effect associated with a particular type of cue might be related to a multitude factors, one of which is the context in which the cue (monaural or binaural) is presented. For example, in segregating sound sources in real environments, children might be adept in using spatial hearing cues in combination with the other cues that often occur in concert (e.g., cues related to harmonicity and/or onset asynchrony), but may be less efficient than adults when required to use isolated cues for sound segregation in less natural listening conditions.

It should also be cautioned that results indicating late development of the ability to use binaural cues in release from informational masking do not mean that children are

1612   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Hall, Buss, and Grose: Informational masking release

generally poor in using binaural cues for masking release, even in the context of highly controlled laboratory experiments. For example, the results of several experiments indicate that children have adult-like masking-level differences (MLDs) (Hirsh, 1948) in a relatively broadband masking noise by age 5–6 years (Hall and Grose, 1990; Grose *et al.*, 1997).

V. CONCLUSIONS

Results of experiment 1 suggest that children aged 4–9 years can make use of monaural cues associated with informational masking release, though at a somewhat reduced level when compared to adults with similar amounts of baseline informational masking. There was some suggestion that the ability to achieve masking release may develop earlier for the cue of temporal asynchrony than for the cue of spectro-temporal coherence. Results from experiment 2 showed that children approximately 6–10 years of age failed to show a masking release (and actually showed increased masking) when interaural level differences were provided that would tend to lateralize the signal and masker to opposite ears. Development of the ability to achieve masking release from informational masking under highly controlled conditions depends strongly upon the kind of cue under study.

ACKNOWLEDGMENTS

Allen, P., and Wightman, F. (**1995**). "Effects of signal and masker uncertainty on children's detection," J. Speech Hear. Res. **38**, 503–511.

ANSI (**1996**). ANSI S3.6–1996, "Specification for Audiometers" (American National Standards Institute, New York).

Arbogast, T. L., Mason, C. R., and Kidd, G. Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organisation of Sound* (Bradford Books, MIT Press, Cambridge, MA).

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Dannenbring, G. L., and Bregman, A. S. (**1978**). "Streaming vs fusion of sinusoidal components of complex waves," Percept. Psychophys. **24**, 369–376.

Darwin, C. J. (**1984**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.

Durlach, N. I., Mason, C. R., Kidd, G. Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L.,

Colburn, H. S., and Kidd, G. Jr. (**2003b**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Grose, J., Hall, J., and Dev, M. (**1997**). "The MLD in children: Effects of signal and masker bandwidths," J. Speech Lang. Hear. Res. **40**, 955–959.

Hall, J. W., and Grose, J. H. (**1990**). "The masking-level difference in children," J. Am. Acad. Audiol **1**, 81–88.

Hirsh, I. J. (**1948**). "Influence of interaural phase on interaural summation and inhibition," J. Acoust. Soc. Am. **20**, 536–544.

Kidd, G. Jr., Arbogast, T. L., Mason, C. R., and Walsh, M. (**2002**). "Informational masking in listeners with sensorineural hearing loss," J. Assoc. Res. Otolaryngol. **3**, 107–19.

Kidd, G. Jr., Mason, C. R., Arbogast, T. L., Brungart, D. S., and Simpson, B. D. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.

Kidd, G. Jr., 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.

Levene, H. (**1960**). "Robust tests for equality of variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, edited by I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (Stanford University Press, Stanford, CA).

Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lutfi, R. A. (**1990**). "How much masking is informational masking," J. Acoust. Soc. Am. **88**, 2607–2610.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Callaghan, B. P. (**1987**). "Simultaneous masking by small numbers of sinusoids under conditions of uncertainty," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Mahwah, NJ).

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., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

Richards, V. M., Huang, R., and Kidd, G. Jr. (**2004**). "Masker-first advantage for cues in informational masking," J. Acoust. Soc. Am. **116**, 2278–2288.

Richards, V. M., Tang, Z., and Kidd, G. D. Jr. (**2002**). "Informational masking with small set sizes," J. Acoust. Soc. Am. **111**, 1359–1366.

Veloso, K., Hall, J. W., and Grose, J. H. (**1990**). "Frequency selectivity and comodulation masking release in adults and in six-year old children," J. Speech Hear. Res. **33**, 96–102.

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, Mahwah, NJ).

Werner, L., and Bargones, J. (**1991**). "Sources of auditory masking in infants: Distraction effects," Percept. Psychophys. **50**, 405–412.

Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.

Yost, W. A. (**1991**). "Auditory image perception and analysis: The basis for hearing," Hear. Res. **56**, 8–18.

Binaural release from informational masking in a speech identification task[a)]

Frederick J. Gallun,[b)] Christine R. Mason, and Gerald Kidd, Jr.
Communication Disorders and Hearing Research Center, Sargent College, Boston University,
635 Commonwealth Avenue, Boston, Massachusetts, 02215

Binaural release from informational masking (IM) was examined in a speech identification task. Target and masker sentences were processed into mutually exclusive frequency bands, thus limiting energetic masking (EM), and presented over headphones. In a baseline condition, both were presented monotically to the same ear (T_mM_m). Despite minimal frequency overlap between target and masker, the presence of the masker resulted in reduced performance, or IM. Presenting the target monotically and the masker diotically (T_mM_0) resulted in a release from IM. Release was also obtained by imposing interaural differences in level (ILDs) and in time (ITDs) on the maskers (T_mM_{ILD}, T_mM_{ITD}). Any masker with a perceived lateral position that differed from that of a truly monaural stimulus resulted in a similar amount of release from IM relative to T_mM_m. For binaural targets and maskers (T_0M_{ILD}, T_0M_{ITD}), release was seen whenever ITDs or ILDs differed between target and masker. These results suggest that binaural cues can be very effective in reducing IM. Because mechanisms based on differences in perceived location make predictions that are similar to those of nonlocation-based binaural mechanisms, a variant of the equalization-cancellation model is also considered. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1984876]

I. INTRODUCTION

Kidd *et al.* (2005) reported that with stimuli in which the spectral overlap of target and masker was minimized and the binaural parameters fluctuated randomly, a monaural target was identified more easily with a binaural than a monaural masker. This study attempted to systematically examine if perceived location, which was uncontrolled in that study, could have allowed listeners to distinguish between the auditory stimuli. Situations in which performance is governed by the presence of masker energy in the frequency regions occupied by the target (i.e., "energetic masking" or EM) make it difficult to differentiate the role of perceived location from the effects of nonlocation-based mechanisms that result in binaural release from masking (Colburn and Durlach, 1978, pp. 498–501). Using stimuli that fall into distinct auditory filters reduces the likelihood that observed effects can be explained by mechanisms that improve signal-to-noise ratio within a filter. Stimuli that meet these criteria include the spectrally processed speech described in Arbogast *et al.* (2002) and used in Kidd *et al.* (2005). In those studies, the target and masker were sentences comprised of sets of very narrow, mutually exclusive frequency bands, with which target identification performance was substantially reduced relative to the performance obtained with a noise-masker processed to have the same long-term spectrum as the speech. This additional masking has been attributed to "informational" masking (IM).

For the stimuli used in Arbogast *et al.* (2002) and Kidd *et al.* (2005), IM was due largely to listener uncertainty about which sentence was the target. Unlike with a noise masker, the listener was confronted with two very similar sentences, only one of which was the target. Although the talkers speaking the sentences were always different for masker and target, the narrowband processing disrupted fundamental frequency and intonation information. Both of these factors are known to provide important cues that listeners use to segregate natural competing speech (Darwin and Hukin, 2004). The result of the processing, then, was that listeners were forced to rely upon the temporal correlations between bands and timbral differences to group the bands successfully. Faced with this difficulty, listeners were motivated to make use of any additional information available—such as perceived location—to determine which bands corresponded to the target and which to the masker.

In Kidd *et al.* (2005), listeners were asked to report the keywords associated with a target sentence in the presence of a masker sentence processed into mutually exclusive frequency bands (called "different-band speech" or DBS). When both sentences were presented to the same ear, performance declined significantly relative to when no masker was present. The surprising result that motivated the experiments described here occurred when a second noise masker occupying the same bands as the speech masker (but different bands from the target) was added in the opposite (contralateral) ear. The addition of this "different-band noise" (DBN) masker in the contralateral ear *reduced* the amount of observed masking significantly by *increasing* identification performance. This result was unexpected because it is at odds with the assumption that noise in one ear does not energeti-

---

cally mask information in the opposite ear. Nonetheless, these results resembled those obtained when the noise was added to the same ear as the masking speech. Thus the noise "masked the masker" even though it was in the opposite ear. The reason for this is almost certainly that the two maskers, speech in the target ear and noise in the nontarget ear, shared the same narrow frequency bands. This resulted in binaural cues that were fluctuating randomly within each masker band and were uncorrelated between bands. It seems likely that some aspect of this situation resulted in either decreased intelligibility or—the possibility examined in this study—increased separation for the masking speech.

In general, it has been found that adding noise to the contralateral ear has very little effect on speech recognition performance in the ipsilateral ear. Brungart and Simpson (2002), for example, reported that performance was not changed for an ipsilateral speech identification task when a broadband noise was added to the opposite ear. Similarly, Brungart *et al.* (2005) found that when no speech masker was present, adding different-band noise to the contralateral ear produced performance statistically identical to that obtained when no masker was present. In the cases where adding a contralateral masker does have an impact, the finding is that adding a speech masker to the contralateral ear *increases* the amount of masking caused by an ipsilateral speech masker (e.g., Brungart and Simpson, 2002; Brungart *et al.*, 2005). The main difference between broadband and narrowband maskers is that a broadband masker has the same effect on both target and masker, but a narrowband masker has the ability to affect only one. This may explain why a release from masking has not been seen previously when broadband maskers have been added to the contralateral ear.

Kidd *et al.* (2005) suggested that the interaction of the narrowband DBS and DBN stimuli presented to opposite ears may have created a binaural image that allowed listeners to distinguish monaural-only target bands from masker bands on the basis of differences in perceived location. In addition, the combination of DBS and DBN maskers across ears may have reduced the intelligibility of the speech masker by some form of binaural interaction (for example, binaural summation) thus reducing its ability to compete with the target. The purpose of the experiments described in this paper is to examine whether perceived lateralization derived from narrowband binaural interactions is capable of explaining the finding reported by Kidd *et al.* (2005). Two main difficulties with evaluating the role of lateralization in the original study were that (1) the DBN masker may have been interfering with the intelligibility of the speech masker (second possible explanation above) and (2) it is not clear exactly what perceived lateralizations were associated with those stimuli. Because the DBN and DBS maskers were uncorrelated, the interaural differences that were created would have differed among the bands. Furthermore, the main finding occurred for cases where the target was monaural and the masker was either monaural or was binaural, leaving open the possible complication of a monaural versus binaural effect regardless of any differences that occurred in lateral position. The current experiments were intended to evaluate the above-offered lateral

position explanation for the Kidd *et al.* findings while minimizing or controlling for these potentially confounding variables.

II. GENERAL METHODS

The experiments described in this paper were carried out largely simultaneously with the data collected for the experiments described in Kidd *et al.* (2005). The perceived lateralization study (presented as experiment 1) was conducted several months later. The three listeners were the same and the design of the experiments was influenced by the fact that these conditions were introduced as part of a larger group of conditions already in progress.

A. Listeners

Three females, aged 22–25 years, participated in 2 h sessions three to four times each week. All were very familiar with the stimuli through several thousand trials of listening experience prior to the introduction of these conditions. All had audiometric thresholds of 20 dB HL or better in each ear for octave frequencies from 250 to 8000 Hz.

B. Stimuli

The stimuli resembled those of Arbogast *et al.* (2002). Sentences spoken by four male talkers from the coordinate response measure (CRM) corpus of Bolia *et al.* (2000) were used. Each sentence has the structure: "Ready [callsign] go to [color] [number] now." There are eight callsigns, four colors ("white," "red," "green," and "blue") and eight numbers (1–8). Each sentence wave form was digitally preprocessed to restrict the frequency content. Down-sampling from 40 to 20 kHz was followed by a high-frequency emphasis using a first-order high-pass butterworth filter with a cut-off of 1200 Hz. Each wave form was then passed through a bank of fifteen, approximately one-third octave, fourth-order butterworth filters. The filters were evenly spaced on a logarithmic scale from 215 to 4891 Hz with successive center frequencies spaced at a ratio of 1.25, making them approximately one-third of an octave apart as well. Half-wave rectification and low-pass filtering at 50 Hz by a fourth-order butterworth filter extracted the amplitude envelope within each band.

Thus, each wave form was reduced to a set of amplitude envelopes, each corresponding to one of the fifteen frequency bands. Processed sentences were generated by randomly choosing a subset of the fifteen envelopes and using that subset to modulate pure tones with frequencies equal to the center frequencies of the bands, amplitudes of one, and starting phases of zero. Digitally adding the envelope-modulated tones resulted in a sentence with restricted frequency energy and reduced harmonic structure but with temporal content that reflected the original sentence. The processed sentences (when composed of at least six bands and presented at an adequate level) were perfectly intelligible to listeners (see also Brungart *et al.*, 2005) with only a few minutes of exposure—although this is probably due in part to the closed-set nature of the task. After preprocessing, the sentences were up-sampled to 50 kHz, scaled to equal rms amplitude, and saved to disk for later presentation.

Processed sentences were presented via a Tucker-Davis Technologies (TDT) System II hardware array. The digital to analog conversion was at 50 kHz with 16 bits amplitude quantization and all signals were low-pass filtered at 7.5 kHz. The attenuation and signal routing as well as the relative timing of the signals were varied across conditions and are discussed in the following. Stimuli were presented to the listener over TDH-50 headphones in a double-walled sound-attenuating (IAC) booth.

III. EXPERIMENT 1: PERCEIVED LATERAL POSITION

In order to examine whether increasing masker lateralization can cause release from informational masking, it was first important to show that variations in interaural parameters resulted in changes in perceived lateralization for these listeners with these stimuli. Consequently, an experiment was conducted in which listeners were not required to identify the sentences, but simply to rate their lateral positions.

A. Methods

On all trials, the stimulus was a sentence from the CRM corpus processed into six narrow frequency bands using the method described earlier (Sec. II B). In the first experiment, the sentence was presented to the right ear at 50, 60, or 70 dB SPL and the sentence in the left ear was either (1) not present (monaural presentation), (2) presented at a different level from the sentence at the right ear [interaural differences in level (ILD) condition], or (3) presented at the same level as the sentence at the right ear but either advanced or delayed in time relative to the onset of the sentence in the right ear [interaural differences in time (ITD) condition].

For the ILD trials, the left-ear sentence was presented at levels that varied in 10 dB steps from 30 to 70 dB SPL. To calculate the ILD created for a given combination, the level at the left ear (L_L) is subtracted from the level at the right ear (L_R). A positive ILD denotes a situation in which the ILD favors the right ear (L_R exceeds L_L), while a "negative" ILD indicates that L_L exceeds L_R.

For the ITD trials, onset and ongoing ITDs were created by delaying the onset of one of the sentences (left or right) by either 300 or 600 μs. To calculate ITD, the onset time for the left-ear sentence (T_L) is subtracted from the masker onset time for the right-ear sentence (T_R). In cases where the right-ear sentence was delayed, T_L exceeded T_R and the ITD was negative. The ITD values used were 600, 300, 0, −300, and −600 μs.

On 80% of the trials, the speech was binaural. On the remaining trials, the speech was monaural at the right ear. Listeners reported the perceived lateralization of a sound source on a rating scale—a method similar to that of Sayers and Toole (1964). After listening to each sentence, listeners rated the location on a nine-point scale. A midline location was to be indicated with a rating of five, all the way to the left with a rating of one and all the way to the right with nine. No feedback was given and listeners were not informed of the relative likelihoods of monaural and binaural sentences. The precise distribution was random and each listener completed 15 blocks of 50 trials each. Due to the truly ran-
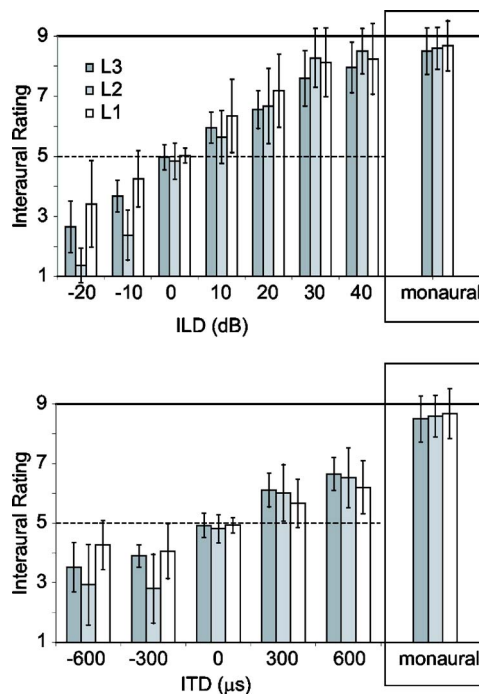
FIG. 1. Average ratings associated with the perceived interaural locations of the stimuli presented in experiment 1. The top panel indicates the ratings associated with ILDs and the bottom those associated with ITDs. Error bars indicate ± one standard deviation around the mean rating.

dom design, the actual numbers of each sentence type were not equal, but this method did result in at least 15 ratings of each combination of interaural parameters for each listener. The monaural signals were rated at least 75 times by each listener.

B. Results

The top panel of Fig. 1 shows the average ILD ratings and the bottom panel the average ITD ratings. The bars on the far right of each panel show the ratings for monaural stimuli. Both data sets show a monotonic increase in ratings that occurs for all subjects for both ILDs and ITDs. The effect of ILD on listener ratings was examined by a two-way ANOVA on masker level at the right and left ears. The effect of level at the right ear was significant [$F_{(2,4)}=51.097$, $p=0.001$] as was level at the left [$F_{(4,8)}=85.939$, $p<0.001$]. The lack of a significant interaction [$F_{(8,16)}=1.929$, $p=0.125$] shows that as the level at one ear increased, the effect of the relative level was not changed. This confirms the use of ILD as a measure of lateralization independent of level at either ear. A two-way ANOVA on masker level and ITD showed no significant effect of level [$F_{(2,4)}=0.544$, $p=0.588$], but a significant effect of ITD [$F_{(4,8)}=27.414$, $p<0.001$]. This confirms that although the range of ratings was smaller than for ILD, the various ITDs were in fact perceived as different.

Regardless of level at the left ear, ILDs between −20 and 20 dB were rated in a manner significantly different from monaural ($p<0.05$) but 30 and 40 dB were not ($p>0.10$). All of the mean ratings at the various ITDs were statistically different ($p<0.01$) from the monaural ratings. Thus, listen-

ers rated all of the ITDs as producing perceptions that differ in laterality from a monaural stimulus, but only stimuli with ILDs of greater than 20 dB are rated as occupying a lateral position similar to that of a monaural stimulus. These results are in accordance with those of Hafter and Kimball (1980), who found that the ability to discriminate a large ILD from a truly monaural signal declines to chance once the ILD exceeds 45 dB. The second set of comparisons involving the ILDs and ITDs used in experiment 3 found that 10 and 20 dB ILDs were rated as different from 0 dB ($p < 0.05$), but that while 600 μs was different from 0 μs ($p < 0.05$), 300 μs was not ($p = 0.07$).

C. Trading ratios

One method of calculating the amount of ILD needed to produce the same lateralization as a given ITD (and vice versa) is by presenting a signal lateralized with one cue (ITD or ILD) and asking listeners to point to the location of the signal with an acoustic pointer that is moved by varying the other cue. The values found with this method tend to fall between 20 and 50 μs/dB [reviewed, along with methods that give divergent results, in Hafter and Jeffress (1968)]. The rating responses from the first experiment can also be used to establish a rough estimate of the "trading ratio" between interaural differences in level and in time. Since ratings of lateral position were assigned to all of the ITDs and ILDs using the same scale, the trading ratio was calculated by performing three linear regressions: first on the ratings of ILD and the ILDs presented, second on the ITD ratings and the ITDs presented, and third, between the predictions generated by the first two functions. This allowed an estimate of which ILD and ITD values result in the same rating of lateral position. For these listeners and these stimuli and equipment, the rating data produced a fairly large tradeoff between ITD and ILD: 47.6, 32.1, and 36 μs/dB for L1, L2 and L3, respectively. The mean was 38.6 μs/dB. In addition, the values associated with a rating of five (midline) were displaced from zero for the three listeners by 103.2, −67.3, and −139.7 μs, respectively. This corresponds to values of 2 to 4 dB in ILD, according to the trading ratios. While large, these values are not outside the expected range, especially given the fact that no ITDs were presented between 0 and 300 μs and no ILDs were presented between 0 and 10 dB. Given this very coarse sampling of interaural differences, the trading ratios obtained seem quite reasonable. Most significantly for interpreting the data from experiments 2 and 3, it can be seen from these values that the maximum ITDs employed, −600 and 600 μs, cover a range of roughly −20 to 20 dB in ILD.

IV. EXPERIMENT 2: IDENTIFICATION, TARGET MONAURAL

The second experiment was designed to examine the ways in which listeners make use of interaural differences in the masker to improve performance on an IM task where the target is monaural. In particular, the goal was to examine whether release from IM can be predicted from the ratings of lateral position obtained in experiment 1.
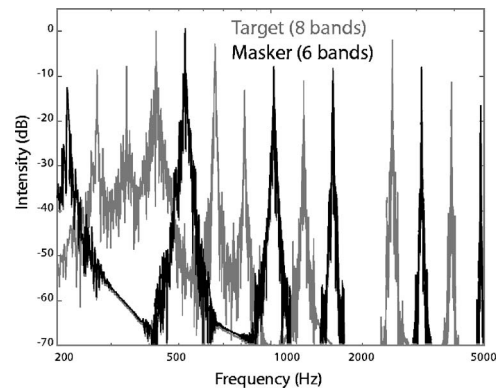
FIG. 2. Example of the magnitude spectra of target (grey) and masker (black).

A. Methods

The task was single-interval forced-choice closed-set speech identification using processed versions of the coordinate response measure (CRM) corpus of spoken sentences (Bolia *et al.*, 2000). Each trial involved presentation of two sentences: a "target" (identified by the presence of the callsign "Baron") and a "masker." Each unprocessed target sentence was first paired with a random masker sentence seven times, resulting in a total of 896 target-masker pairs. Target-masker pairs were constrained to differ on talker, callsign, color, and number. Each sentence was then processed into a collection of randomly chosen narrowband signals using the method described in Sec. II B. In order to ensure that the frequency bands comprising the target and the masker were mutually exclusive, eight target bands were randomly chosen and then six of the remaining seven bands were assigned to the masker. Target and masker-sentence lengths were not equalized. An example of the frequency content of a target and masker pair appears in Fig. 2. It can be seen that the target sentence (grey) and the masker sentence (black) each contain mutually exclusive bands and that the peaks are well separated. For most combinations of bands, the masker was at least 40 dB lower than the target at the center frequency of the target band and was always at least 20 dB lower.

On each trial, the listener was presented with a single target-masker pair and was instructed to report the color and number keywords associated with the target, designated by the presence of the callsign "Baron." In the masker sentence, any of the other callsigns could appear. The four talkers, four colors, and eight numbers resulted in 128 unique target sentences. After stimulus presentation, the listener made two responses, one indicating the color and the other the number associated with the target sentence. The first response (color) had four alternatives and the second response (number) had eight. Response was via a keyboard with the numbers and color names labeled on the keys. Feedback was given via an LCD screen.

Three conditions were tested. In the first condition, the monaural 60 dB SPL target was added to a monaural masker that could be at one of three levels: 50, 60, or 70 dB SPL. This resulted in three T/M ratios at the target ear ("ipsilateral" T/M values): 10, 0, and −10 dB, respectively. Adapting terminology from the binaural masking literature, this condi-

tion is referred to as T_mM_m: target monotic, masker monotic. The second condition, T_mM_α, consisted of presenting the masker sentence to both the right (target or ipsilateral) and left (contralateral) ears with an interaural level difference (ILD). The target sentence was always presented monaurally at 60 dB SPL to the right ear and the masker at the ipsilateral ear was set to 50, 60, or 70 dB SPL. The ILDs were constructed in the same manner as for experiment 1. The range of ILDs was from −20 to 40 dB, although the range was smaller for each ipsilateral T/M due to the constraint that the masker in the contralateral ear would take on the same set of values (30–70 dB SPL) regardless of masker level at the target ear. In the third experimental condition, T_mM_τ, masker ITD was investigated. A target was presented to the right ear at 60 dB SPL while the masker was presented to both ears with no ILD (the masker was presented to the two ears at 50, 60, or 70 dB for any given trial). ITDs were created in the same manner as for experiment 1. Again, the ITD values used were 600, 300, 0, −300, and −600 μs. The levels for each condition were presented randomly within a block of 50 trials. Blocks were generally repeated until a minimum of 100 trials had been obtained at every combination of levels for each listener.

B. Results

The condition in which target and masker were both monaural, T_mM_m, was the same as that reported in Kidd *et al.* (2005), but additional data were collected in order to ensure that improvements were not due to learning effects. A comparison of the first half of the data with the second showed no significant effects. The data were combined to improve the estimates of baseline performance. This led to a greater number of trials for the monaural conditions (∼400 points per listener). For each masker level in each condition, proportion correct was calculated for each listener by dividing the number of trials on which a listener was correct (correctly reported both the color and number of the target sentence) by the total number of trials at that masker level for that condition. ANOVAs were conducted on these "proportion correct" data.[1] The data presented as T_mM_m are the same as those reported in Kidd *et al.* (2005). The individual data for all three listeners for the ILD condition (T_mM_α) are plotted in the successive panels of Fig. 3. Each listener occupies one row of panels and the columns of panels indicate the level of the ipsilateral masker (50, 60, or 70 dB). The solid horizontal line in each panel indicates performance in the baseline condition T_mM_m. By comparing the solid lines across rows in Fig. 3 it can be seen that decreases in T/M in the target ear are associated with decreases in performance, despite the nonoverlapping target/masker bands. The solid lines are included in this and later figures as a monaural masker baseline reference: Any points that are higher than this baseline indicate release from masking due to the presence of a contralateral masker.

The first dichotic condition, T_mM_α, involved a monaural target sentence and a binaural masker sentence. Figure 3 summarizes the effect of changing ILD, the level of which is shown in decibels (dB) along the abscissa for each panel. A
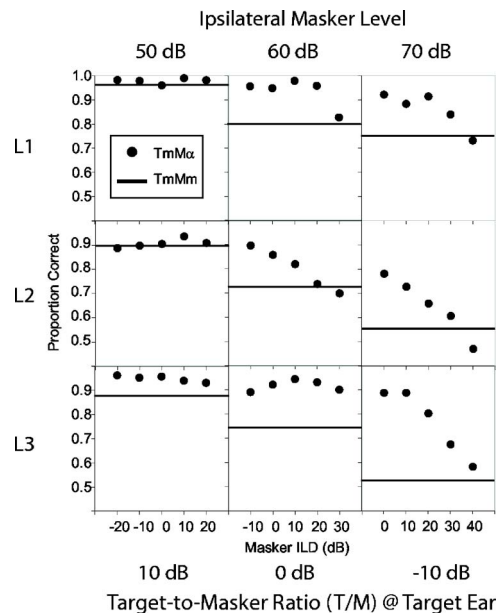
FIG. 3. Proportion correct in the masker ILD condition of experiment 2. Each row of panels contains the results for a single listener and each column contains those for a given T/M. The solid lines represent performance when the masker was presented only at the target ear and the symbols represent performance when the same masking sentence was also presented at the contralateral ear. The range of ILDs varies for the three T/M values because ILD was dependent upon the relationship between the masker levels at the two ears and the range of masker levels at the contralateral ear was always 30–70 dB (see the text).

two-way repeated-measures ANOVA found significant main effects of both ipsilateral masker level [$F_{(2,4)}=19.10$, $p<0.01$] and contralateral masker level [$F_{(4,8)}=24.27$, $p<0.01$]. The interaction was also significant [$F_{(8,16)}=5.50$, $p<0.01$]. These results show that for a monaural target, it is the ILD that matters, and not just the T/M in the target ear. As can be seen in Fig. 3, both increasing ipsilateral masker level and decreasing contralateral masker level (expressed as increasing ILD) were associated with decreasing performance. The interaction occurred because higher ipsilateral masker levels caused increasing ILD to have a larger impact on performance. Correlations were calculated between the performance values obtained with a 70 dB SPL ipsilateral masker and the mean ratings associated with ILD from experiment 1. The correlations were negative, indicating that for increasing ILDs, ratings increased and performance decreased. The values were −0.753, −0.946, and −0.962 for the listeners L1, L2, and L3. The variation in mean ratings accounted for 79% of the variance in performance (R^2 averaged across the three listeners).

The second dichotic condition, T_mM_τ, explored the effect of masker ITD in the presence of a monaural target. The individual data for all three listeners are plotted in the successive panels of Fig. 4. As in Fig. 3, each listener occupies one row of panels and the columns indicate the level of the dichotic masker (50, 60, or 70 dB in both ears). Again, the solid line in each panel indicates the effect of a monaural masker at that level and points that fall above the line indicate an improvement relative to baseline. The symbols in this case indicate what occurs when the masker is binaural, with

Gallun *et al.*: Binaural release from informational masking
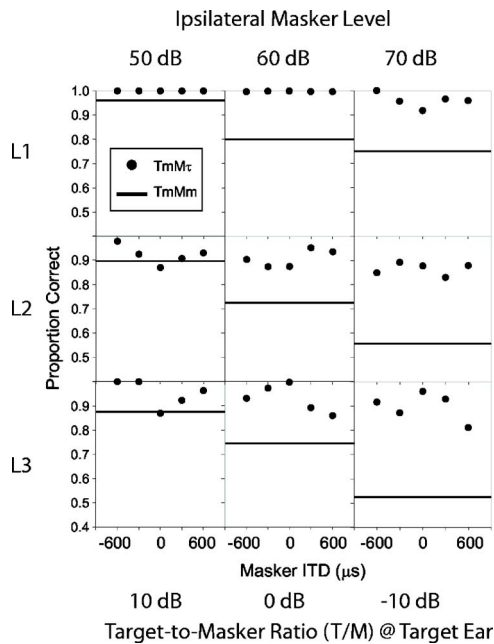
FIG. 4. Proportion correct for the masker ITD condition in experiment 2. Each row of panels contains the results for a single listener and each column contains those for a given T/M. Solid lines indicate performance when only the masker at the target ear was presented and symbols when a masker was presented at both ears at the same level. The ITD values indicate the interaural delay (in microseconds) for the masker. Positive ITDs indicate that the ipsilateral masker led.

the ITD values plotted along the abscissa. Each panel in Fig. 4 indicates that there was substantial release from masking for all of the ITD values used in the study and that, unlike with ILD, changes in masker ITD had essentially no systematic effect. For all three listeners at all three masker levels, the function is somewhat irregular and essentially independent of ITD. Although increasing ITD led to a shift in perceived masker location towards that of the monaural target (see Fig. 2), and the effect of ITD on perceived lateralization was significant in experiment 1, identification performance was largely unaffected. A two-way repeated-measures ANOVA conducted on T/M at the target ear and ITD found a reliable effect of T/M [$F_{(2,4)}=61.93$, $p<0.01$], but no effect of ITD [$F_{(4,8)}=0.78$, $p=0.569$]. Unlike T_mM_α, ceiling performance at a masker level of 50 dB SPL did not cause an interaction because performance was essentially independent of ITD at all levels. Another difference from T_mM_α is that correlations between performance and mean lateral position rating from experiment 1 were very small, accounting for only 4% of the variance in performance for the 70 dB SPL masker (average R^2 across listeners). To summarize, these results indicate that *all* ITDs produced a release from masking and that shifts toward and away from the location of the target (as shown by the results of experiment 1) were not enough to reliably change the amount of release.

By combining the results of the trading ratio analysis (described in Sec. III C) with the results from experiment 2, it can be seen that the amount of release from masking is dependent largely on the lateral position of the masker. Since the lateral positions associated with the ITD manipulations were restricted to values between −20 and 20 dB, it is not surprising that all the amount of release was quite large across all ITD conditions—release was also quite large for ILDs between −20 and 20 dB.

C. Masking release in dB

To compare our result more directly with those of other researchers, the averaged performance data from experiment 2 were used to plot three-point psychometric functions for each ITD and for a subset of the ILDs (0, 10, and 20 dB). These functions relate performance in the various conditions to the T/M at the target ear. The difficulty is that the slopes of the psychometric functions were much steeper for the monaural masking condition than for the binaural conditions. Unequal slopes result in T/M differences that depend heavily on which performance value is used for the comparison. The performance value associated with 2-down, 1-up adaptive tracking (70.7%) results in extremely large estimates of masking release (20–50 dB for the ILDs and 50–120 dB for the ITDs), suggesting that 70.7% would be an inappropriate value. A more reasonable value is 90%, because it requires little or no extrapolation of the functions—each data set contains a value near 90% correct. In order for a function to be included in this analysis, the correlation between predicted and actual performance values was required to exceed 0.85. "Release from masking" expressed in dB was calculated by subtracting the T/M that was associated with 90% performance in the monaural masker condition from the T/M associated with 90% correct in each binaural condition. When this analysis was performed for the ILD functions, 5.2 dB of release was estimated for the 0 dB ILD, 9.6 dB for the 10 dB ILD, and 10.9 for the 20 dB ILD. The other ILD functions (−10, −20, 30, and 40 dB ILDs) had fewer than three points and so accurate psychometric functions could not be estimated. For the ITD functions, the 300 μs ITD function only correlated 0.79 with the actual data and the 0 μs ITD function was nonmonotonic and so only correlated 0.10. Consequently, neither was included. The remaining functions (600, −300, and −600 μs ITDs) resulted in release estimates of 16.1, 21.7, and 24.4 dB, respectively. While large, these are the most conservative estimates possible given this data set.

D. Summary

Experiment 2 found that: (1) with monaural targets, the introduction of binaural maskers resulted in substantial release from masking relative to the monaural maskers (note the large number of points that fall above the horizontal lines in Figs. 3 and 4); (2) the correlations between ratings of lateral position and performance accounted for a large percentage of the variation in performance for the ILD manipulation (when the masker was at 70 dB SPL) but almost none of the variance for the ITD manipulation.

There are two conflicting explanations for these data. The first is that differences in lateral position *do* increase the ability of listeners to distinguish between target and masker bands, but that the lateral position associated with the monaural target was substantially different from the positions associated with all of the ITD values and the majority of the ILD values. This is supported by the finding that all of the

lateral position ratings obtained for the ITDs in experiment 1 were statistically different from the monaural ratings. This was also true for many of the ILD values and the few that had similar location ratings to the monaural stimulus accounted for the only reduced performance found in experiment 2.

The second explanation is that differences in lateral position are not useful *per se*, but rather that the ILD values that caused reductions in performance resulted in percepts that were not perceived as binaural. This argument asserts that the majority of the maskers were perceived as binaural and that the very salient difference between a binaural masker and a monaural target was responsible for the improved performance at all of the ITD and the majority of the ILD values. Experiment 3 was designed to gather evidence on the specific question of whether a binaural/monaural difference is required to obtain the high levels of performance seen in experiment 2.

V. EXPERIMENT 3: IDENTIFICATION, TARGET DIOTIC

It is possible that the results from the second experiment depend heavily on the fact that the target was monaural and the maskers binaural. In this experiment, the target and masker were both binaural, allowing a test of whether differences in T/M, ITD, and ILD are all effective at causing a release from masking. This situation is much more similar to real-world listening conditions in which all targets and maskers are usually binaural. Experiment 3 serves as a test of the hypothesis that while a perceptible monaural/binaural difference is sufficient to produce maximal release from masking, it may not be necessary

A. Methods

The target and masker sentences were drawn from the same pool of processed sentences as were used in experiment 2. The task and response method were also identical. The primary difference in experiment 3 was that the target was no longer presented monaurally. In the first condition, T_0M_α, the target was presented diotically (same level at both ears with no interaural delay) at 60 dB and the masker was presented to both ears with an average level of 50, 60, or 70 dB across the two ears. There was no ITD on target or masker in the ILD condition, T_0M_α. Three ILDs were created at each nominal masker level: 0, 10, and 20 dB. For the 0 dB ILD, the masker sentence was at the same level in both ears. When a 10 or 20 dB ILD was imposed, however, the masker level at the left ear was raised and the level at the right ear was lowered. Thus, for an average masker level of 60 dB and a 20 dB ILD, the left ear would receive the masker sentence at 70 dB and the right ear would receive the same sentence at 50 dB. Hence the T/M was different at the two ears. All three average masker levels and all three ILDs were presented in random order on every block of ILD trials. Each listener participated in an average of 100 trials at each combination of average masker level and ILD.

For experiment 3, the target was presented to both ears, so ILDs had to be constructed in order to minimize the mismatch between the T/M at the two ears. To produce average

T/M's of -10, 0, and 10 dB and ILDs of 0, 10, and 20 dB, on successive trials the masker at the left ear was presented at levels ranging between 45 and 70 dB (in 5 dB steps) and the masker at the right ear at levels between 55 and 80 dB, also in 5 dB steps. Thus, an average T/M of -10 dB and an ILD of 20 dB was created by presenting the left masker at 60 dB and the right masker at 80 dB. This also created a subset of conditions that allow ILDs to be examined with T/M held constant at one ear.

For the second condition, T_0M_τ, varying ITD, the target was presented as described earlier and the masker was presented to both ears with no ILD (both maskers were presented at 50, 60, or 70 dB SPL for any given trial), hence the same T/M of -10, 0, or 10 dB was present at both ears. The difference between the maskers was created by adding a delay of 300 or 600 μs to the beginning of the masker presented to the right ear. This meant that the masker sentence in the left ear started playing before the one in the right, producing both an onset and an ongoing ITD (as well as an offset ITD). The ITD values tested were 600, 300, and 0 μs relative to the target sentence. All three masker levels and all three ITDs were presented in random order on every block of ITD trials. Each listener participated in an average of 100 trials at each combination of masker level and ITD.

The results of experiment 1 suggested that maskers with these ILDs and ITDs were experienced at a variety of interaural locations and that maskers with ILDs of 10 and 20 dB were rated in a manner statistically different from those with 0 dB ILDs. Ratings of maskers with 600 μs ITDs were statistically different from ratings of those with 0 μs ITDs, but those with 300 μs ITDs were not. This difference seems to have been due to the increased variability of the ratings of masker with ITDs relative to those with ILDs and could perhaps have been overcome with the collection of more data.

B. Results

The masking data will be described as in the second experiment. "Proportion correct" refers to the proportion of the total trials at any one combination of parameters for which both the color and number of the target sentence were correctly reported by a given listener. Figure 5, the results of the T_0M_α condition, shows the effect on performance of introducing a difference in level in the speech masker presented to the two ears when the target sentence was presented diotically. The individual data for all three listeners are plotted in successive panels along each row and the columns of panels indicate the level of the masker. Masker ILD is plotted along the abscissa of each panel, with the 0 dB value corresponding to the case in which there was no interaural difference in masker level (T_0M_0). Performance in T_0M_0 (indicated by the dashed line as well as the leftmost point in each panel) serves as a baseline for the other two ILD values. It also serves as a comparison with T_mM_m (indicated by the solid line in each panel), both of which, it seems reasonable to assume, would lead to identical performance. As can be seen, this was generally the case. The exceptions are discussed in Sec. V C. The question that motivated this portion of the experiment was whether or not masker ILD
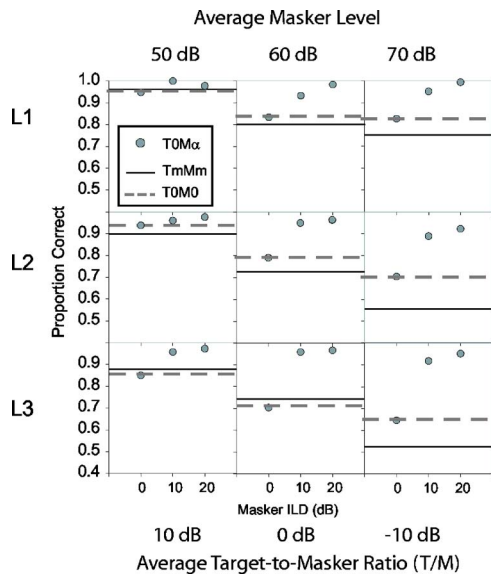
FIG. 5. Proportion correct as a function of masker ILD for the diotic targets in experiment 3. Each row of panels contains the results for a single listener and each column contains those for a given T/M. The solid lines show performance when both target and masker were monaural. The symbol at a masker ILD of 0 (as well as the dashed line) indicates that both target and masker were diotic. The remaining symbols indicate the result of lowering the level at one ear and raising it at the other to produce ILDs of 10 or 20 dB.

FIG. 6. Proportion correct for the ITD condition in experiment 3. Each row of panels contains the results for a single listener and each column contains those for a given T/M. The solid lines show performance when both target and masker were monaural. The symbol at a masker ITD of 0 (as well as the dashed line) indicates that both target and masker were diotic. The remaining symbols indicate the result of delaying the onset of the masker at one ear in order to produce onset and ongoing masker ITDs of 300 or 600 μs.

supports a release from masking when the target is also binaural. Focusing again on the condition where performance is least influenced by ceiling effects (average T/M of -10 dB) shows that average proportion correct performance in T_0M_0 was 0.73. Creating a 10 dB ILD in the masker increased performance to 0.92 and a 20 dB ILD increased it to 0.95. A repeated measures two-way ANOVA confirmed the effects of T/M [$F_{(2,4)}=15.14$, $p<0.05$] and ILD ($F_{(2,4)}=19.32$, $p<0.01$). The interaction of the two factors was also reliable [$F_{(4,8)}=22.66$, $p<0.01$]. As in experiment 2, the interaction was due to the near ceiling performance found at large T/M values. Correlations between lateral position ratings and performance for the -10 dB T/M data were 0.998, 0.994, and 0.836 for L1, L2, and L3, respectively. Thus lateral position can account for 90% of the variance in performance (average R^2 across listeners).

Figure 6 shows the performance values obtained for each of the three listeners in the T_0M_τ condition, in which the target was diotic and various ITDs were imposed upon the masker. Each listener's data are confined to a single row and the columns contain all of the data associated with a given masker level (50, 60, or 70 dB). As with the 0 dB masker ILD, the 0 μs masker ITD, T_0M_0 (indicated by the dashed line as well as the leftmost point in each panel) serves as a baseline for the other two ITD values. A comparison between T_mM_m (indicated by the solid line in each panel) and T_0M_0 shows that, as in the ILD conditions, there are a number of panels in which diotic performance is better than monaural, primarily for the T/M of -10 dB (see the discussion in Sec. V C for more on this point). As with ILD, there was substantial release associated with the imposition of masker ITD, indicated by points that fall above the horizontal lines.

In the majority of panels, the improvement for a 300 or 600 μs ITD is obvious relative to the 0 μs condition. Those not showing improvement typically have near ceiling performance in T_0M_0. A repeated-measures two-way ANOVA confirmed the effect of T/M [$F_{(2,4)}=15.81$, $p<0.05$] and ITD [$F_{(2,4)}=9.53$, $p<0.05$]. The interaction of the two factors was not reliable in this condition [$F_{(4,8)}=2.06$, $p=0.178$]. Correlations calculated between lateral position ratings and performance in the -10 dB T/M condition were 0.671, 0.654, and 0.954 for L1, L2, and L3, respectively, thus accounting for 60% of the variance in performance (average R^2 across listeners).

C. Masking release in dB

Using the same method as in experiment 2, an estimate of the amount of masking release in dB was obtained by fitting curves to the psychometric functions for the T_0M_0 conditions and then subtracting the T/M associated with 90% correct on those functions from the T/M associated with 90% correct on the functions associated with the dichotic conditions. For these functions, the correlations were required to exceed 0.85. Only the correlation for the 300 μs ITD function, which was nonmonotonic, failed this test. The amount of release estimated from this difference was 27.7 dB for the 10 dB ILD and 73.5 dB for the 20 dB ILD condition. The estimate of release associated with the 600 μs ITD condition was 21.6 dB. Clearly, these estimates are made unreliable by the large differences in the slope of the functions associated with diotic and dichotic maskers. Nonetheless, it is also the case that these values are no less than the values found by Arbogast et al. (2002) and they give an approximate idea of the size of the effects expressed in dB. For a

more precise assessment, performance should be held constant while masker level is varied or, equivalently, complete psychometric functions should be obtained for all conditions.

D. Better-ear advantage

In comparing binaural to monaural masking, it is important to consider the "better-ear advantage" (Bronkhorst and Plomp, 1988). If a binaural masking situation is such that the T/M at one ear becomes better than it was in the monaural case, improvements in performance could be a simple result of this improvement in T/M rather than a result of binaural processing or a difference in perceived location. For experiment 2, this was not an issue because the T/M at the target ear was independent of binaural manipulations. For the ILD condition in experiment 3, however, the actual T/M at one of the ears was equal to the diotic T/M for that condition plus half of the ILD imposed. Thus, a 20 dB ILD created both an interaural difference and a change in T/M at one ear by 10 dB relative to the T/M in that ear for T_0M_0. In order to help separate the effect of binaural interactions from the effects of an improved T/M at the better ear, there are several relevant performance values that can be compared. If the T/M at the better ear was the sole determining factor, performance would be unaffected by changing the T/M at the contralateral ear. The average performance across listeners for the case where ILD was 0 dB and T/M at both ears was 0 dB was 0.78. When the better ear T/M was kept at 0 dB and the worse ear was lowered to −20 dB, performance increased to 0.95. This improvement of 0.17 cannot be due to a better-ear advantage because the T/M at the better ear remained the same. It cannot be due to a binaural effect based on a time difference because the ITD was 0 μs in both cases. Consequently, it is concluded that, at least for the stimuli used in this experiment, an ILD alone is capable of improving performance substantially and the effect is not mediated by the T/M at the better ear.

E. Summary

The performance obtained when target and masker were diotic was compared to that obtained when masker ILD and ITD were systematically varied. Substantial improvement was found for all listeners at nearly every combination of masker level and interaural difference. In general, less improvement was obtained when the masker level was 50 dB, probably due to ceiling effects. These results suggest that the release from masking found in experiment 2 was not due entirely to the monaural/binaural comparison between target and masker. Rather, it seems that interaural differences *can* support substantial release from IM. As with the first experiment, there was little difference between the release observed for a 10 and a 20 dB ILD or a 300 and a 600 μs ITD. This is in accordance with the trading ratios obtained in Sec. III C, which suggested that these listeners found these values of ILDs and ITDs to be roughly equivalent in terms of the lateral positions they produce.

VI. DISCUSSION

A. Lateral position and release from masking

The primary question driving these experiments was whether or not listeners can use differences in the perceived lateralization of target and masker to improve their performance on a speech identification task. The use of spectrally nonoverlapping stimuli allowed this question to be addressed in a configuration that substantially reduced the amount of energetic masking. This was important because, as discussed by Durlach and Colburn (1978), if the target and masker fall in the same auditory filters then performance that correlates with differences in perceived location is also the performance predicted by nonlocation-based binaural mechanisms. Despite the lack of EM, the amount of binaural release in this study was anticipated to be high based on similar work by Arbogast *et al.* (2002). That study showed that stimuli processed in this way resulted in substantial release from masking when target and masker each were presented from spatially separated loudspeakers compared to both from the same loudspeaker. In that study the stimuli were presented in a mildly reverberant chamber rather than over headphones, meaning that the ITD and ILD information occurred together naturally and the effect of better-ear T/M could not be ascertained. This study had the advantage that it could examine the effects of ITD and ILD in isolation while controlling for better-ear advantage.

The first experiment showed that ILD and ITD manipulations with these stimuli were sufficient to produce a wide range of lateral position ratings. Statistical tests showed that there were significant effects of ILD and ITD on the lateralization ratings obtained from the listeners. A trading ratio between ITD and ILD was estimated for these stimuli and the average value was 38.6 μs/dB. This suggests that the largest ITD values presented (±600 μs) were equivalent to ILD values of only 20 dB.

The results of experiment 2 showed that listener performance was related to whether or not the masker had a binaural configuration that had been reliably rated as different from the monaural target. The correlations between lateral position ratings and performance were high for the ILD conditions in experiment 2, but not for the ITD conditions (all of which supported good performance). The ambiguity of the relationship between release and perceived location raised the possibility that release might be based on the difference between a monaural target and a binaural masker. The results of the third experiment did not support this interpretation, however. When the target was diotic, the introduction of ILDs and ITDs still produced release from masking and the correlations between performance and lateral position accounted for at least 60% of the variance across all listeners and both manipulations.

B. Mechanisms based on perceived location

These results are concordant with those of Carhart *et al.* (1969), who showed that masker ITDs of 800 μs produced release from speech-on-speech masking with wideband stimuli. In that study, four wideband maskers, two of speech and two of modulated noise, were presented at opposing time

delays with a diotic target. The introduction of these time delays produced release from masking similar to that obtained when all of the maskers had the same (parallel) time delays. This similarity of release poses problems for any mechanism that relies upon a single compensatory time delay for each frequency channel. Since a single delay would only compensate for one of the two sets of maskers, the authors argue that such a mechanism cannot completely explain the results. Instead, they suggest that release from masking based on signal location acts in parallel with binaural unmasking based on interaural compensation.

There is conflict between our results and those of Culling and Summerfield (1995), however. The stimuli in that study were four noise bursts with narrow bandwidths and different center frequencies that listeners were taught to identify based on their similarity to vowel spectra. If only two were presented, a recognizable vowel was perceived, but if all four were presented at the same level, vowel identification was at chance. Introducing a 390 μs ITD into one pair did not improve performance, but presenting one pair monaurally did cause improvement. This result is in conflict with those of experiment 3 in which a 300 μs ITD was effective in changing the amount of masking obtained relative to the T_0M_0 condition.

Darwin and Hukin (1999) proposed a solution to this apparent contradiction, based on a difference they had observed between segregating one time-varying signal from another (target-word identification) and segregating the steady portions of a complex stimulus (segregating a harmonic from a vowel). While the identification task showed an effect of ITD, the task of segregating a harmonic did not. They suggested that the harmonic was grouped with the other vowel components on the basis of grouping factors such as harmonicity and common-onset and that the ITD of the grouped object was assigned later by a weighted averaging of ITD across its frequency components. The target word, on the other hand, was grouped as a separate object from the remainder of the sentence and thus a difference in ITD between the target and the sentence was able to serve as a cue for tracking one object among others.

The arguments of Darwin and Hukin (1999) could be applied to this study if we assume that listeners used interaural differences to focus attention on the target instead of the masker. This would imply that both sentences were available for processing (consistent with reports of typical error patterns; e.g., Brungart and Simpson, 2002; Arbogast et al., 2002), but that the listeners used interaural differences to track the information in the target and ignore the masker. Evidence that a successful grouping of the frequency components is possible when the target and masker are both presented monaurally is supported by the fact that listeners performed the task at above-chance levels in the T_mM_m condition. Cues that listeners could have been using monaurally included correlations between component envelopes (of which common-onset is an example) as well as "goodness-of-fit" tests on various band combinations, perhaps based on the fact that the target bands were constrained to produce the callsign "Baron" as the second word of the sentence. Once the bands have been successfully grouped, the main diffi-

culty facing the listener is to track the proper sentence. For the monaural case, the only cue that identifies the target sentence is the set of bands that were associated with the word "Baron." Once the callsigns have been uttered, the listener must combat uncertainty about which sentence to track by focusing attention only on the target bands. When there is an interaural difference between target and masker, however, an additional cue is available for tracking the target sentence, an operation that Darwin and Hukin (1999) describe as selective attention.

C. Mechanisms based on cancellation

An alternative argument is that performance is improved by a binaural cancellation operation conceptually similar to the equalization-cancellation (E-C) mechanism (Durlach, 1963; also discussion of "Listener-Min" in Durlach et al., 2003) that acts on the masker bands only, leaving the target bands unaffected. According to this explanation, initial grouping still occurs on the basis of monaural cues as described earlier, but once the masker bands are identified they are cancelled and thus removed. For the stimuli in experiment 2, all masker combinations were binaural and all targets were monaural, so if all the binaural bands could be equalized and cancelled the resulting stimulus would—ideally—contain only the target. While this usage of the E-C model is different from how it has been traditionally considered, it is in accordance with the basic "nulling" analogy from radar operations that was referenced from the earliest discussions of the model (e.g., Durlach, 1963). In most cases, the E-C model has been used to explain how binaural manipulations could lead to reductions in the effectiveness of a masker overlapping the frequencies of the target, a situation in which the sizes of differences between the interaural parameters of the target and masker correlate with the reduction in effectiveness. For these stimuli, neither the predictions nor the results follow this pattern. As long as the binaural masker bands can be successfully identified (and the mechanism can select the equalization parameters for just the masker bands), the masker may be "cancelled." This could explain why changes in the interaural parameters had so little effect on the amount of release obtained.

The only case where there seemed to be a relationship between ILD and binaural release was when the target was monaural and the ILD was so great that the ratings of lateral position were not significantly different from monaural. In this case, the system was unable to generate a release from masking. This limit is actually in agreement with data on release from masking with broadband binaural noise maskers and monaural tonal targets (reviewed in Durlach and Colburn, 1978). In that case, it has been found that although release does occur based on the noise ILD, the amount of release declines with increasing ILD. The limit seems to be similar to that seen in experiment 2—roughly 30 dB. If, as this suggests, the level-equalization simply has a limited range, then it is perhaps the case that the masker can only be completely equalized when the difference between the ears falls within that range. In fact, Breebaart et al. (2001) have modeled these data and suggest that, with noise maskers,

large ILDs lead to a decorrelation at the two ears. The mechanism is based on peripheral nonlinearities and the suggestion is that the level adjustment alone is no longer sufficient to match the noise at the two ears.[2] If such a mechanism were to operate for stimuli such as those employed in this study, this would allow an E-C approach to explain the ILD data from experiment 2 as well.

Recent work by Akeroyd (2004) examined the central question of whether or not the band parameters can be selected independently, allowing (in our case) cancellation of only the masker. The stimuli were multiple-component signals in wideband noise that was interaurally 180° out of phase from each signal component. In the baseline condition, the noise had no interaural-phase difference (IPD) applied to it (and thus ITD was constant across frequency), while the signal components were each 180° out of phase at the two ears. Applying a constant equalization parameter to these signals would result in optimal cancellation of the noise. The test conditions involved noise that also had a constant IPD, still 180° out of phase from the signal components. In those cases, a single equalization parameter would not permit optimal cancellation of the noise. Nonetheless, the test conditions had thresholds indistinguishable from baseline performance, supporting a version of the E-C model in which the equalization and cancellation process selects equalization parameters independently in each band. This is also the conclusion reached by Edmonds and Culling (2005).

D. The diotic advantage

One aspect of the data that is not captured by either an E-C mechanism or an attentional explanation is the consistent finding that presenting target and masker to both ears results in better performance than presenting them to only one ear. Improvement due to diotic presentation is not the usual pattern of results, but it does occur (Langhans and Kohlrausch, 1992). Several of the factors that tend to result in a diotic improvement were present in this study: a cross-frequency analysis was required and the target had limited spectral overlap with the masker. It is possible that the two ears allow two independent samples to be taken of the stimuli, thus reducing the variability of the estimated signal parameter (as suggested by Zwicker and Henning, 1985 and by Schooneveldt and Moore, 1989). Additionally, as suggested by Langhans and Kohlrausch, interactions between the left and right hearing pathways could be responsible for an increased sensitivity in binaural but not monaural conditions. Another of their suggestions, that the headphones were not actually equivalent, was eliminated by a series of acoustical measurements. None of this explains why a difference should occur primarily for the −10 dB T/M (see Figs. 5 and 6).

E. Better-ear listening

These conditions allowed us to test the hypothesis, initially proposed by Bronkhorst and Plomp (1988), that the effects of interaural level differences can be completely described as better-ear listening. While this describes their data quite well (and that of Hawley *et al.*, 2004), if the only effect of ILD is to change the T/M at the better ear, then for the ILD conditions in experiment 2 of this study there should have been no effect of ILD. This is because T/M at the better ear was held constant as ILD was manipulated. As is clear in Fig. 3, however, ILDs clearly lead to improvements despite the constant T/M at the "better" ear. Further support comes from Shinn-Cunningham *et al.* (2005), who showed that, with stimuli very similar to those used in this study, the improvements in performance obtained by separating the source locations of target and masker contained a large component related to differences in perceived location. In addition, these differences could be obtained by introducing variations in just ILD or just ITD. In fact, combining ILD and ITD information produced no greater release from masking.

The conclusion from the Shinn-Cunningham *et al.* (2005) study as well as from this one is that, for IM, the traditional models of spatial unmasking may encourage researchers to focus on the wrong mechanisms. Since there is relatively little masker energy in the target bands for these stimuli, the decorrelation caused by introducing a target and the effect of improved T/M at one ear may be fairly unimportant compared to traditional binaural conditions in which target and masker occupy the same frequency regions. This parallels the conclusion of Carhart *et al.* (1969) that binaural mechanisms that ignore perceived location may be missing an important part of the explanation for why interaural differences improve performance. It is important to also consider the differences in perceived location that serve to enhance the segregation of two very similar sentences, one of which is the target and the other of which is the masker.

F. Summary

This study found that for a monaural target masker ILDs and ITDs that favored the target ear were just as effective as cues that favored the opposite ear—so long as the lateral positions were rated as significantly different. This is in accord with the data that directly inspired this study, in which a binaural interaction with no long-term correlation between the ears still supported substantial release from IM. The results of experiment 2 can be summarized by the statement that when a sufficiently intense masker is perceived as occupying a lateral position different from that of a monaural stimulus, listeners somehow exploit the pairing of a monaural target with a binaural masker to overcome IM. The third experiment showed that the result is not entirely dependent on distinguishing a binaural masker from a monaural target, however. Substantial release due to the presence of ILDs and ITDs was obtained when the target and masker were both binaural. Two mechanisms were proposed to explain the data. According to the first mechanism, listeners use the perceptual differences arising from the binaural/monaural comparison and/or the differences in perceived location to maintain their attention on the target sentence and thus report the correct words. In the second mechanism, listeners actively suppress the activity caused by the masker by relying on an equalization-cancellation mechanism similar to that suggested by Durlach (1963).

[1]It has been argued that a rationalized arcsine transform should be applied to proportion correct data prior to statistical analyses in order to normalize the data (see Studebaker, 1985). For this experiment and for the data in experiment 3, this transformation was performed and the statistical analyses were repeated. There were no differences in the results. In particular, the differences between conditions at T/M values of 10 dB, where performance approached ceiling values, were still not statistically reliable. As the results did not vary from those for the untransformed data, proportion correct is reported throughout for ease of interpretation.

[2]The applicability of the Breebaart *et al.* (2001) model to our data was suggested by Dr. Armin Kohlrausch, to whom we are grateful.

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

Arbogast, T. L., Mason, C. R. and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Bolia, R. S., Nelson, W. T., Ericson, M. A. and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Breebaart, J., van der Par, S. and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. I. Model structure." J. Acoust. Soc. Am. **110**, 1074–1088.

Bronkhorst, A. W. and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.

Brungart, D. S. and Simpson, B. D. (**2002**). "The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal," J. Acoust. Soc. Am. **112**, 664–676.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd Jr., G. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Carhart, R., Tillman, T. W., and Greetis, E. (**1969**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Colburn, H. S. and Durlach, N. I. (**1978**). "Models of binaural interaction" in *Hearing*, Handbook of Perception Vol. IV, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Culling, J. F. and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds: Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.

Darwin, C. J. and Hukin, R. W. (**1999**). "Auditory objects of attention: The role of interaural time differences," J. Exp. Psychol.: Hum. Percept. Perform. (HPP) **25**, 617–629.

Darwin, C. J. and Hukin, R. W. (**2004**). "Limits to the role of a common fundamental frequency in the fusion of two sounds with different spatial cues," J. Acoust. Soc. Am. **116**, 502–506.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Durlach, N. I. and Colburn, H. S. (**1978**). "Binaural phenomena" in *Hearing, Handbook of Perception* Vol. IV, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr., (**2003**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Edmonds, B. A. and Culling, J. F. (**2005**). "The spatial unmasking of speech: Evidence for within-channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.

Hafter, E. R. and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.

Hafter, E. R. and Kimball, P. (**1980**). "The threshold for binaural interaction," J. Acoust. Soc. Am. **67**, 1823–1825.

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (**2004**). "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**, 833–843.

Kidd Jr., G., Mason, C. R., and Gallun, F. J. (**2005**). "Combining informational and energetic masking for speech identification," J. Acoust. Soc. Am. (in press).

Langhans, A. and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.

Sayers, B. McA. and Toole, F. E. (**1964**). "Acoustic-image lateralization judgments with binaural transients," J. Acoust. Soc. Am. **36**, 1199–1205.

Schooneveldt, G. P. and Moore, B. C. J. (**1989**). "Comodulation masking release for various monaural and binaural combinations of the signal, on-frequency, and flanking bands," J. Acoust. Soc. Am. **85**, 262–272.

Shinn-Cunningham, B. G., Ihelfeld, A., Satyavarta, S., and Larson, E. (**2005**). "Bottom-up and top-down influences on masking," Acta. Acust. Acust. (submitted).

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Zwicker, E. and Henning, G. B. (**1985**). "The four factors leading to binaural masking-level differences," Hear. Res. **19**, 29–47.

Measures of extents of laterality for high-frequency "transposed" stimuli under conditions of binaural interference

Leslie R. Bernstein and Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology),
University of Connecticut Health Center, Farmington, Connecticut 06030

Our purpose in this study was to determine whether across-frequency binaural interference would occur if ITD-based extents of laterality were measured using high-frequency transposed stimuli as targets. The results of an earlier study [L. R. Bernstein and C. Trahiotis, J. Acoust. Soc. Am. **116**, 3062–3069 (2004)], which focused on threshold-ITDs, rather than extents of laterality, suggested that high-frequency transposed stimuli might be "immune" to binaural interference effects resulting from the addition of a spectrally remote, low-frequency interferer. In contrast to the earlier findings, the data from this study indicate that high-frequency transposed targets are susceptible to binaural interference. Nevertheless, high-frequency transposed targets, even when presented along with an interferer, yielded greater extents of ITD-based laterality than did high-frequency Gaussian noise targets presented in isolation. That is, the "enhanced potency" of ITDs conveyed by transposed stimuli persisted, even in the presence of a low-frequency interferer. Predictions made using an extension of the model of Heller and Trahiotis [L. M. Heller and C. Trahiotis, J. Acoust. Soc. Am. **99**, 3632–3637 (1996)] accounted well for across-frequency binaural interference obtained with conventional Gaussian noise targets but, in all but one case, overpredicted the amounts of interference found with the transposed targets. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1984827]

I. INTRODUCTION

Several experiments concerning binaural hearing have demonstrated that sensitivity to an ongoing interaural temporal disparity (ITD) conveyed by a high-frequency "target" can be degraded by presenting a simultaneously gated, diotic, low-frequency "interferer." This general outcome is termed binaural interference in order to differentiate it from other deleterious effects, such as masking. Binaural interference has been demonstrated both in experiments concerning the detection of ITDs and in experiments concerning extents of the laterality produced by ITDs. (e.g., McFadden and Pasanen, 1976; Davis, 1985; Zurek, 1985; Dye, 1990; Trahiotis and Bernstein, 1990; Bernstein, 1991; Buell and Hafter, 1991; Woods and Colburn, 1992; Buell and Trahiotis, 1993; Stellmack and Dye, 1993; Bernstein and Trahiotis, 1992; 1995; Heller and Trahiotis, 1995a, 1996; Bernstein and Trahiotis, 2004).

One explanation offered to account for binaural interference is that it stems from a "nonoptimal" combination of ITD information conveyed by the target and the interferer, respectively (e.g., Buell and Hafter, 1991; Heller and Trahiotis, 1995a, 1996; Bernstein and Trahiotis, 2004). Within that view, the ITD value of zero carried by the low-frequency diotic interferer is combined with the ITD carried by the high-frequency target. As a consequence, the "internal" effective magnitude of the ITD conveyed by the target is diminished, as compared to what it would be in the absence of the interferer. For that reason, ITDs imposed on the high-frequency target must be increased: (1) in order to reach threshold, and (2) in order to produce an extent of laterality comparable to that obtained in the absence of the interferer.

Bernstein and Trahiotis (2004) have recently shown that across-frequency binaural interference effects as measured via ITD thresholds were, for all practical purposes, absent when certain "transposed" stimuli served as high-frequency targets. Transposed stimuli are specially constructed in order to provide high-frequency auditory channels with envelope-based temporal information that mimics waveform-based temporal information normally available only in low-frequency channels (see van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002; 2003; 2004).

Our purpose in this study was to determine to what degree binaural interference would occur if extents of laterality, rather than the resolution of ITD, were measured using high-frequency transposed stimuli as targets. It has long been recognized, using conventional stimuli, that one cannot predict extents of laterality solely from measurements of threshold ITDs made with the same stimuli (e.g., Stern and Colburn, 1978; Bernstein and Trahiotis, 1985; Trahiotis *et al.*, 2001). Therefore, it seemed important to determine whether interference effects would be revealed when extents of laterality, a type of the "suprathreshold" measure of the potency of ITD, were the focus.

II. EXPERIMENT

A. Procedure

Two sets of four high-frequency stimuli served as targets and each set of four will be described and justified in turn. The first set of four consisted of (1) a 200-Hz-wide band of Gaussian noise centered at 4 kHz; (2) a 128-Hz tone transposed to 4 kHz; (3) a 50-Hz-wide band of Gaussian noise
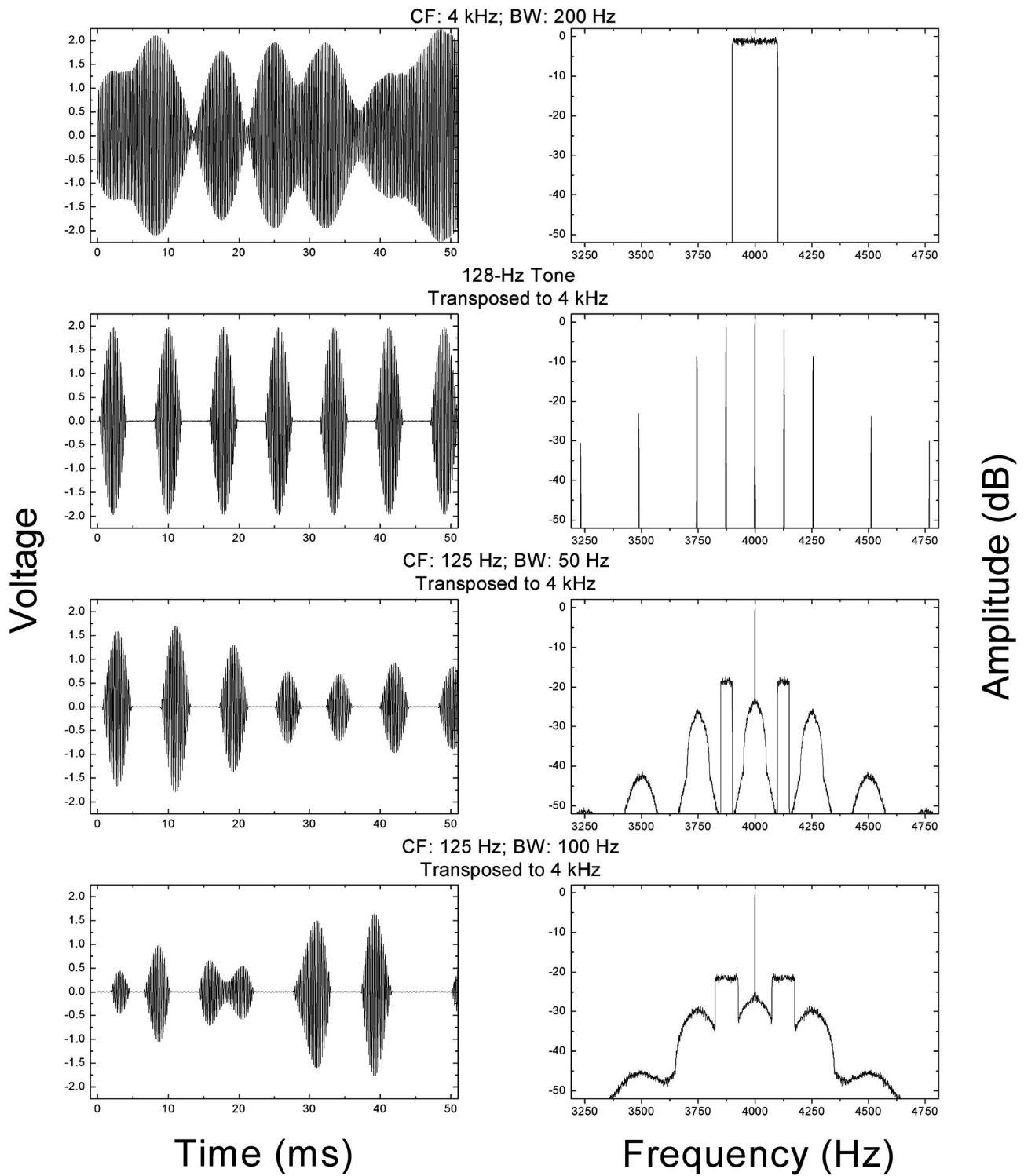
FIG. 1. Left-hand panels: Randomly chosen 50-ms epochs of the four types of targets employed in the main experiment. Right-hand panels: The long-term power spectra of each of the four types of targets.

centered at 125 Hz transposed to 4 kHz; (4) a 100-Hz-wide band of Gaussian noise centered at 125 Hz transposed to 4 kHz.[1]

Examples of each of the four types of high-frequency target stimuli composing this first set are depicted in Fig. 1. The left-hand panels show randomly chosen 50-ms epochs of each type of target; the right-hand panels show their (long-term) power spectra calculated over tens of seconds. Begin-

ning with the left-hand panels, note that six or seven envelope maxima occur for each type of target during the 50-ms epochs. That is so because the number of envelope-maxima per second for each type of target is about 125. This corresponds to the rate of the low-frequency stimuli that were transposed to 4 kHz and is the expected number of envelope maxima per second for a 200-Hz-wide band of Gaussian noise (see Rice, 1954).

The primary reason for choosing a rate of envelope fluctuation near 125 Hz was that transposed stimuli having a similar rate of envelope fluctuation have been shown to produce the greatest enhancements (relative to their conventional counterparts) of ITD processing at high frequencies (e.g., van de Par and Kohlrausch, 1997; Bernstein and Trahiotis, 2002; 2003). Therefore, using such a rate would help to maximize observed differences in the extents of laterality between the transposed and conventional stimuli. Furthermore, using such a rate causes more than 99% of the energy of the transposed stimuli to fall between the sidebands centered at 3750 Hz and 4250 Hz (see the right-hand panels of Fig. 1). As a result, only minimal amounts of energy associated with each transposed target would fall outside the approximately 500-Hz-wide auditory filter centered at 4 kHz (see Moore, 1997). It was judged that having all stimuli satisfy this latter constraint would enhance the validity of comparisons among the data obtained with the various targets and, perhaps, simplify interpretations of the results.

The second set of four targets consisted of (1) a 400-Hz-wide band of Gaussian noise centered at 4 kHz; (2) a 256-Hz tone transposed to 4 kHz; (3) a 50-Hz-wide band of Gaussian noise centered at 250 Hz transposed to 4 kHz; (4) a 100-Hz-wide band of Gaussian noise centered at 250 Hz transposed to 4 kHz. Note that the rates of envelope fluctuation of these four targets are twice those of their counterparts in the first set of targets. Despite this difference, 90% of the energy associated with each of the transposed targets in the second set also fell within the sidebands centered at 3750 and 4250 Hz.

The major reason for employing the second set of targets was that increasing the rate of fluctuation of transposed noises from 125 to 250 Hz has been shown to result in *somewhat smaller* extents of laterality (Bernstein and Trahiotis, 2003). In contrast, for conventional bands of Gaussian noise, increasing the rate of envelope fluctuation from about 125 Hz to about 250 Hz by increasing bandwidth from 200 to 400 Hz has been shown to result in *somewhat greater* extents of laterality across values of ITD (e.g., Bernstein and Trahiotis, 2003). Mindful of these results, it seemed prudent to assess the degree to which amounts of interference measured with the second set of stimuli would be commensurate with those measured with the first set of stimuli. Of special interest was whether transposed noises in the second set, which have been shown to yield relatively less enhancement of ITD processing in terms of the extent of laterality, would be relatively more susceptible to across-frequency binaural interference. Posed differently, would the "less-robust" processing of ITD supported by the transposed noises in the second set of targets result in more binaural interference?

The transposed stimuli were generated in the manner described by Bernstein and Trahiotis (2002; 2003; 2004). Briefly, the time-domain representation of a narrow band of low-frequency Gaussian noise or a low-frequency tone was (linearly) half-wave rectified. This was accomplished by setting all negative values to zero. Then, the rectified waveforms were transformed to the frequency domain and all spectral components above 2 kHz were removed by setting their magnitudes to zero [see pp. 1027–1028 of Bernstein

and Trahiotis (2002) for a justification of the choice of this type of rectification and filtering]. Following this, each rectified and filtered waveform was transformed back to the time domain and multiplied by a 4-kHz sinusoidal "carrier." This procedure resulted in transposed stimuli, each having an envelope with a time signature that mimicked that of the rectified and filtered low-frequency narrow band of noise (or low-frequency tone) used to generate it.

The stimulus that served as the interferer was a diotic 400-Hz-wide band of Gaussian noise centered at 500 Hz. That stimulus has been shown to produce substantial amounts of binaural interference, as indexed by threshold ITDs, for conventional stimuli centered at 4 kHz (e.g., Bernstein and Trahiotis, 1995; 2004).

All stimuli were generated digitally as seconds-long buffers using a sampling rate of 20 kHz (TDT AP2), were low-pass filtered at 8.5 kHz (TDT FLT2), and were presented via Etymotic ER-2 insert earphones. Targets and interferers were each presented at a level matching 72 dB SPL, as produced by TDH-39 earphones in a 6-cm^3 coupler.[2] For all stimuli that served as targets, ongoing ITDs (0, 200, 400, 600 μs, left ear leading) were imposed by applying linear phase-shifts to the representation of the signals in the frequency domain and then gating the signals destined for the left and right ears coincidentally after transformation to the time domain.

Four normal-hearing young adults (one male and three female) who had each participated in previous binaural experiments served as listeners. Extents of laterality were measured via an acoustic pointing task in which the listeners varied the interaural intensitive difference (IID) of a 200-Hz-wide band of noise centered at 500 Hz (the pointer) so that it matched the intracranial position of a second, experimenter-controlled, stimulus (the target). This procedure has been used previously in several studies (e.g., Trahiotis and Stern, 1989; Buell *et al.*, 1991; Heller and Trahiotis, 1996) and is described fully in Bernstein and Trahiotis (1985). The overall level of the pointer, when presented diotically (IID=0 dB), was 65 dB SPL. Listeners adjusted the intracranial position of the pointer by rotating a knob. Rotation of the knob produced symmetric changes of the IID (in dB) of the pointer (i.e., increases in level at one ear and decreases in level at the other ear). The IID adjusted by the listener served as a metric of the intracranial position of the target. A value of IID was chosen randomly from the range±4 dB (in 0.5 dB steps) and was inserted in the pointer prior to each match. This served to randomize the initial position of the pointer with respect to the absolute position of the knob. Each sequence of stimuli consisted of three presentations of the target (each separated by 50 ms), a pause of 300 ms, three presentations of the pointer (each separated by 200 ms) and a pause of 650 ms. The duration of target and pointer stimuli was 100 ms, including 20 ms cos^2 rise/decay, and each was extracted from the appropriate longer buffer.

Targets and pointers were repeated until the listeners indicated that they had matched the intracranial positions of the target and pointer. Prior to completing a match, listeners

had the option of halting, and then restarting, the sequence in order to "check" their adjustments after a period of silence.

Testing began by measuring extents of laterality in the absence of the interferer. All of the aforementioned targets were visited in pseudorandom order. Having chosen a particular stimulus condition as the target, a random process was used to select a value of ITD from the set to be tested until the listeners had completed three independent matches for each value of ITD. The mean IID inserted by the listener to match the diotic targets (ITD=0) served as a "correction factor." That is, it was subtracted from the IIDs resulting from all the matches in the run. Any series of matches that yielded a correction factor with a magnitude greater than 5 dB was discarded and rerun until a "valid" series was obtained with a correction factor less than that value. After matches were made in all of the "without-interferer" conditions, the same stimulus conditions were re-visited in reverse order. Finally, the entire process was repeated for targets presented in the presence of the simultaneously-gated low-frequency interferer. The data reported in the figures represent the mean "corrected" value of IID of the pointer across the six valid matches made by each listener for a particular combination of target and ITD.

B. Results and discussion

In all of the plots of the data that follow (Figs. 2–4), filled squares represent data obtained in the absence of the interferer and open triangles represent data obtained in the presence of the interferer. Only the data averaged across the four listeners will be presented because between-listener differences were small, as indicated by the error bars that represent ± one standard error of the mean. Table I contains information regarding the within-listener precision with which the matches were made. The average standard deviation was computed for each value of ITD tested. This was done separately for the no-interference and interference conditions. Each entry in the table represents the mean of the 32 (eight targets × four listeners) measures of standard deviation computed across the six matches made by each listener for each experimental condition. Note that the average, within-listener standard deviation increases with ITD, being only slightly larger than 1 dB when the targets were presented diotically and increasing to about 2.5±0.25 dB when the ITD was 600 μs. This indicates that the precision of binaural processing is greater for stimuli perceived near midline than for stimuli that are perceived to be lateralized well toward the ear. This is in accord with the findings of previous studies (e.g., Mills, 1958; Hafter and DeMaio, 1975; Domnitz and Colburn, 1977). The standard errors of the means indicated in the figures, taken together with the standard deviations appearing in Table I, indicate that both the within- and between-listener variance was small in comparison to, and need not constrain interpretations of, the magnitude of the differences found between conditions.

The solid lines within the figures are the best linear-regression fits to the data with the intercept forced to be zero. An intercept of zero is consistent with our use of "correction factors" for matches made with diotic targets. Using the cor-
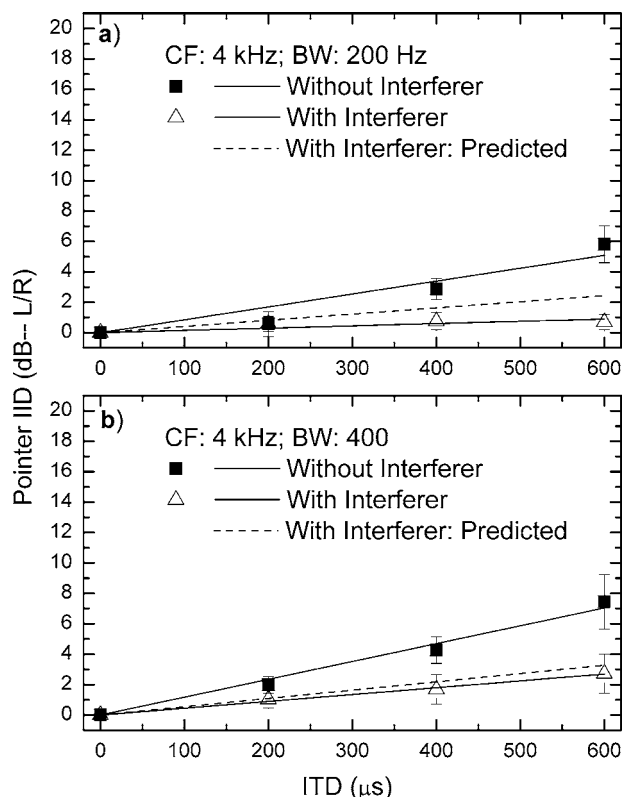
FIG. 2. IID of the pointer (in dB) required to match the intracranial position of the target as a function of the ongoing ITD (left ear leading) imposed on a 200-Hz-wide [panel (a)] or 400 Hz wide [panel (b)] band of Gaussian noise. The data points represent the mean values computed across the four listeners. The error bars represent ±1 standard error of the mean. Squares represent data obtained when the target was presented in isolation; triangles represent data obtained when the target was presented along with the low-frequency interferer. Solid lines represent the best linear-regression fits to the data with the intercept forced to be zero. Dashed lines represent quantitative predictions of the extents of laterality in the presence of the interferer. Predictions were derived from an extension of the Heller and Trahiotis (1996) model described in Sec. III C of the text.

rection factors forces the empirical match to an ITD of 0 μs to be 0 dB of IID of the pointer. The dashed lines within the figures represent quantitative predictions of extents of laterality in the presence of the interferer. Those predictions are discussed in Sec. II C.

A summary of the statistical analyses that will be used to help interpret the data is presented in Table II. The first column of the table indicates the type of target. The second column of the table contains the slopes of the best-fitting lines to the data for each target when it was presented in the absence of the interferer. The third column indicates the amount of variance accounted for by the fitted line forced through the origin.[3] The fourth and fifth columns contain the same types of information for the conditions in which the interferer was presented along with the target. The sixth column contains the ratio of the "without-interferer" slope to the "with-interferer" slope for each target. Because, as will be discussed, the data are well-fit by straight lines passing through the origin, the ratio of the value of the pointer-IID obtained in the without-interferer condition to the value obtained in the with-interferer condition can be taken to be constant for all values of ITD of the target. Therefore, the
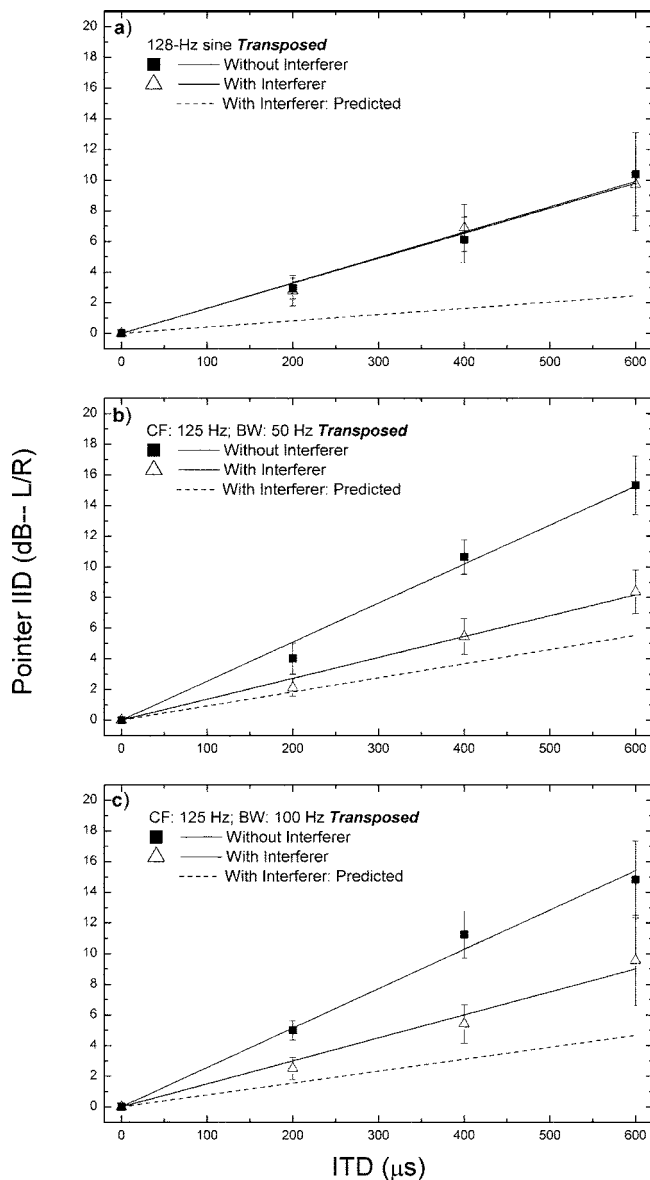
FIG. 3. Similar to Fig. 2, but for the "125 Hz" transposed stimuli.
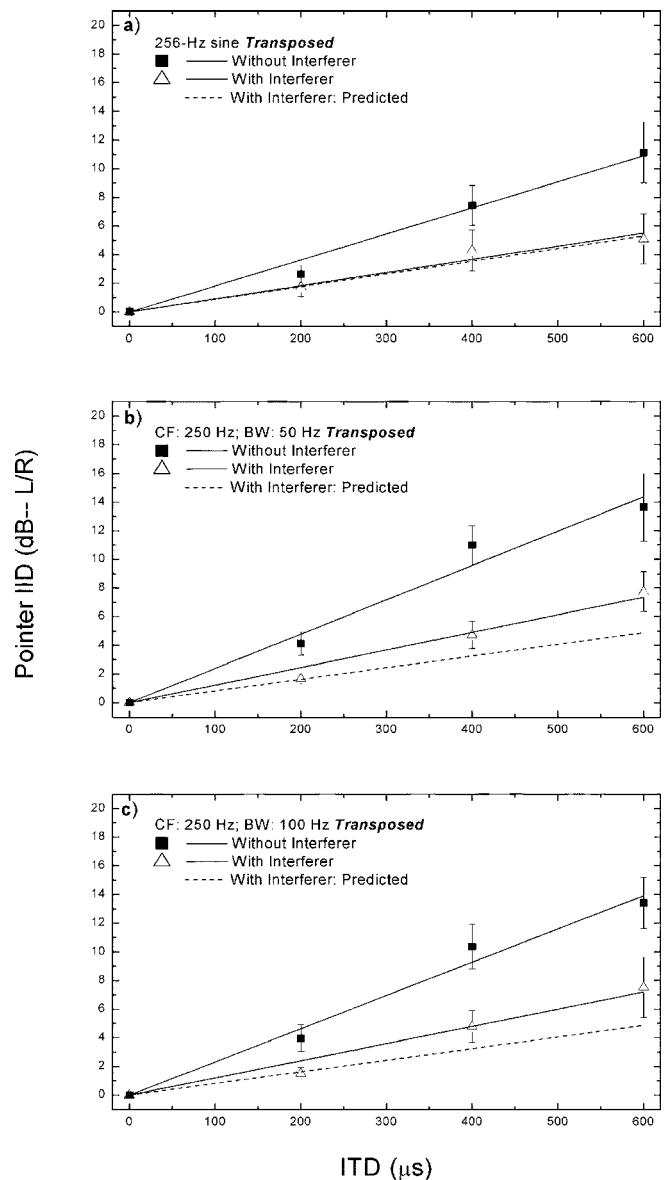
FIG. 4. The same as Fig. 3, but for the "250 Hz" transposed stimuli.

ratio of the slopes provides a consistent measure of binaural interference independent of the perceived intracranial loci of the targets. Bernstein and Trahiotis (1995) used similar types of ratios to quantify binaural interference in an experiment measuring threshold ITD, threshold IID, and thresholds for the detection of a signal in noise.

The seventh and eighth columns show the F-ratios and their associated probabilities derived from a statistical analysis (Edwards, 1964) that yields the significance of the differences of the regression slopes (calculated using the mean data) obtained for the targets presented in isolation versus when they were presented along with the interferer. This type of statistical analysis was deemed appropriate because it is the amount of the disparity between the two slopes that directly (i.e., independently of the magnitude of ITD) reflects the amount of binaural interference that occurred.

Asterisks indicate significance using an alpha level of $p < 0.05$. The percentages of variance accounted for by the linear fits (columns 3 and 5), being, in the great majority of

cases between 97% and 99%, indicate that the data are well fit by straight lines having an intercept of zero. Given the quality of the linear fits, the ratio of the slopes obtained for "without-interference" and for "with-interference" conditions can be taken as valid descriptors of the amounts of interference produced for each target.

TABLE I. The average standard deviation (in dB) obtained for each ITD in the "no-interference" and "interference" conditions. Each entry in the table represents the mean of the 32 (eight targets × four listeners) measures of standard deviation computed across the six matches made by each listener for each experimental condition.

ITD	No Interference	Interference
0	1.31	1.19
200	1.94	1.54
400	2.66	2.06
600	2.78	2.25

1630   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Bernstein and Trahiotis: Binaural interference and transposed stimuli

TABLE II. The second through fifth columns report the slopes (dB/μs) and variance accounted for by the best-fitting lines to the data obtained in each experimental condition. The sixth column contains the ratios of the "without-interferer" to "with interferer" slopes. The two rightmost columns of the table show the results of the statistical analysis (Edwards, 1964) yielding the significance of the differences of the regression slopes obtained when the targets were presented in isolation versus when they were presented along with the interferer. Asterisks indicate significance using an alpha level of $p < 0.05$.

Target	Without-interferer slope	% var. acc.	With-interferer slope	% var. acc.	Slope ratio	F	p
Gaussian- CF: 4 kHz; BW: 200 Hz	0.0085	90.1	0.0015	60.5	5.65	19.43	0.0116*
Gaussian- CF: 4 kHz; BW: 400 Hz	0.0117	98.4	0.0045	99.3	2.61	62.06	0.0014*
Transposed 128 Hz tone	0.0165	99.0	0.0163	99.3	1.01	0.13	0.7367
Transposed CF: 125 Hz; BW: 50 Hz	0.0255	99.0	0.0136	99.0	1.87	40.07	0.0032*
Transposed CF: 125 Hz; BW: 100 Hz	0.0257	99.0	0.0150	98.2	1.71	18.85	0.0122*
Transposed 256 Hz tone	0.0182	98.5	0.0092	96.6	1.98	30.15	0.0054*
Transposed CF: 250 Hz; BW: 50 Hz	0.0239	97.4	0.0123	97.7	1.95	12.97	0.0227*
Transposed CF: 250 Hz; BW: 100 Hz	0.0232	98.3	0.0120	97.8	1.93	17.10	0.0144*

Panels (a) and (b) of Fig. 2 display the results obtained when 200-Hz-wide and 400-Hz-wide bands of Gaussian noise, respectively, served as targets. Note that only modest extents of laterality were obtained when the high-frequency Gaussian-noise targets were presented in isolation (filled squares), the largest values of the IID of the pointer being only 6–7.5 dB when the ITD was as large as 600 μs. Such values of IID are substantially less than the 10–12 dB or so that have been commonly found to indicate intracranial images lateralized near the leading ear (see Watson and Mittler, 1965; Yost, 1981; Trahiotis and Bernstein, 1986). Comparison of the data in panel (b) to those displayed in panel (a) reveals that slightly greater extents of laterality were obtained when the bandwidth of the target was increased from 200 to 400 Hz.

Most important for our purposes, for both of the Gaussian-noise targets, extents of laterality were diminished with the addition of the simultaneously gated low-frequency noise (open triangles), thereby indicating binaural interference. The degree of across-frequency binaural interference is quantified by the ratios of the without-interferer to with-interferer slopes that can be found in the fifth column of Table II. For the data obtained with the two Gaussian noise targets, the ratios of the slopes were 5.65 (200-Hz-wide noise) and 2.61 (400-Hz-wide noise). As indicated by the rightmost column of the table, the underlying pairs of slopes for these two targets were significantly different. Note that the slope ratios for these Gaussian-noise targets are the largest in the table. This indicates that the largest interference effects occurred with these two conventional Gaussian-noise targets. In fact, the relatively small values of the IID of the pointer required to make matches in the presence of the interferer for those two targets indicates intracranial positions of the targets quite close to or at the midline. The only seemingly poor linear fit to the data occurred when the 200-Hz-wide Gaussian-noise target was presented along with the interferer. As indicated in Table II, the linear fit to these data accounted for only 60% of the variance. This outcome should be interpreted with caution for two reasons. First, as shown in Fig. 2(a), in the presence of the interferer, there was virtually no ITD-based lateralization and, therefore, virtually no variance above and beyond the noise of measurement for which to account. Second, the best fitting line deviates, at most, by about 0.5 dB from the data. Overall, the

extents of laterality and the amounts of interference obtained with the Gaussian-noise targets are entirely consistent with and essentially replicate patterns of data reported in our earlier studies (Bernstein and Trahiotis, 1995; 2003; Heller and Trahiotis, 1996).

We now turn to the data of most interest, those obtained with high-frequency, transposed stimuli. Panel (a) of Fig. 3 displays data obtained when the target was a 128-Hz pure tone transposed to 4 kHz. Two effects are apparent. First, ITDs conveyed by the transposed target presented in isolation resulted in larger extents of laterality as compared to the Gaussian-noise targets when they were presented in isolation. For example, an IID of the pointer of 10 dB was required to match the intracranial position of the target when the ITD was 600 μs. This indicates an intracranial image near the leading ear and is typical of the extent of laterality produced by a 600-μs ITD conveyed by a *low-frequency* stimulus. Second, the addition of the interferer had no effect. This is confirmed quantitatively in Table II in that the ratio of the without-interferer to with-interferer slopes is 1.01 for this target. Furthermore, as also indicated in the table, the 128-Hz transposed target is the *only* one for which the difference between the two underlying slopes is not significant.

This set of observations, taken together, supports the conclusion that adding the diotic low-frequency interferer had *no effect* on ITD-based lateralization for this target. A similar apparent "immunity" to across-frequency binaural interference was found with the same stimulus when threshold ITD, rather than the extent of laterality, was the dependent variable (Bernstein and Trahiotis, 2004).

Panels (b) and (c) of Fig. 3 display data obtained when the target was either a 50-Hz-wide or a 100-Hz-wide band of noise centered at 125 Hz and transposed to 4 kHz. Note that these two transposed-noise targets consistently produced larger extents of laterality than did the transposed tone [panel (a)] in the absence of the interferer (filled squares). For example, for the transposed noises, the listeners required about 11 dB of IID in the pointer, indicating an image just at the leading ear when the ITD was only 400 μs. Furthermore, an IID of about 15 dB was required to match an image when the ITD was 600 μs. These extents of laterality are very similar to those reported with the same stimuli by Bernstein and Trahiotis (2003). At this time, we cannot account for the finding that the transposed noises produced slightly greater

extents of laterality than did the transposed tone. Perhaps this outcome is related to an earlier finding with conventional stimuli, namely that high-frequency Gaussian noises produce greater extents of ITD-based laterality than do high-frequency SAM tones having the same "bandwidth" (Trahiotis and Bernstein, 1986).

Unlike what was found with the transposed tone [panel (a)], the addition of the low-frequency, diotic noise to the transposed *noises* resulted in interference, in that extents of laterality were reduced. As indicated in Table II, the ratios of the lateralization slopes for the two bandwidths were 1.87 and 1.71, respectively, indicating substantial (and statistically significant) binaural interference.

Interestingly, as can be seen by visual inspection of Fig. 3 and by referring to the lateralization slopes in Table II, the extents of laterality measured in the presence of the interferer when the targets were transposed noises were very much like those measured with the 128-Hz transposed-tonal target. In all four cases, the slopes are near 0.015. Thus, the increments in lateralization measured when the ITDs were conveyed by the transposed noises, rather than by a transposed tone, were essentially "neutralized" in the presence of the interferer. That is, in the presence of the interferer, the extents of laterality measured with the transposed tone and the two transposed noises are virtually identical. Note that, even under interference conditions, these transposed stimuli yielded greater (i.e., enhanced) extents of laterality than did conventional high-frequency Gaussian-noise targets *presented in isolation* [compare Fig. 3(b) with Figs. 2(a) and 2(b)].

The data obtained when the low-frequency stimuli to be transposed were centered near 250 Hz are displayed in Fig. 4. Once again, panel (a) contains data obtained for targets generated by transposing a pure tone, in this case 256 Hz, to 4 kHz. Panels (b) and (c) contain data obtained when 50-Hz-wide [panel (b)] or 100-Hz-wide [panel (c)] Gaussian noise centered at 250 Hz and transposed to 4 kHz served as targets.

The comparisons to be made among the data displayed in the panels of Fig. 4 directly parallel those made among the data presented in Fig. 3. The reader is reminded that the primary reason for employing this second set of transposed stimuli was that low-frequency noises centered on 250 Hz and transposed to 4 kHz have been shown to yield smaller extents of laterality than low-frequency noises centered on 125 Hz (Bernstein and Trahiotis, 2003). The question of interest here was whether such "less-robust" processing of ITD would result in more binaural interference.

For ease of exposition, we begin with the data in panels (b) and (c) of Fig. 4. The data show that these transposed targets, when presented in isolation, result in extents of laterality that are substantially *greater* than those obtained with conventional Gaussian-noise targets (compare to Fig. 2) but only slightly *smaller* than extents of laterality measured with their counterparts derived from bands of noises centered on 125 [compare to panels (b) and (c) of Fig. 3]. These differences are consistent with those reported by Bernstein and Trahiotis (2003). The *amounts of interference*, as quantified by the ratio of the slopes obtained with these targets (1.95 and 1.93, respectively), were slightly greater than those ob-

tained with transposed noise targets derived from noises centered at 125 Hz (1.87 and 1.71, respectively).

Finally, we turn to the data obtained with a 256-Hz pure tone transposed to 4 kHz [Fig. 4, panel (a)]. Note that the extents of laterality measured in the absence of the interferer are virtually the same as those obtained with the 128-Hz pure tone transposed to 4 kHz when it was also presented alone [see panel (a) of Fig. 3]. As seen by considering their respective lateralization slopes of 0.0165 and 0.0182, those two stimuli, when presented in isolation, were lateralized essentially identically. Nevertheless, for the 256-Hz pure tone transposed target, the addition of the low-frequency diotic noise produced *substantial* interference, as evidenced by the ratio of the without-interferer to with-interferer slope-ratio of 1.98. This clearly indicates that this target, unlike its 128 Hz transposed counterpart [Fig. 3(a)] was, indeed, "susceptible" to across-frequency binaural interference.

The comparisons among the data discussed above do not appear to yield insights regarding (1) which features of transposed targets affect their ITD-based extents of laterality and (2) which features of the transposed targets render them more or less susceptible to across-frequency binaural interference. For example, as can be seen in Fig. 1, the difference between a transposed tone and a transposed noise is manifest as both a change from a deterministic to a stochastic envelope and a change from a discrete "line spectrum" to one having sidebands characterized by spectral extent or "breadth." In future experiments, we plan to investigate whether and to what degree various temporal and/or spectral aspects of high-frequency complex waveforms affect their susceptibility to across-frequency binaural interference.

All of these issues notwithstanding, the main findings of the study are quite clear: high-frequency transposed stimuli not only produce greater ITD-based extents of laterality than do high-frequency Gaussian noises, but they are, generally speaking, less susceptible to across-frequency binaural interference.

C. Extending and testing a laterality-based quantitative model of binaural interference

In this section, we describe the application to the data described above of an extension of the quantitative model of across-frequency binaural interference described by Heller and Trahiotis (1996). To date, the model has been found to account quantitatively for binaural interference measured via threshold-ITDs when the targets and interferers were sinusoidally amplitude-modulated tones (Heller and Trahiotis, 1996) and when the targets and interferers were narrow bands of Gaussian noise (Bernstein and Trahiotis, 1994). A fundamental underlying assumption of that model is that binaural interference arises from an obligatory, nonoptimal, across-frequency-channel, combination of binaural cues. The combining of information is nonoptimal because the low-frequency diotic interferers, which convey no information useful for detecting or lateralizing on the basis of ITDs within high-frequency targets, are not ignored. A second important assumption of the model is that it is *the extent of laterality* produced at threshold-ITD that is the crucial variable, and not the magnitude of the threshold ITD itself. The

extent of laterality at threshold (in terms of IID of the pointer) is estimated by multiplying the threshold ITD (in μs) by the lateralization slope (in IID of the pointer/μs). That extent of laterality is, in turn, used to compute the salience or "weight" of the ITD-based information carried by the target or interferer when each is presented in isolation [see pp. 3066–3068 of Bernstein and Trahiotis (2004) for a more formal and much more detailed discussion]. Computing the weights via empirical measures obtained with targets and interferers presented in isolation allows the model to *predict* the expected amounts of interference when targets and interferers are presented simultaneously. This approach should be differentiated from others in the general literature that compute "weights" derived in a post-hoc manner that are then used as *descriptors* of the relative potencies of the elements in compound stimuli.

In order to compute the predicted lateralization slope for each target in the presence of the interferer, the Heller and Trahiotis (1996) model had to be extended. This was done by assuming that

$$effs_{t+i} = \frac{a * effs_t + b * effs_i}{a + b}, \qquad (1)$$

where a and b represent the weights applied to the ITD-based information conveyed by the target and interferer, respectively. The term $effs_{t+i}$ represents the *effective* lateralization slope of the target in the presence of the interferer, $effs_t$ and $effs_i$ represent the *effective* lateralization slopes for the target and interferer, respectively. The effective lateralization slope for the target, $effs_t$, is simply the lateralization slope measured when the target is presented in isolation. The effective lateralization slope for the interferer, $effs_i$, however, is *not* the lateralization slope measured when the extent of laterality of the interferer is measured in isolation because, in the experimental conditions, the interferer carried an ITD of zero *regardless of the value of the ITD of the target*.

The *effective* slope of either the target or the interferer is obtained by multiplying that particular stimulus' lateralization slope (i.e., the one obtained in isolation) by the ratio relating the ITD actually imposed on that particular stimulus to the ITD of the target. In our case, the ITD actually imposed on the target was, of course, the ITD of the target itself and, therefore, that ratio is one. This means, as stated above, that the effective slope of the target, $effs_t$, is the same as its lateralization slope measured in isolation. Formally,

$$effs_t = s_t * 1, \qquad (2)$$

where s_t is the lateralization slope of the target measured in isolation.

Now, in the interference conditions for which predictions are to be made, the ITD actually imposed on the interferer was always zero. Therefore, the ratio relating the ITD actually imposed on the interferer to the ITD imposed on the target was always zero and, therefore, the effective lateralization slope of the interferer was zero. Formally,

$$effs_i = s_i * 0, \qquad (3)$$

where s_i is the lateralization slope of the interferer measured in isolation.

It should be understood that this form of the model takes into account the fact that the targets had ITDs ranging up to 600 μs while the interferers always carried an ITD of zero, independent of the ITD imposed on the target. On an intuitive level, the model can be seen as attempting to capture the combination of the positive lateralization slope of the target with the effective "zero" lateralization slope of the diotic interferer.

Because only lateralization slopes, and not threshold ITDs, were measured in the current study, the weights (a and b) were derived from measures obtained in an earlier study that employed stimuli that were essentially identical to those employed here. In the earlier study, both threshold ITDs and lateralization slopes were measured (Bernstein and Trahiotis, 2004). Three of the four listeners in the current study also participated in that earlier study. The lateralization slopes obtained in the prior and current study for common stimulus conditions were functionally identical. In order to arrive at a prediction for the lateralization slopes measured with targets in the presence of the interferer, the weights, a and b, were entered into Eq. (1) along with the effective lateralization slopes estimated, as described above.

Inspection of Fig. 2 reveals that the predictions of the model (dashed lines) are fairly accurate (within less than 1 dB or so of the obtained data) when the targets were conventional high-frequency bands of Gaussian noise. The model slightly, but consistently, *underpredicts* effects resulting from the addition of the interferer. That is, the model predicts slightly greater extents of laterality in the presence of the interferer than were obtained.

In contrast, inspection of Fig. 3 reveals that, for the "125 Hz" transposed targets, the model consistently and, in many instances substantially, *overpredicts* the effects of adding the interferer. Especially noteworthy is the failure of the model for the data depicted in panel a, when the target was a 128-Hz tone transposed to 4 kHz. In this condition, no interference was observed and the model predicted as much as 5–7 dB of interference depending on the ITD imposed was 400 or 600 μs.

The finding that the model accounts fairly well for interference effects obtained with conventional high-frequency targets, but overpredicts the degradation produced by the interferer for high-frequency *transposed* stimuli, is consistent with the results of our earlier study, which focused on threshold ITDs (Bernstein and Trahiotis, 2004). Inspection of Fig. 4 reveals that the model accurately predicts the effects of adding the interferer to the 256 tone transposed to 4 kHz and moderately overpredicts the effects of adding the interferer to the two transposed targets derived from Gaussian bands of noise centered at 250 Hz. Overall, the success and failure of predictions of across-frequency interference effects via the Heller and Trahiotis (1996) model in this study parallel those of Bernstein and Trahiotis (2004).

III. GENERAL DISCUSSION AND SUMMARY

As a follow-up to Bernstein and Trahiotis (2004), the purpose of this study was to determine whether across-frequency binaural interference would occur if ITD-based extents of laterality were measured using high-frequency transposed stimuli as targets. In the earlier study, which focused on threshold ITDs rather than the extents of laterality, no binaural interference was observed when the targets were high-frequency transposed stimuli. That outcome suggested that transposed high-frequency stimuli might, somehow, be "immune" to across-frequency interference effects commonly found with conventional high-frequency targets.

The data from the current study indicate that high-frequency transposed targets can, indeed, be susceptible to binaural interference when extents of laterality, rather than threshold ITDs, are measured. That is, it appears that whether one observes binaural interference with high-frequency transposed targets depends upon whether threshold ITDs or ITD-based extents of laterality are measured. It should be emphasized, however, that even when interference occurred, high-frequency transposed targets yielded greater extents of ITD-based laterality than did Gaussian noise targets presented in isolation. That is, enhancements of ITD-based lateralization obtained with high-frequency transposed stimuli, as compared to their conventional counterparts presented in isolation, were found even under conditions of across-frequency binaural interference. Predictions made using an extension of the model of Heller and Trahiotis (1996) accounted quite well for binaural interference obtained with conventional Gaussian noise targets but, in all but one case, overpredicted the amounts of interference found with the transposed targets.

One broad implication of the findings in this study, taken together with those of Bernstein and Trahiotis (2004), is that across-frequency binaural interference effects cannot, in general, be accounted for solely by assessing the relative potency and/or salience of ITDs conveyed by targets and interferers when each is presented in isolation. It appears that the types of "blackbox" models that have been used to account for across-frequency binaural interference with conventional stimuli (e.g., Buell and Hafter, 1991; Heller and Trahiotis, 1995a, 1996) will have to be augmented or superceded by a more sophisticated, and probably more mechanistic, approach. The success of such models will depend upon their ability to capture how the stimuli, as processed, determine the "strength" of across-frequency interactions that appear to occur within the binaural system. Based on previous studies, (e.g., Trahiotis and Bernstein, 1990; Heller and Trahiotis, 1995a) across-frequency binaural interference appears, in the main, to be central, rather than peripheral, in origin.

ACKNOWLEDGMENTS

[1]The frequency of the tone that was transposed to 4 kHz was chosen to be 128 Hz in order to match the frequency that was employed in our earlier studies (Bernstein and Trahiotis, 2002; 2004). The center frequency of the low-frequency bands of noise that were transposed to 4 kHz was chosen to be 125 Hz in order to match the center frequency employed in our earlier study of extents of laterality using transposed noises. The same considerations dictated the choice of 256 and 250 Hz for the transposed tone and noises, respectively, which composed the second set of targets described later in the text. These very small disparities within each set of targets between the frequencies of the tones and the center frequencies of the noises that were transposed to 4 kHz are, for all practical purposes, inconsequential.

[2]Stimulus levels produced according to the calibration supplied with the Etymotic ER-2 earphones sounded less loud than stimuli presented at nominally the same level via TDH-39 earphones, according to their calibration. Dr. Mead Killion, of Etymotic Research, validated our listening experience and agreed with us that the two respective methods of calibration would be expected to produce levels of stimulation differing by about 10 dB. We chose to "calibrate" the outputs of the Etymotic earphones to the nominal levels produced by the TDH-39s so that listeners in this study would receive levels of stimulation directly comparable to those utilized by us and others in prior psychophysical experiments employing TDH-39s. We verified that the levels from the Etymotic earphones were appropriate by presenting a high-frequency, stimulus to one ear via an Etymotic ER-2 earphone and simultaneously to the other ear via a TDH-39 earphone. We then adjusted the relative levels between the two ears to produce a "centered" intracranial image, as is produced by diotic stimuli in normal-hearing listeners. In order to produce a centered image, it was necessary to impose a 10 dB larger voltage on the Etymotic ER-2 than would be expected on the basis of its calibration. Incidentally, the same type of ear-to-ear comparison allows one to compare and to cross-calibrate any earphone to any other one, notwithstanding local variations in the frequency response of the earphones.

[3]The formula used to compute the percentage of the variance for which our predicted values of pointer IID (in dB) accounted was $100 \times (1 - [\Sigma (O_i - P_i)^2]/[\Sigma (O_i - \bar{O})^2])$, where O_i represents the average observed values and P_i represents the predicted values. $\bar{O}$ represents the grand mean of the observed values of IID.

Bernstein, L. R. (**1991**). "Spectral interference in a binaural detection task," J. Acoust. Soc. Am. **89**, 1306–1313.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.

Bernstein, L. R., and Trahiotis, C. (**1992**). "Spectral interference in a binaural detection task: Effects of temporal fringe and masking bandwidth," J. Acoust. Soc. Am. **94**, 735–742.

Bernstein, L. R., and Trahiotis, C. (**1995**). "Binaural interference effects measured with masking-level difference and with ITD- and IID-discrimination paradigms," J. Acoust. Soc. Am. **98**, 155–163.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **112**, 1026–1036.

Bernstein, L. R., and Trahiotis, C. (**2003**). "Enhancing interaural-delay-based extents of laterality at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **113**, 3335–3347.

Bernstein, L. R., and Trahiotis, C. (**2004**). "The apparent immunity of high-frequency "transposed" stimuli to low-frequency binaural interference," J. Acoust. Soc. Am. **116**, 3062–3069.

Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.

Buell, T. N., and Trahiotis, C. (**1993**). "Interaural temporal discrimination using two sinusoidally amplitude-modulated, high-frequency tones: Conditions of summation and interference," J. Acoust. Soc. Am. **93**, 480–487.

Davis, J. B. (**1985**). "Remote frequency masking: Differential effects in binaural versus monaural detection," unpublished thesis, Boston University, Boston, MA.

Domnitz, R. H., and Colburn, H. S. (**1977**). "Lateral position and interaural discrimination," J. Acoust. Soc. Am. **61**, 1586–1598.

Dye, R. H., Jr. (**1990**). "The combination of interaural information across frequencies: Lateralization on the basis of interaural delay," J. Acoust. Soc. Am. **88**, 2159–2170.

Edwards, A. (**1964**). *Experimental Design in Psychological Research* (Holt, Rinehart and Winston, New York).

Hafter, E. R., and De Maio, J. (**1975**). "Difference thresholds for interaural delay," J. Acoust. Soc. Am. **57**, 181–187.

Heller, L. M., and Trahiotis, C. (**1995a**). "Interference in detection of interaural delay in a SAM tone produced by a second, spectrally remote SAM tone," J. Acoust. Soc. Am. **97**, 1808–1816.

Heller, L. M., and Trahiotis, C. (**1995b**). "Interference in detection of interaural delay in a SAM tone produced by a second, spectrally remote SAM tone," J. Acoust. Soc. Am. **97**, 1808–1816.

Heller, L. M., and Trahiotis, C. (**1996**). "Extents of laterality and binaural interference effects," J. Acoust. Soc. Am. **99**, 3632–3637.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Moore, B. C. J. (**1997**). "Frequency analysis and pitch perception," *Handbook of Acoustics*, edited by M. Crocker (Wiley, New York), Sec. III V. III, pp. 1447–1460.

van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.

Rice, S. O. (**1954**). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–294.

Stellmack, M. A., and Dye, R. H. (**1993**). "The combination of interaural information across frequencies: The effects of number and spacing of components, onset asynchrony, and harmonicity," J. Acoust. Soc. Am. **93**, 2933–2947.

Stern, R. M., and Colburn, H. S. (**1978**). "Theory of binaural interaction based on auditory-nerve data. IV. A model for subjective lateral position," J. Acoust. Soc. Am. **64**, 127–140.

Trahiotis, C., and Bernstein, L. R. (**1986**). "Lateralization of bands of noise and sinusoidally amplitude-modulated tones: Effects of spectral locus and bandwidth," J. Acoust. Soc. Am. **79**, 1950–1957.

Trahiotis, C., and Bernstein, L. R. (**1990**). "Detectability of interaural delays over select spectral regions: Effects of flanking noise," J. Acoust. Soc. Am. **87**, 810–813.

Trahiotis, C., Bernstein, L. R., and Akeroyd, M. A. (**2001**). "Manipulating the "straightness" and "curvature" of patterns of interaural cross-correlation affects listeners' sensitivity to changes in interaural delay," J. Acoust. Soc. Am. **109**, 321–330.

Trahiotis, C., and Stern, R. M. (**1989**). "Lateralization of bands of noise: Effects of bandwidth and differences of interaural time and phase," J. Acoust. Soc. Am. **86**, 1285–1293.

Watson, C. S., and Mittler, B. (**1965**). "Time-intensity equivalence in auditory lateralization: A graphical method," Psychonomic Sci. **2**, 219–220.

Woods, W. S., and Colburn, H. S. (**1992**). "Test of a model of auditory object formation using intensity and ITD discrimination," J. Acoust. Soc. Am. **91**, 2894–2902.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.

Zurek, P. M. (**1985**). "Spectral dominance in sensitivity to interaural delay for broadband stimuli," J. Acoust. Soc. Am. **78** Suppl. 1, S18.

Effects of bite blocks and hearing status on vowel production[a]

Harlan Lane[b]
Department of Psychology, Northeastern University, Boston, Massachusetts 02116 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Margaret Denny
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Frank H. Guenther
Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Melanie L. Matthies
Department of Health Sciences, Boston University, Boston, Massachusetts 02215 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Lucie Menard
Université de Québec a Montréal, Quebec, Canada and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Joseph S. Perkell
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139; Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215; and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Ellen Stockmann
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Mark Tiede
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Haskins Laboratories, New Haven, Connecticut 06511

Jennell Vick
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98105 and Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

Majid Zandipour
Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139 and Department of Cognitive and Neural Systems, Boston University, Boston, Massachusetts 02215

---

[a]The order of authors after the first is alphabetical.
[b]Electronic mail: lane@neu.edu

This study explores the effects of hearing status and bite blocks on vowel production. Normal-hearing controls and postlingually deaf adults read elicitation lists of /hVd/ syllables with and without bite blocks and auditory feedback. Deaf participants' auditory feedback was provided by a cochlear prosthesis and interrupted by switching off their implant microphones. Recording sessions were held before prosthesis was provided and one month and one year after. Long-term absence of auditory feedback was associated with heightened dispersion of vowel tokens, which was inflated further by inserting bite blocks. The restoration of some hearing with prosthesis reduced dispersion. Deaf speakers' vowel spaces were reduced in size compared to controls. Insertion of bite blocks reduced them further because of the speakers' incomplete compensation. A year of prosthesis use increased vowel contrast with feedback during elicitation. These findings support the inference that models of speech production must assign a role to auditory feedback in error-based correction of feedforward commands for subsequent articulatory gestures. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2001527]

I. INTRODUCTION

As part of a broader inquiry into speech production, our laboratory has had a longstanding interest in determining the role of auditory feedback in that performance (Perkell *et al.*, 1992; Guenther and Perkell, 2004). To that end, our research has investigated speech production in late-deafened adults who have had no auditory feedback for many years. When those adults have some hearing restored with cochlear implants, we measure acoustic parameters of their speech again, as well as their phoneme perception. In some studies such as the present one, we also switch off implant microphones in order to block auditory feedback for short periods of time. Finally, we make measures of speech perception and production in participants with normal hearing, in order to have some normative benchmarks.

One method of investigating the role of auditory feedback in speakers with normal hearing that has attracted wide interest is to measure speech parameters while the subject clenches a bite block between the teeth and produces a set of utterances. It turns out that such speakers' vowel formants are only slightly altered from their unperturbed state. Moreover, evidence for this compensation can be found in measures of the first glottal pulse, indicating that most of the needed reorganization was not accomplished with the aid of closed loop auditory feedback (Lindblom and Sundberg, 1971; Lindblom *et al.*, 1978; Lubker, 1979; Flege *et al.*, 1988; Baum *et al.*, 1996). McFarland and Baum (1995) found that compensation to bite blocks was neither immediate nor complete and, at least for vowels, compensation strategies may develop over time. Flege *et al.* (1988) measured linguapalatal contact patterns in /s/ and /t/ and found that ten minutes' conversation with bite block led to normal acoustic parameter values; they suggest that sensory feedback may have provided error based correction of articulatory gestures. Hoole (1987) gave bite blocks to a German-speaking patient who had suffered loss of oral sensation due to head trauma. Neutralization of vowels was particularly marked in the bite block plus masking condition and there was no evidence of learning. Hoole concludes that afferent information is required to "establish a frame of reference for motor commands" (p. 19).

The combined interventions of bite block and changes in hearing status present an opportunity to test several hypotheses concerning dispersion and vowel contrast. The present experiment employs postlingually deaf adults to test these hypotheses since prelingually deaf participants may not have mastered the phonemic system or the motor commands and feedback mechanisms required for speech. Remarkably, there appears to be only a single published study in which bite blocks were used with postlingually deaf adults (Tye *et al.*, 1983).

Should we expect that late-deafened adults, asked to speak while clenching a bite block between their teeth, will compensate as accurately as their normal-hearing counterparts? On the one hand, yes, for late-deafened adults had the normal opportunity to master speech production and to monitor their own speech aurally (as well as somesthetically). If speakers with normal hearing do not require closed-loop auditory feedback for accurate compensation, then late-deafened adults who have little or no auditory feedback may have no disadvantage in this regard. On the other hand, no, for late deafened adults have been for some time without sufficiently acute hearing to "update" their model of the relations between the configuration of the vocal tract and its auditory consequences. Speakers with normal hearing are able to update their model when there is drift, when oral morphology has changed, or indeed, when speaking with something in their mouths (Jones and Munhall, 2003).

A. Theoretical background

In thinking about such issues and in formulating experimental hypotheses, our theoretical framework has been guided by the DIVA model of speech motor planning (Guenther and Perkell, 2004). When controlling a model of the vocal tract, DIVA accounts qualitatively for a wide range of observations on speech production, including speech sound acquisition, articulatory variability, motor equivalence, coarticulation and rate effects (Guenther, 1994; Guenther, 1995; Guenther *et al.*, 1998).

Figure 1 presents a schematic of how feedforward and feedback mechanisms are hypothesized to interact in speech motor planning according to our framework. Representations of speech sounds, syllables, and words are stored in long-
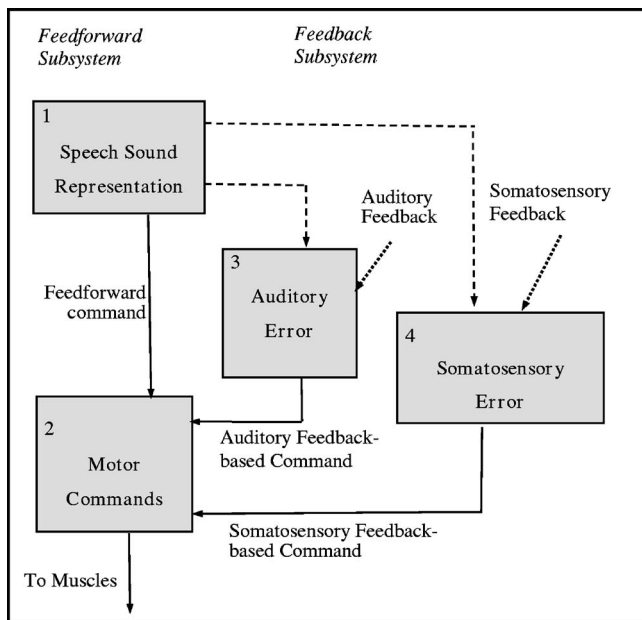
FIG. 1. Schematic diagram of the functionality of the DIVA model. Cells in premotor, motor, and sensory cortical areas are indicated by boxes. Dashed lines: projections from premotor to sensory cortex that encode expected sensory consequences. Dotted lines: projections from sensory periphery to sensory cortex. Solid lines: projections that encode motor commands.

term memory (box 1). (These representations are acquired as the developing child generates vocal-tract movements and begins to form associations among the motor commands, the resulting sounds, and their linguistic significance.) The production of an utterance consisting of a sequence of sounds involves the activation of two sets of projections from the stored representations: (1) projections in the feedback subsystem of *expected auditory consequences* of the movements to be produced. These projections, called *auditory goals*, correspond to regions in multidimensional auditory-temporal space. (2) *Feedforward* projections that result in motor commands to the articulators. As the movements evolve and sound is generated, auditory feedback about the movements' actual consequences is compared with the auditory goals, and if there is a mismatch, an auditory error signal (box 3 in Fig. 1) arises and leads to auditory feedback-based corrective motor commands. These corrective motor commands are also used to help refine feedforward commands for subsequent movement attempts. Eventually, as feedforward commands become well tuned, the control becomes almost entirely feedforward.[1] As the child is learning to produce speech sounds, patterns of somatosensory feedback are also generated. These patterns become incorporated into somatosensory goals, which are used in comparison with somatosensory feedback for error correction as part of the feedback control subsystem.

B. Vowel dispersion hypotheses

Hypothesis (1): A bite block will increase *deaf speakers' dispersion* of vowel tokens around their target means in the formant-1 $\times$ formant-2 plane.

When late-deafened adults have had little or no hearing available for a long time, the bite block is expected to disrupt

their speech production extensively since such speakers' feedforward commands (box 1 to box 2, Fig. 1) were established many years earlier. With a bite block inserted, that mapping is likely to be invalid in some degree but speakers have no way without being able to discriminate the changes in their speech to update such mappings in order to achieve the desired auditory phonemic goals.

Hypothesis (2): The effects of the bite block on *implant users' vowel dispersion* will be less after a year of prosthesis use compared to a month of such use.

When the deaf participant has been using a cochlear implant for an extended period, he or she is better able to retune the articulatory-to-acoustic mappings that will be used to compensate for the subsequent introduction of the bite block.

Hypothesis (3): Speakers with *normal hearing* will respond to the introduction of a bite block with lesser increases in *dispersion* of vowel tokens than implant users.

Speakers with normal hearing are expected to have more accurate articulatory to auditory mappings in the first place since they have had the possibility of updating those mappings continually. Further, their auditory error signal may be more accurate since their hearing does not have the reduced bandwidth and distortion arising from the use of a cochlear implant. Consequently, when movement trajectories for achieving a given vowel target are modified after a bite block has been introduced, the range of those trajectories is more constrained in normal-hearing speakers.

Hypothesis (4): Implant users and speakers with normal hearing, speaking with *bite blocks* inserted, will increase vowel token *dispersion* during short-term interruptions in auditory feedback.

In the presence of the bite block perturbation, movement trajectories are more consistent across trials when an auditory error signal is available to correct subsequent feedforward commands; when that signal is interrupted, dispersion increases.

C. Vowel separation hypotheses

The preceding hypotheses regarding the effects of bite block and hearing status on dispersion are linked to hypotheses concerning their effects on vowel contrasts, for high dispersion leads to low contrast (and conversely) according to the theoretical framework we have sketched. To see the reason for this connection, consider Fig. 2, a schematic of two trajectories that pass through a sequence of goal regions, C1, V1, C2, and V2. It will be seen that the larger goal regions (closed squares) give rise to reduced vowel contrast (closed double-headed arrow). That is because the trajectory, guided by least effort (Lindblom, 1980; 1990), passes through the most proximal parts of the goal regions, thereby reducing its travel and hence vowel separation. Therefore:

Hypotheses (5): Insertion of a *bite block* will decrease vowel *separation*, (6) less so for hearing speakers than for implant users; (7) vowel *separation* will increase after extended *implant use*, and (8) it will diminish with *short-term interruptions* of auditory feedback.
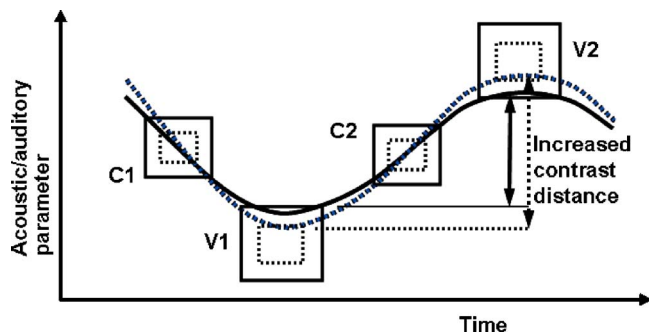
FIG. 2. According to the DIVA model, during speech production an auditory trajectory is planned to pass through the sequence of auditory goals that correspond to the input phoneme string. When the dispersion of the goal region is reduced (dotted squares), the contrast distance between the goals increases (dotted double-headed arrow).

II. METHOD

A. Participants

The experimental group was comprised of six male and two female postlingually deaf, adult, paid volunteers, speakers of Standard American English. Table I shows that the group was heterogeneous in several respects, as is characteristic of this population. Age at onset of hearing loss ranged from birth to 60 years; age at onset of profound loss from 26 to 72. Duration of profound loss varied from less than a year to ten years; the average age of receiving an implant ranged from 28 to 78. Subjects' consonant recognition scores measured before activation of their implants ranged from 21% to 49% correct, and their vowel recognition scores from 25% to 64%. For a description of the phoneme recognition test, see Lane *et al.* (2005) The implant was either the Clarion (Advanced Bionics), cf. Wilson *et al.* (1995); or the Nucleus 24, (Cochlear Corp.), cf. Blamey *et al.* (1987) and McKay and McDermott (1993).

A control group was comprised initially of four male and six female participants, speakers of Standard American English, who reported no speech or hearing anomalies. A screening test was administered to speakers in this group over age 40, in order to determine approximate thresholds at 0.5, 1, 2, and 4 kHz. After a practice tone at 50 dB HL, sound pressure was increased in 5 dB increments from 0 to 25 dB HL or until the subject reported hearing a tone. The series was presented twice at each of the four frequencies to each ear. Subjects who failed to report hearing the tone in any of the 16 series were excluded; this resulted in excluding two male participants, leaving eight subjects in the group. Cochlear implant users participated in four recording sessions, two pre-implant, one approximately a month after activation of their implant speech processors, the other approximately a year after activation. The normal-hearing controls served only once.

B. Procedure

The elicitation set consisted of five /hVd/ words. Before each utterance in a bite block condition, a bite block made of plastic (the kind used in football mouth guards) was inserted between the participant's upper and lower left first bicuspid. The thickness of the bite blocks varied inversely with typical jaw opening for each vowel, as follows: *heed* 2.5 cm; *hid* 2.0 cm, *head* 1.5 cm, *had* 1.0 cm, *hot* 0.5 cm. These sizes were chosen to significantly disrupt the vertical displacement of the jaw relative to its normal position (informally observed) for the particular vowel, e.g., a large bite block was used to force an unnaturally large opening during the high vowel in "heed," whereas a small bite block was used to force an unnaturally small opening for the low vowel in "hot."

Implant users were tested in two sessions prior to activation of the speech processors of their cochlear implants, at one month postactivation and at one year postactivation. For the control group, only a single time sample was recorded. For all subjects, twenty repetitions of each of the five hVd words, uttered in isolation with emphasis on the vowel, were elicited in random order in each of four experimental conditions—the four possible combinations of auditory feedback available or not and bite block inserted or not: (1) No bite block, auditory feedback; (2) No bite block, no feedback; (3) Bite block, no feedback; (4) Bite block, auditory feedback. Since a vowel-specific bite block was inserted for each trial and the vowel order was randomized, the bite block was changed between trials in the bite block conditions. The four experimental conditions were always presented in the order just given with one exception: Feedback was not varied with the deaf participants in the recording

TABLE I. Characteristics of participants with cochlear implants.

Speakers	FI	FJ	MJ	MK	ML	MM	MO	MP
Consonant perception (% correct)	26	25	21	24	21	32	49	21
Vowel perception (% correct)	29	26	35	25	64	32	51	27
Etiology	Autoimmune response	Infection	Hereditary	Unknown	Infection	Noise (WWII)	Blood clot	Hereditary
Age at onset of change in hearing	19	5	Early 30s	18	10	20	60	Birth
Age at onset of profound loss	54	45	43	28	48	72	67	26
Age at cochlear implantation	56	46	49	28	48	78	72	36
Hearing aid used before implantation	None	None	Left	None	None	Both	Left	Both
Implant	Nucleus-24	Clarion	Clarion	Clarion	Nucleus-24	Clarion	Clarion	Clarion

session prior to activation of their implants. They received only the no-bite-block followed by the bite-block conditions. To simulate to a limited degree the implant users' lack of auditory feedback in the processor-off condition, the speakers with normal hearing had their auditory feedback masked by speech-shaped noise presented at 95 dB (SPL).

C. Equipment

Speakers were seated in a sound-attenuating room in a comfortable office chair. A head-mounted electret microphone (Audio-Technica, model AT803B) was placed at a fixed distance of 20 cm from the speaker's lips. Elicitation utterances were presented on a monitor. For the feedback-off condition, participants with normal hearing wore calibrated TDH-39 headphones with ear cushions. The headphones were supplied with noise from a custom-built device (Technical Collaborative, Lexington, MA). The noise was approximately speech-shaped, with a spectral envelope that rolls off at 6 dB per octave. Participants with cochlear prostheses used the normal settings of their speech processors for the feedback-on conditions. For the feedback off conditions, the input to their processors was turned off.

For calibration of sound pressure level, a sound generator (electrolarynx) was placed in front of the speaker's lips while an experimenter observed the sound pressure level on a sound level meter (C scale) placed next to the microphone. The calibration signal and participants' speech were low-pass filtered at 7.2 kHz and digitized in real-time with a 16 kHz sampling rate.

D. Data extraction

The start and end of each token (/hVd/) were labeled automatically based on smoothed rms thresholding exceeding 20% of the peak value for the utterance. The first and second formants were extracted algorithmically from an LPC spectrum around mid-vowel (half the interval from onset to offset) using a 40 ms analysis window for F0 and a 25 ms window for the formants. The algorithm displayed, for each vowel token, (a) LPC-derived formant estimates computed initially at mid-vowel; (b) a broadband spectrogram on which was superimposed the formant trajectories that were detected; and (c) the spectral cross section at the measurement time. In a preliminary pass through the data, the experimenter set the formant bandwidth and LPC filter order to optimize formant identification in each subject's vowel tokens. In the data collection pass, if three formants were detected unambiguously in the regions expected for that vowel target, the experimenter accepted that token with those values as complete and moved on to the next. If the algorithm did not succeed in detecting a formant, the experimenter shifted the time of spectral analysis between one and three pitch periods.

To measure dispersion, we determined for each utterance of a phoneme the Euclidean distance of that single token in the $F1 \times F2$ plane from the average position of all the tokens of that phoneme. Then, those deviations were converted to logarithms (to render the distributions more normal) and averaged across repetitions and subjects to obtain
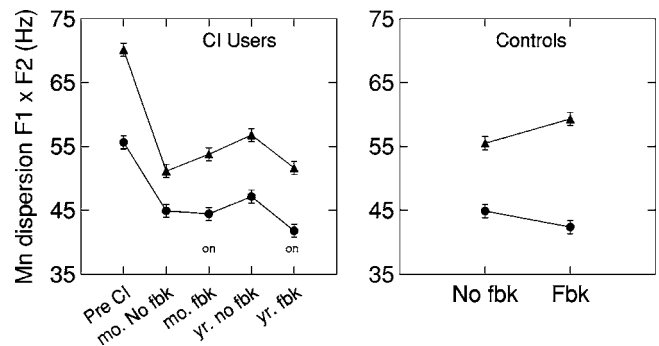
FIG. 3. Mean token dispersion in the formant plane for the five vowels elicited. For the implant users (left pane), dispersion is plotted as a function of the time sample, and presence ("ON") or absence of auditory feedback during elicitation, with bite block insertion (triangles) as a parameter. For the speakers with normal hearing, the treatments were presence or absence of auditory feedback and of bite block (triangles). (For the cochlear implant users, each point is the mean of 800 determinations, 20 repetitions of each of 5 vowels produced by 8 speakers. For the hearing controls, each point is the mean of 700 determinations, 20 repetitions of 5 vowels produced by 7 speakers.) Error bars are one standard error of the mean.

the mean dispersion for that phoneme under the given experimental condition. To measure acoustic contrast distance, Mahalanobis distances were calculated for the ten possible pairs of five vowels, on each of twenty repetitions. Each repetition of a particular vowel on the ith trial is given a Mahalanobis distance to the distribution of each of the other vowels. The square roots of the distances were averaged for each group and listening condition.[2]

III. RESULTS

A. Dispersion

The average dispersion of vowel phones around their mean in the formant plane is plotted in Fig. 3, for implant users and hearing controls separately. For the implant users (left pane), dispersion is plotted as a function of the time sample, and presence ("ON") or absence of auditory feedback during elicitation, with bite block insertion (triangles) as a parameter. For the hearing speakers, the treatments were presence or absence of auditory feedback and of bite block.

Hypothesis (1) states that the bite block will increase implant users' vowel dispersion. As the figure shows, it did so under all conditions of the experiment and especially before speakers had some hearing restored by cochlear prosthesis. Hypothesis (2), which states that the effect of the bite block will decline over time as the implant user gains experience with auditory feedback, is not supported: the effect of time sample (one month or one year after processor activation) and the time $\times$ bite block interaction are not significant. (The results of an ANOVA are reported in Table II). Hypothesis (3) states that vowel dispersion measured on speakers with normal hearing will increase less as a result of introducing the bite blocks than dispersion measured on implant users. These outcomes must be compared cautiously as the groups are small and differ in several respects, including gender and age. Contrary to the hypothesis, as Fig. 3 shows, the smaller increase in dispersion for the hearing controls, 10 Hz, was as large or larger than all the implant users' in-

TABLE II. Analysis of variance of dispersion measures for eight implant users (*=$p<0.05$).

Source	df	F	Eta^2 (%)
Between subjects	7 115	41.6*	71
Time sample	1 115	0.3	0.2
Time*subject	7 115	9.0*	35
Vowel	4 460	4.0*	3
Vowel*subject	28 460	3.2*	16
Feedback	1 115	5.7*	4
Feedback*subject	7 115	2.6*	13
Bite block	1 115	97.8*	45
Bite block*subject	7 115	12.6*	43
Time*vowel	4 460	2.1	1.7
Time*vowel*subject	28 460	3.7*	18
Time*feedback	1 115	3.5	2
Time*feedback*subject	7 115	2.2*	11
Time*bite block	1 115	1.3	1
Time*bite block*subject	7 115	5.6*	25
Vowel*feedback	4 460	0.4	0.3
Vowel*feedback*subject	28 460	2.8*	14
Vowel*bite block	4 460	6.3*	5
Vowel*bite block*subject	28 460	2.8*	14
Feedback*bite block	1 115	1.0	0
Feedback*bite block*subject	7 115	0.5	2
Time*vowel*feedback	4 460	2.9*	2
Time*vowel*feedback*subject	28 460	1.4	7
Time*vowel*bite block	4 460	2.1	1
Time*vowel*bite block*subject	28 460	2.7*	14
Time*feedback*bite block	1 115	0.1	0.1
Time*feedback*bite block*subject	7 115	2.7*	14
Vowel*feedback*bite block	4 460	1.1	0.9
Vowel*feedback*bite block*subject	28 460	1.3	7
Time*vowel*feedback*bite block	4 460	0.8	0.6
Time*vowel*feedback*bite block*subject	28 460	2.6*	13

TABLE III. Analysis of variance of dispersion measures for seven control subjects with normal hearing (*=$p<0.05$).

Source	df	F	Eta^2 (%)
Between subjects	6 130	36.1*	62
Vowel	4 520	4.1*	3
Vowel*subject	24 520	3.6*	14
Feedback	1 130	0.0	0
Feedback*subject	6 130	3.6*	14
Bite block	1 130	86.0*	39
Bite block*subject	6 130	6.5*	23
Vowel*feedback	4 520	0.7	0.5
Vowel*feedback*subject	24 520	3.3*	13
Vowel*bite block	4 520	6.9*	5
Vowel*bite block*subject	24 520	2.0*	8
Feedback*bite block	1 130	4.7*	3
Feedback*bite block*subject	6 130	6.5*	23
Vowel*feedback*bite block	4 520	2.7*	2
Vowel*feedback*bite block*subject	24 520	3.2*	12

dence that this enlarged dispersion declined after greater experience hearing with the implant. Interrupting feedback did increase implant users' vowel dispersion but only in the one-year sample. It also increased the controls' dispersion but only without the bite blocks. The bite blocks were approximately equally disruptive for implant users and hearing controls.

B. Vowel contrast

Figure 4 presents a measure of speakers' compensations for the bite blocks. The vertical axes plot the average effect on F1 of inserting the bite block, compared to measures

creases in dispersion except for the preimplant condition. The mean increases were 13.7 Hz for the hearing subjects and 9.9 Hz for the implant users. The very high dispersion in the preimplant condition with no auditory feedback and bite blocks may be due to the elevated dispersion found without bite blocks in that condition. The increase in dispersion in that condition due to bite block, 26%, is comparable to the parallel increase obtained from the controls, 24%.

Hypothesis (4) states that dispersion of vowel tokens will be greater without auditory feedback than with it. It is supported by the findings with the implant users who yielded a significant effect of feedback condition (Table II), due primarily to their vowel dispersion measured one year after their implant speech processors had been activated. For the hearing speakers (Table III), the effects of interrupting auditory feedback on vowel dispersion depended on whether the bite blocks were present or not. Without bite blocks, dispersion was greater in the absence of feedback, consistent with hypothesis (4) ($F=3.8, df=7, 130, p<0.01$). However, with bite blocks, dispersion was higher in the presence of feedback than in its absence ($F=5.8, df=7, 130, p<0.01$), as was the case with the implant users in the one month sample.

To summarize: the bite blocks did indeed increase the dispersion of implant users' vowel phones, especially prior to their receiving their prostheses. However, there was no evi-

FIG. 4. (Color online) Compensation for bite block in three time samples for implant users and one for controls. The vertical axis plots the average effect on F1 of inserting the bite block, compared to measures without a bite block, and normalized by the latter (percent change). The horizontal axis plots the comparable measure for F2. Closed circles represent measures made with auditory feedback available; open, no feedback. (For the cochlear implant users, each point is the mean of 160 determinations, 20 repetitions by 8 speakers. For the hearing controls, each point is the mean of 140 determinations, 20 repetitions by 8 speakers.)

TABLE IV. Analysis of variance for the measures of average vowel space for implant users ($*=p<0.05$).

Source	df	F	Eta² (%)
Between subjects	7 115	4 053*	99
Time sample	1 115	580*	83
Time*subject	7 115	1 354*	98
Vowel	4 460	11 900*	99
Vowel*subject	28 460	339*	95
Feedback	1 115	503*	81
Feedback*subject	7 115	515*	96
Bite block	1 115	14 500*	99
Bite block*subject	7 115	1 202*	98
Time*vowel	4 460	3*	2
Time*vowel*subject	28 460	29*	63
Time*feedback	1 115	1 824*	94
Time*feedback*subject	7 115	410*	96
Time*bite block	1 115	114*	49
Time*bite block*subject	7 115	576*	97
Vowel*feedback	4 460	28*	19
Vowel*feedback*subject	28 460	28*	63
Vowel*bite block	4 460	679*	85
Vowel*bite block*subject	28 460	50*	75
Feedback*bite block	1 115	516*	81
Feedback*bite block*subject	7 115	250*	93
Time*vowel*feedback	4 460	56*	32
Time*vowel*feedback*subject	28 460	18*	52
Time*vowel*bite block	4 460	50*	30
Time*vowel*bite block*subject	28 460	13*	44
Time*feedback*bite block	1 115	6*	4.9
Time*feedback*bite block*subject	7 115	186*	91
Vowel*feedback*bite block	4 460	20*	14
Vowel*feedback*bite block*subject	28 460	5*	23
Time*vowel*feedback*bite block	4 460	1	0.8
Time*vowel*feedback*bite block*subject	28 460	10*	37

FIG. 5. The average size of the vowel space (AVS) is expressed as the mean of (the square root of) the Mahalanobis distances between the members of all ten possible pairs of the five vowels. For the implant users (left pane), AVS is plotted as a function of the time sample, and presence ("ON") or absence of auditory feedback during elicitation, with bite block insertion (triangles) as a parameter. For the controls, the treatments were presence or absence of auditory feedback and of bite block (triangles). (For the cochlear implant users, each point is the mean of 800 determinations of average vowel spacing. The 4 distances from each vowel token to the other tokens in each repetition set were averaged, and a grand mean obtained for the 20 repetitions of each of the 5 vowels produced by 8 speakers. For the hearing controls, each point is the mean of 700 determinations, 20 repetitions of 5 vowels produced by 7 speakers.) Error bars are one standard error of the mean.

without a bite block, and normalized by the latter (in percent) ("re: normal"). The horizontal axis plots the comparable measure for F2. Closed circles represent measures made with auditory feedback available; open, no feedback. The finding that most points lie within a few percent of the dotted lines at F1 % change=0 and F2 % change=0 indicates that our speakers, both implant users and controls, compensated for the bite block in a way that allowed them nearly to restore their formant frequencies in the unperturbed state. Nevertheless, the mean percent changes in F1 and F2, pooling across subjects and vowels (N=approximately 750 values) were significantly different from zero in every case but one (viz., F2 at one year without feedback). Those changes were, on average, almost identical comparing feedback and no feedback conditions. Not surprisingly, /i/ is one of the vowels most resistant to compensation as it is the vowel with the highest place of articulation in this study and thus particularly hindered by its 2.5 cm bite block.

The effects of the experimental interventions on the vowel space will be seen in Fig. 5 and Table IV. Findings for implant users and controls are shown separately. The average size of the vowel space (AVS) is expressed as the average Mahalanobis distance between the members of all ten possible pairs of the five vowels. For the implant users (left pane), AVS is plotted as a function of the time sample, and

presence ("ON") or absence of auditory feedback during elicitation, with bite block insertion (triangles) as a parameter. For the controls, the treatments were the presence or absence of auditory feedback and of bite block (Table V).

Hypothesis (5) states that the insertion of a bite block will reduce the distances in the formant plane separating vowels. It is supported by visual inspection of Fig. 5, where the triangles fall below the circles in both panes, and by ANOVAs for both groups of speakers. Hypothesis (6) states that the reduction in size of the vowel space due to the bite block will be less for the controls than for the implant users. This expectation is clearly not borne out, whether one measures the decrease in AVS in absolute terms (4.3 units for controls, 2.5 for implant users, averaging across all values) or relative terms (31% for controls, 22% for implant users). The reduction in the controls' vowel space with a bite block appears to be so large, compared to the implant users, at least in part because they had a larger vowel space in the first place (14.0 vs 11.2 units). This view is supported by the observation that AVS for the controls with a bite block is higher than the AVS for implant users with a bite block (9.7 units vs 9.0), despite the bigger percent reduction in AVS for normal-hearing speakers ($t=4.2, df=199, p<0.01$).

Hypothesis (7) states that vowel contrast will improve after extended implant use. Figure 5 shows that with auditory feedback (indicated by "on"), the size of the implant users' vowel space increased from one month to one year (both with and without bite block). However, without feedback no such improvement appeared: the vowel space was unchanged in size (no bite blocks, $t=1.2, df=784, p=0.25$) or became a half unit smaller (bite blocks, $t=5.7, df=783, p<0.01$). Examining normal-hearing speakers' productions of /s/ with and without masking while they wore a dental prosthesis, Jones and Munhall (2003) report a parallel result: There was improvement in compensating for the prosthesis but only when speakers could hear their speech.

TABLE V. Analysis of variance for the measures of average vowel space for control speakers with normal hearing (*=$p<0.05$).

Source	df	F	Eta^2 (%)
Between subjects	6 130	3 574*	99
Vowel	4 520	9 921*	98
Vowel*subject	24 520	208*	90
Feedback	1 130	257*	66
Feedback*subject	6 130	1 517*	98
Bite block	1 130	25 100*	99
Bite block*subject	6 130	1 269*	98
Vowel*feedback	4 520	65*	33
Vowel*feedback*subject	24 520	36*	62
Vowel*bite block	4 520	964*	88
Vowel*bite block*subject	24 520	70*	76
Feedback*bite block	1 130	78*	37
Feedback*bite block*subject	6 130	1 492*	98
Vowel*feedback*bite block	4 520	5*	3
Vowel*feedback*bite block*subject	24 520	23*	51

Hypothesis (8) states that the size of the vowel space will be greater with auditory feedback than without it. This was confirmed by measures taken, both with and without bite blocks, at implant users' one year session but was not observed at the one month session where, on the contrary, AVS with bite blocks inserted was larger in the absence of feedback ($t=15.3, df=787, p<0.01$). Likewise, the controls showed significantly larger vowel spaces without auditory feedback across bite block condition (Table V).

To summarize: the implant users and hearing controls in this experiment did not compensate fully for the shift in formant values caused by insertion of the bite blocks. Consequently there were net changes to the size of the vowel space and those changes varied with the samples of time since implantation and the presence or absence of auditory feedback. The insertion of the bite blocks did indeed reduce the distances separating vowels for both groups. However, the hearing controls did not appear to be better compensators; on the contrary, they had greater compression of the vowel space in both relative and absolute terms, though their AVS with the bite block was still higher than that for the implant users with the bite block. The implant users' vowel spaces had grown by the one-year session but not in the first month. Finally, interrupting implant users' auditory feedback at one year led to a reduction in their vowel spaces, consistent with the hypothesis, but interruption produced the opposite effect in the controls.

IV. DISCUSSION

A. Dispersion

The dispersion of vowel tokens around their means (Fig. 3) was substantially increased by insertion of the bite blocks, confirming hypothesis (1). For the implant users, dispersion of vowel productions was highest before they received prosthetic hearing; at this time most of the speakers had too little audition to distinguish among vowels and that had been the case for some years. Then the bite blocks were introduced exacerbating the already elevated vowel dispersion. In terms

of the model in Fig. 1, the deaf speakers' poor discriminative capacity deprived them of accurate auditory error signals (box 3) that would otherwise modify future feedforward commands (setting aside for now the issue of closed-loop error correction within a syllable). The lack of an auditory error signal caused the deaf speakers to produce a greater range of trajectories (Fig. 2) passing through the phonemic goal regions, and that greater range was reflected in the mid-vowel formant measures and thus in the measured dispersion. Despite this variability, the late-deafened speakers compensated astonishingly well to bite blocks after decades with degraded or no auditory feedback (cf., Fig. 4). This may be a reflection of goal regions that are highly stable over time and over feedback conditions, while the changes in dispersion due to bite block and changes in feedback may be attributable to changes in trajectories passing through the outskirts of those goal regions. The DIVA model postulates that goal regions and feedforward commands degrade very slowly over time with extensive hearing loss. [On the stability of the goal regions, see Perkell *et al.* (1992).]

Consistent with our findings, the sole study of the effects of bite block on late-deafened adults that has come to our attention (Tye *et al.*, 1983) reports that two deaf speakers showed more variable tongue shapes and positions, both with and without bite blocks, than did three hearing controls. The dramatic drop in implant users' vowel dispersion in the present study that occurred a month after they had some hearing restored, and the long-term decrease of their dispersion into the normal range, confirm that auditory feedback reduces production variability (Harris *et al.*, 1985; McGarr and Campbell, 1995), possibly by making trajectories smoother and more accurate.

One month of auditory experience with the implant markedly reduced dispersion both with and without the bite blocks. The enlarged vowel dispersion due to the bite blocks did not diminish significantly as the implant users gained auditory experience from one month to one year (Fig. 3), disconfirming hypothesis (2). (If auditory feedback were a guarantor of reduced variability, the speakers with normal hearing would not have increased dispersion as much as they did with the introduction of the bite block.) At one year after activation of their speech processors, however, implant users' dispersion reached a minimum without the bite blocks, indicating a greater role for auditory feedback after longer experience with the implant.

The controls were also more variable in their vowel productions with introduction of the bite blocks. The increase in their dispersion compared to their vowel productions without bite blocks was not less than that of implant users, contrary to hypothesis (3).

Finally, do short-term interruptions of auditory feedback increase vowel token dispersion, as hypothesis (4) posits? The answer is a qualified "yes." For the implant users, dispersion was greater without auditory feedback in the sample at one year but not that at one month. Auditory feedback may not have reduced dispersion at one month because prosthetic feedback is quite different from the speaker's former auditory feedback. Thus it takes some time before the auditory system reorganizes to the point where it can make use of the

new auditory information for guiding speech. For the control speakers, dispersion without bite blocks was also greater without auditory feedback. However, with the bite blocks in place, the controls' vowel dispersion was unexpectedly less without feedback than with it, possibly because speakers had adopted a "clear speech" strategy in the face of bite blocks and no feedback (Lane and Tranel, 1971; Picheny *et al.*, 1986).

B. Vowel contrast

Tye *et al.* (1983) found that two deaf speakers (one congenitally deaf, the other postlingually deaf), exhibited articulatory compensation for a bite block. Campbell (1999) reported that a group of six congenitally, profoundly deaf speakers not only compensated but in fact overcompensated for a bite block in production of high vowels. In the present study, the implant users' compensation for the bite blocks was excellent (Fig. 4), with most vowel means falling within 5% of perfect compensation. It is questionable whether the implant users, given the limited fidelity of their prosthetic hearing, were able to detect that they had undercompensated. With speakers having normal hearing, Baum *et al.* (1996) found no effect of bite block on the speaker's vowel intelligibility and rated vowel quality (except for /i/). Fowler and Turvey (1980) found that bite block vowels were slightly less identifiable than unperturbed vowels for three of their four speakers, but the bite block had a substantial effect on vowel identification with only one speaker.

The significant shortfall in bite block compensation that remained after our speakers' compensatory movements is reflected in the average vowel spacing as a reduction (Fig. 5), as predicted by hypothesis (5). Turning to the control speakers, Fig. 5 shows that with bite blocks in place they also reduced their vowel spaces but to a much greater degree than the implant users thus disconfirming hypothesis (6) that their vowel contrast would be less affected than that of the implant users.

There is evidence, confirming hypothesis (7), that vowel contrast improved with extended use of the auditory prosthesis: with auditory feedback, the average vowel space grew from the one-month to the one-year sample (Fig. 5) both with and without bite blocks. With sustained auditory feedback and hearing speakers, Baum *et al.* (1996) found that 15 min of conversation with bite blocks in place enhanced compensation. Flege *et al.* (1988) obtained a similar result. We examined our data for a short-term practice effect by comparing percent bite block compensation on the first and last (twentieth) trial in the one-month and one-year samples; however, we found none. Most likely, it was the extended experience with their prostheses that led implant users to increase contrast distance by the time of the one-year sample. Economou *et al.* (1992) reported on a postlingually deaf preadolescent who had a failed single-channel implant replaced by a multichannel device. Eighteen days after the failure of the single-channel implant, his vowel space was reduced and after a year's experience with the multichannel implant, it was significantly expanded.

Was the measured vowel space smaller without auditory feedback than with it, as hypothesis (8) claims? This proved to be true of the implant users at one year (but not at one month). Svirsky *et al.* (1992) found with implant users that vowel contrast decreased when auditory feedback was interrupted and expanded when it was restored. Matthies *et al.* (1996) reported contrast reductions for fricatives with interrupted auditory feedback in implant users and Lane *et al.* (1995) found the same for plosive voicing.

However, turning to the speakers in the present study with normal hearing, Fig. 5 shows they did not have smaller vowel spaces when their auditory feedback was masked. It is possible that there were differences among the vowels that their two-formant specification did not capture. In the study cited earlier of /s/ productions while wearing a dental prosthesis, Jones and Munhall (2003) found the spectra of the speakers' /s/ productions moved toward that of normal, unperturbed production with increased experience with the prosthesis. Although an acoustic analysis did not show any significant differences depending on auditory feedback, listener ratings proved to be higher for utterances recorded with feedback present.

C. Relating the results to the DIVA model

The disconfirmed hypotheses from these experiments suggest refinements to the DIVA model. In particular, disconfirmation of hypothesis (8) for speakers with normal hearing (they did *not* show greater vowel spacing with auditory feedback) suggests that a change in hearing status may not immediately change average vowel spacing. Instead, hearing status appears to affect average vowel spacing more slowly, as evidenced by the larger AVS of normal-hearing speakers (with or without auditory feedback) compared to the AVS of the deaf speakers prior to implantation (cf. Fig. 5). With regard to the DIVA model, this suggests that the greater vowel space of normal-hearing speakers relative to the deaf is not due to the hearing speakers' online feedback control mechanisms (otherwise AVS would have increased with auditory feedback in these speakers), but instead due to the slow degradation of the deaf speakers' feedforward commands. Currently the DIVA model does not explicitly include degradation of feedforward commands in the absence of auditory feedback unless the vocal tract morphology changes; instead the commands remain tuned indefinitely when feedback is removed.[3] Such degradation could be incorporated by including a very slow decay term for the synaptic weights encoding feedforward commands in the model.

One year after implantation, implant users still show a relatively small AVS in the absence of auditory feedback, but their AVS increases when the implant is turned on. This suggests that their feedforward commands are still not fully tuned, but experience with the implant has allowed them to learn new auditory targets (corresponding to the initially novel auditory signals produced by the implant), which in turn allows their auditory feedback control subsystems to perform online corrections to their still-degraded feedforward commands. According to the DIVA model, continued practice should eventually lead to improved feedforward commands in these individuals due to this corrective influ-

ence of the auditory feedback control circuit. If this account is correct, this process apparently takes longer than one year.

Disconfirmation of hypothesis (6) (that the reduction in vowel space due to bite block insertion will be smaller for controls than for implant users) appears to be largely explained by the observation that AVS is much larger for controls than implant users before bite block insertion, and thus the reduced range of motion caused by the bite blocks is more disruptive in controls. The greater AVS of hearing speakers compared to deaf speakers is consistent with the idea, outlined in the previous paragraph, that the feedforward commands of hearing-impaired individuals become degraded slowly over time, leading to a reduction of vowel spacing.

Disconfirmation of hypothesis (3) (which states that the effect of the bite block on vowel dispersion should be less for normally hearing speakers than implant users) appears to indicate that hearing does not significantly affect the increased dispersion of vowel tokens caused by a bite block, even after long periods of time with little or no hearing. That is, both hearing and bite block affect dispersion but there is no significant interaction (cf. Table II for confirmation). If hearing does not affect bite block-induced dispersion, this would also account for the disconfirmation of hypothesis (2) (that the deleterious effect of the bite block on vowel dispersion will decrease with time after implantation) and hypothesis (4) (that auditory feedback should reduce the effect of the bite block on dispersion). In the DIVA model, many factors can affect the dispersion of repeated vowel tokens, including (i) size and location of the auditory target region for the vowel, (ii) phonetic context, (iii) change in speaking style and/or rate, (iv) presence or absence of articulator perturbations such as bite blocks, and (v) presence or absence of auditory feedback. Because the dynamics of the model depend in a complex way on each of these factors, their net effect on variability can only be determined through computer simulations of the model on the specific tokens and conditions of the experiment, which is beyond the scope of the current article.

D. Summary

In order to explore the role of audition during speech production, several interventions were made with postlingually deaf speakers and hearing controls: the deaf speakers received cochlear implants and were tested before, one month after, and one year after activation of their speech processors. Both groups received bite blocks and short-term interruptions of auditory feedback. Our theoretical framework assigned to auditory feedback the role of revising feedforward commands controlling trajectories between phoneme goals. With long-term absence of auditory feedback, dispersion of vowel tokens was expected to be large and inflated further by inserting bite blocks, as speakers would have no way to adequately inform themselves of their auditory effects. Once these speakers receive cochlear prostheses, they should be able to use auditory feedback and the dispersion would fall. These expectations were supported. To a degree, the effects of short-term interruptions of feedback paralleled these long-term changes.

The deaf speakers compensated extensively, even before they had prosthetic hearing; however, the compensation was not total. Consequently, insertion of the bite blocks reduced the size of their vowel spaces. The vowel space had grown when it was measured with auditory feedback present a year after processor activation, both without and with bite blocks. Two processes may have been at work after a year of audition: improved vowel contrasts (seen in the no bite block condition) and improved compensation for the bite block perturbation.

Thus auditory experience with a prosthesis brought a reduction of the heightened dispersion when deaf. It reduced dispersion over a year (Fig. 3) and also expanded the vowel space (Fig. 5). These findings indicate that any model of speech production must assign a role to auditory feedback in error-based correction of feedforward commands for subsequent articulatory gestures.

[1]There is a body of literature showing compensatory responses to alterations in auditory feedback, consistent with the hypothesis that it serves to correct perceived errors produced by feedforward commands. However, it remains unclear when auditory feedback updates feedforward commands for the next utterance and when it introduces corrections during a speech sound. Among the variables that have been manipulated with compensatory consequences are amplitude (Lane and Tranel *et al.*, 1971; Siegel *et al.*, 1982); fundamental frequency (Kawahara and Williams, 1996; Burnett *et al.*, 1998; Hain *et al.*, 2001; Natke and Kalveram, 2001); and vowel spectra (Houde and Jordan, 1998, 2002).

[2]The Mahalanobis Distance, D^2, is defined as $(x_i - y_i)' S_i^{-1}(x_i - y_i)$, where x are the individual's scores on a number of measures, y are the corresponding means from a prespecified group, S^{-1} is the inverse covariance matrix. Unlike Euclidean distance, D^2 takes into account not only the average value but also the variance and covariance of the variables measured and it does not assume equal weight for the two predictors, F1 and F2. D^2 measures the overlap of two distributions; it is the basis of a maximum probability classification; and is proportionate to the F ratio as a measure of separation (Taguchi and Jugulum, 2002).

[3]In simulations to date, once the DIVA model has been tuned with a particular vocal tract morphology, removing auditory feedback does not affect the learned target regions, and the feedforward commands remain in their current state unless the vocal tract morphology changes. If such changes occur, the somatosensory feedback subcircuit will cause changes in the feedforward commands, and in the absence of the auditory feedback subcircuit, these changes may be deleterious to the acoustic signal. This may be responsible for the degraded vowel spacing found in pre-implant measurements, but this is impossible to verify without measurements indicating the amount of change in morphology since onset of deafness.

Baum, S. R., McFarland, D. H., and Diab, M. (**1996**). "Compensation to articulatory perturbation: Perceptual data," J. Acoust. Soc. Am. **99**, 3791–3794.

Blamey, P. J., Dowell, R. C., Brown, A. M., Clark, G. M., and Seligman, P. M. (**1987**). "Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors," J. Acoust. Soc. Am. **82**,

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Lane *et al.*: Hearing status and effects of bite blocks   1645

48–57.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Campbell, M. (**1999**). "Articulatory compensation for a biteblock, with and without auditory feedback in hearing and hearing-impaired speakers," Doctoral dissertation, City University of New York (unpublished).

Economou, A., Tartter, V., Chute, P., and Hellman, S. (**1992**). "Speech changes following reimplantation from a single-channel to a multichannel cochlear implant," J. Acoust. Soc. Am. **92**, 1310–1323.

Flege, J. E., Fletcher, S. G., and Homiedan, A. (**1988**). "Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data," J. Acoust. Soc. Am. **83**, 212–228.

Fowler, C. A., and Turvey, M. T. (**1980**). "Immediate compensation in bite-block speech," Phonetica **37**, 306–326.

Guenther, F., and Perkell, J. (**2004**). "A neural model of speech production and its application to studies of the role of auditory feedback in speech," in *Speech Motor Control in Normal and Disordered Speech*, edited by B. Maassen, R. Kent, H. F. M. Peters, P. Lieshout, and W. Hulstijn (Oxford University Press, Oxford), pp. 29–50.

Guenther, F. H. (**1994**). "A neural network model of speech acquisition and motor equivalent speech production," Biol. Cybern. **72**, 43–53.

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.

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.

Harris, K. S., Rubin-Spitz, J., and McGarr, N. (**1985**). "The role of production variability in normal and deviant developing speech," ASHA Reports **15**, 50–57.

Hoole, P. (**1987**). "Bite-block speech in the absence of oral sensibility," in *Proceedings of the XIth International Congress of Phonetic Sciences Tallinn*, pp. 16–19.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech. I. Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Munhall, K. G. (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal fold Physiology*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 263–278.

Lane, H., Matthies, M., Denny, M., Guenther, F. H., Perkell, J. S., Stockmann, E. *et al.* (**2005**). "Effects of short- and long-term changes in auditory feedback on vowel and sibilant contrasts," Ear Hear. (submitted).

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lane, H., Wozniak, J., Matthies, M., Svirsky, M., and Perkell, J. (**1995**). "Phonemic resetting vs. postural adjustments in the speech of cochlear implant users: An exploration of voice-onset time," J. Acoust. Soc. Am. **98**, 3096–3106.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Lindblom, B., Lubker, J., and Gay, T. (**1978**). "Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation," J. Phonetics **7**, 147–161.

Lindblom, B. E., and Sundberg, J. E. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.

Lindblom, B. E. F. (**1980**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 13–44.

Lubker, J. F. (**1979**). "The reorganization times of bite-block vowels," Phonetica **36**, 273–293.

Matthies, M., Svirsky, M., Perkell, J., and Lane, H. (**1996**). "Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant," J. Speech Hear. Res. **39**, 936–946.

McFarland, D. H., and Baum, S. R. (**1995**). "Incomplete compensation to articulatory perturbation," J. Acoust. Soc. Am. **97**, 1865–1873.

McGarr, N. S., and Campbell, M. (**1995**). "Ephphatha: Opening inroads to understanding articulatory organization in persons with hearing impairment," in *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP Press, New York), pp. 433–404.

McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the spectral maxima sound processor (SMSP) and the mini speech processor (MSP)," Ear Hear. **14**, 350–367.

Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.

Perkell, J., Lane, H., Svirsky, M., and Webster, J. (**1992**). "Speech of cochlear implant patients: A longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2978.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard-of-hearing II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Siegel, G. M., Schork, E. J., Pick, H. L., and Garber, S. R. (**1982**). "Parameters of auditory feedback," J. Speech Hear. Res. **25**, 473–475.

Svirsky, M., Lane, H., Perkell, J., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.

Taguchi, G., and Jugulum, R. (**2002**). *The Mahalanobis-Taguchi Strategy : A Pattern Technology System* (Wiley, New York).

Tye, N., Zimmermann, G., and Kelso, S. (**1983**). " 'Compensatory articulation' in hearing-impaired speakers: A cinefluorographic study," J. Phonetics **11**, 101–115.

Wilson, B., Lawson, D., Zerbi, M., Finley, C., and Wolford, R. (**1995**). "New processing strategies in cochlear implantation," Am. J. Otol. **16**, 669–681.

Effects of speaking rate on the single/geminate stop distinction in Japanese[a)]

Yukari Hirata[b)] and Jacob Whiton
Department of East Asian Languages and Literatures, Colgate University, Hamilton, New York 13346

This study addressed an issue in the theory of acoustic invariance. The question was whether an invariant acoustic property exists for distinguishing Japanese single and geminate voiceless stops across rates and speakers. Four native Japanese speakers produced disyllabic words with single and geminate stops (e.g., /kako/ and /kak:o/) spoken in a carrier sentence at three speaking rates. Durations of sentences, words, stop closures, vowels preceding the contrasting stops, and voice onset times were measured. Ratios of geminate to single stop closures, geminate words to singleton words, closures to preceding vowels, and closures to words were calculated. The effect of rate on closure duration was to yield overlap between the singleton and geminate categories, and to lengthen geminate closures more than single closures as rate decreased. However, the ratio of geminate to single closure duration was unaffected by rate. Furthermore, the ratio of closure to word duration (0.35 as an optimal boundary) best classified all singleton and geminate tokens with 95.7%–98% accuracy. Thus, in spite of overlap in absolute closure duration between single and geminate voiceless stops, there is a relationally invariant measure that divides the two phonemic categories across rates and speakers, supporting the theory of relational acoustic invariance. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000807]

I. INTRODUCTION

The acoustic invariance theory claims that there is a direct mapping between phonetic features and acoustic signals that transcends languages, contexts, and speakers (Stevens and Blumstein, 1978; Lahiri *et al.*, 1984; Sussman *et al.*, 1991; Pickett *et al.*, 1999). The form of acoustic invariance has been found to be relational (Pind, 1999; Pickett *et al.*, 1999; Boucher, 2002). Pickett *et al.* found that, even though duration of Italian single and geminate stop closure overlapped across speaking rates, the ratio of stop closure to preceding vowel durations was a stable indicator of this distinction. In Hirata (2004), for comparing phonemic short and long vowels in Japanese disyllabic words, the ratio of vowel to word duration remained unchanged and the two categories were classified with high accuracy across speaking rates.

The present study examined whether there is an invariant acoustic measure that distinguishes Japanese single and geminate voiceless stops in disyllables spoken at three speaking rates by four speakers. Consonant length is phonemic in Japanese (Vance, 1987), as in many other languages (e.g., reviews in Pickett *et al.*, 1999, for Italian; Lahiri and Hankamer, 1988, for Bengali and Turkish; Hansen, 2004, for Persian; Dunn, 1993, for Finnish). Japanese consonant length is reflected in the number of moras of the words, e.g., /ita/ (single, two moras) "stayed" versus /it:a/ (geminate, three moras) "went," or /kako/ (single, two moras) "past" versus /kak:o/ (geminate, three moras) "parenthesis." The major acoustic correlate for the distinction of single and geminate

stops in Japanese is closure duration, similar to what has been found for Turkish and Bengali (Hankamer *et al.*, 1989; Lahiri and Hankamer, 1988), Persian (Hansen, 2004), and Italian (Pickett *et al.*, 1999). The closure duration for Japanese geminate stops has been found to be 2.6 to 3 times longer than the closure duration for single stops (Han, 1962; Homma, 1981; Beckman, 1982). The major perceptual cue to distinguish this contrast is the closure duration as well (Ohtsubo, 1981), and the closure duration boundary between the two categories changes linearly and positively with the preceding vowel duration (Watanabe and Hirato, 1985; Hirata, 1990).

Even though closure duration is known to be the major correlate for this contrast, little is known about how speaking rate affects this durational measure. In order to investigate the question of whether acoustic invariance exists, it is necessary to first characterize the effect of speaking rate. For example, if closure duration overlaps between single and geminate stops when they are spoken at different rates, closure duration would not be a good indicator for acoustic invariance. The ratio of single and geminate closure duration measured at one rate for Japanese, i.e., 1: 2.6–3 (Han, 1962; Homma, 1981; Beckman, 1982),[1] is greater than the ratios found for other languages, e.g., 1: 1.72 for Icelandic (Pind, 1999); 1: 2.11 for Italian (Pickett *et al.*, 1999); 1: 2.21 for Persian (Hansen, 2004); and 1: 2.28 for Finnish (Dunn, 1993). It could be that the difference between Japanese single and geminate closure duration is so great that there might not be any overlap, even across varying speaking rates. On the other hand, some studies (e.g., Hirata, 2004; Port *et al.*, 1980) suggest that closure duration might overlap between the two categories across rates.

If absolute duration values such as stop closure duration significantly overlap, it is still possible to look elsewhere (such as ratios) for invariant measures. The following durational measures, besides closure duration, were examined to test the acoustic invariance theory: ratio of single versus geminate closure (GC/SC), duration of the disyllabic words containing single and geminate stops (two- and three-mora words), ratio of three- to two-mora words (3M/2M), duration of voice-onset time (VOT), ratio of closure to preceding vowel duration (C/V), and ratio of closure to word duration (C/W). Below, we summarize previous findings on these measures.

As for the GC/SC ratio, Han (1962), Homma (1981), and Beckman (1982) found that the ratio of single versus geminate closure is 1: 2.6–3. However, this ratio was derived from only one speaking rate in each of these studies, and it is not known whether this ratio remains unchanged or changes with speaking rate variations. If this ratio is acoustically invariant, it would be expected not to vary with speaking rate changes.

Another measure that is predicted to reflect single and geminate distinction is duration of words. Port et al. (1987) found that word duration in Japanese increases linearly with the number of moras in a word. Homma (1981) and Han (1992 and 1994) found that duration of words that contain geminate stops is consistently longer than those that contain single stops and, in addition, that the ratio of the durations of disyllabic word pairs containing single and geminate stops was close to 2:3, reflecting the number of moras. However, those data too were measured at only one speaking rate. Given that a durational increment due to the addition of a mora was smaller for a fast than slow speaking rate in Port et al. (1987), it is possible that the ratio of two- and three-mora words is smaller for faster speech. The present study therefore examined how the word duration and the ratios of two- and three-mora words are affected by variations in speaking rate.

Homma (1981) found that VOT does not reflect the distinction of Japanese single and geminate stops. VOT for Japanese stops, categorized as middle-lag VOT (Vance, 1987), has been found to be 30 ms for /t/ and 66 ms for /k/ word-initially (Shimizu, 1996), and 15 ms for /t/ and 29 ms for /k/ word-medially spoken with carrier sentences (Sagisaka and Tohkura, 1984). Effects of speaking rate on these values of Japanese stops are not known, however, although they have been widely studied for English VOT (Miller et al., 1986; Kessinger and Blumstein, 1998; Boucher, 2002). The present study examined how Japanese word-medial VOT varies with the speaking rate, and whether it has any correlation with single and geminate distinction.

Acoustic invariance has been proposed to be defined in relational terms (Pickett et al., 1999). That is, the duration of one part of an utterance will have a consistent relationship with the duration of another part of the same utterance, leading to a constant ratio. For example, in Pickett et al. (1999), this constant ratio was that of closure to the preceding vowels (C/V) for Italian single/geminate distinction. In Pind (1995), ratios of vowel to the rhyme in Icelandic were the stable measure in the distinction of consonant and vowel

length in VC: versus V:C. On the other hand, Hansen (2004) found that C/V ratio was not a good indicator for the distinction of single and geminate stops in Persian. The present study examined the effect of speaking rate on the C/V ratio in Japanese single and geminate distinction.

The final measure we examined was C/W (closure to word) ratio. In Boucher (2002), the ratios of VOTs to the syllable in /ˈtæp/ and /ˈdæp/ was constant across different speaking rates. Similarly, in Hansen (2004), although C/V ratio was not a stable measure, the ratio of the closure to the syllable (i.e., a higher-level unit than the preceding vowel) was found to be a better measure for the distinction of Persian single and geminate consonants. In examining the distinction between short and long vowels in Japanese disyllabic words, Hirata (2004) found that the duration ratio of the vowel to the entire word remained stable across different speaking rates. As noted earlier, Japanese word duration reflects the number of moras and has been found to be a stable durational measure (Port et al., 1987; Hirata, 2004). For this reason, the present study examined the ratio of closure to word duration. The question was to what degree this measure was affected by variations in speaking rate.

After identifying effects of speaking rate on these measures described above, the next goal was to determine whether there is a single acoustic value that could accurately classify the two categories of single and geminate stops. For this goal, we adopted the method of finding optimal boundaries used by Miller et al. (1986) and Pind (1995). These studies sought to find a boundary value that classified the two members of a contrast with highest accuracy. In the present study, three measures were evaluated on their effectiveness in correctly classifying singleton and geminate tokens: closure duration, C/V ratio, and C/W ratio. If there is a significant amount of overlap between single and geminate closure durations across three rates, absolute closure duration would not classify the two categories accurately. If this is the case, C/V and C/W ratios would be more useful measures. If a single value such as C/V or C/W ratio can classify the two categories with high accuracy across rates and speakers, this would imply that relational acoustic invariance exists for the distinction of single and geminate stops in Japanese.

Finally, it is important to note that this study examined solely the acoustic properties of Japanese single and geminate stops and did not investigate how they were perceived. The acoustic invariance theory (Lahiri et al., 1984) has two claims: that invariance exists in acoustic signals, and that native speakers' perceptual system is sensitive to that invariant acoustic property. It would merit further inquiry to investigate how the perceptions of native speakers of Japanese correspond with the findings of this study.

II. METHOD

A. Participants

The participants consisted of four monolingual native speakers of Japanese. Two were male (M1 and M2) and two were female (W1 and W2). The men were from Saga and Niigata prefectures, respectively, and both women were from Fukushima prefecture. They reported that they speak stan-

dard Japanese in formal settings.[2] All four were students or employees at the University of Aizu and had a mean age of 20 (range: 18–26). None reported living outside of Japan for more than a year.

B. Materials and procedure

Materials in set 1 consisted of 6 pairs of nonwords and those in set 2 consisted of 30 pairs of real Japanese words (see the Appendix). In each pair, one word contained a voiceless single stop, and the other contained a voiceless geminate stop at the word-medial position.[3] All of the words were disyllabic, and each pair consisted of two or three moras. Initial consonants of the words were only /k/ and /t/ for set 1, but set 2 included a variety of initial consonants: /s/, /h/, /m/, /n/, /dʒ/, /j/, /w/, /b/, and /ɾ/, as well as /k/ and /t/ (six out of 30 pairs of words in set 2 had no initial consonant). Each word in set 1 contained two identical vowels (either /e/, /a/, or /o/), but vowels used in set 2 varied among the five Japanese vowels (/i/, /e/, /a/, /o/, and /u/). Participants read all nonwords in set 1 with the first vowel accented. For set 2, it was up to each participant where to put an accent on the real words (see footnote 2). The purpose of examining set 1 was to characterize the pattern of single and geminate distinction when other segments in the words were controlled. The purpose of examining set 2 was to find to what extent the results with controlled segments (set 1) generalized to words that had a variety of consonants and vowels that are known to have different inherent durations (Klatt, 1976; Sagisaka and Tohkura, 1984). The nonwords in set 1 were written in Japanese *katakana* orthography. The words in set 2 were written in a combination of *kanji, hiragana*, and *katakana*, depending upon their usual rendering in Japanese orthography.

The speakers read each word in a carrier sentence /sokowa _____ to jomimasu/ ("that is read _____"). With two-mora or three-mora words inserted, there was a total of 10 or 11 moras in the sentence. The participants were instructed to speak the sentence at "slow," "normal," and "fast" rates. The written criteria for producing these separate rates were derived from Port (1977), as they have been successful in soliciting three distinct rates (Port, 1977; Port *et al.*, 1980; Svastikula, 1986). A normal speaking rate was defined to be a "relaxed and comfortable" tempo. A slow speaking rate was defined as the "slowest tempo possible that the speaker could produce" while refraining from adding obvious breaks between words. Finally, a fast speaking rate was defined as "the fastest tempo possible without making an excessive amount of errors." This was the only rate instruction given to the speakers. Words and the carrier sentence were all written in a handout, and participants read them at their own pace while following the instruction on speaking rates. The participants practiced reading sentences at several rates before recording.

The list of words was read in the carrier sentence five times in set 1 and three times in set 2 for each of the three rates. The word order was random in each repetition. To avoid the effect of list intonation, two additional filler tokens were placed at the beginning and end of each repetition's word list. The speakers read each set first at the normal, then slow, then fast rates. This resulted in a total of 2880 tokens (set 1: 12 words×3 rates×5 repetitions×4 speakers; set 2: 60 words×3 rates×3 repetitions×4 speakers) that were recorded using a digital audiotape recorder (Sony TCD-D10PROII) with a stereo microphone in a sound booth. All speech materials were digitized at a 22.05-kHz sampling rate with an automatic 11-kHz low-pass filter setting and 16-bit quantization using the Kay Elemetrics' COMPUTER SPEECH LAB (CSL 4300B).

C. Analysis
1. Criteria for segmentation

The durations were measured from spectrograms and waveforms using the PRAAT (v. 4.0.24) speech analysis software (Boersma and Weenink, 2002). Measures taken were durations of sentences, words, closures of the intervocalic stops, vowels preceding the closure (not word-final vowels), and VOTs of the intervocalic stops. Duration of the sentence was measured as the beginning and end of high-frequency noise in the spectrogram corresponding with /s/ (the final /u/ was voiceless without exception, as would be normal for Japanese).

If the target word began with a /k/ or /t/, the beginning of the word was determined to be the left edge of the sharp burst of energy. In other words, the closure of /k/ and /t/ at the beginning of the word was not included as part of the word. This was because speakers occasionally inserted a pause between the preceding phase /sokowa/ and the target words, especially at the slow rate, and thus it was hard to make a consistent measurement. This method was the same as that used in Homma (1981). There was a concern that the words beginning with /k/ and /t/ might be measured shorter than the rest of the words on average, and this is addressed in footnote 7. For all other words, duration of the target words was measured from the abrupt termination of all visible formants for /a/ in the preceding carrier phrase /sokowa/ to the abrupt termination of all visible formants of the final vowel of the target word prior to the following carrier phrase /to jomimasu/.[4]

In most tokens, the beginning of the vowel preceding the stop closure was marked by the beginning of all formant frequencies and the beginning of a periodic waveform. This was more common in words beginning with /k/, /t/, /m/, /n/, and /ɾ/. The end of the first vowel was marked using the same criteria as the end of the word, the termination of all visible formants.[5]

The beginning of the stop closure was marked at the end of the first vowel. For the end of the closure, the beginning of the burst for /k/ or /t/ was marked. Occasionally some fricative noise or sporadic striations would occur prior to the burst. If this noise was contiguous enough to be considered part of the burst for /k/ or /t/, it was considered the end of the closure.
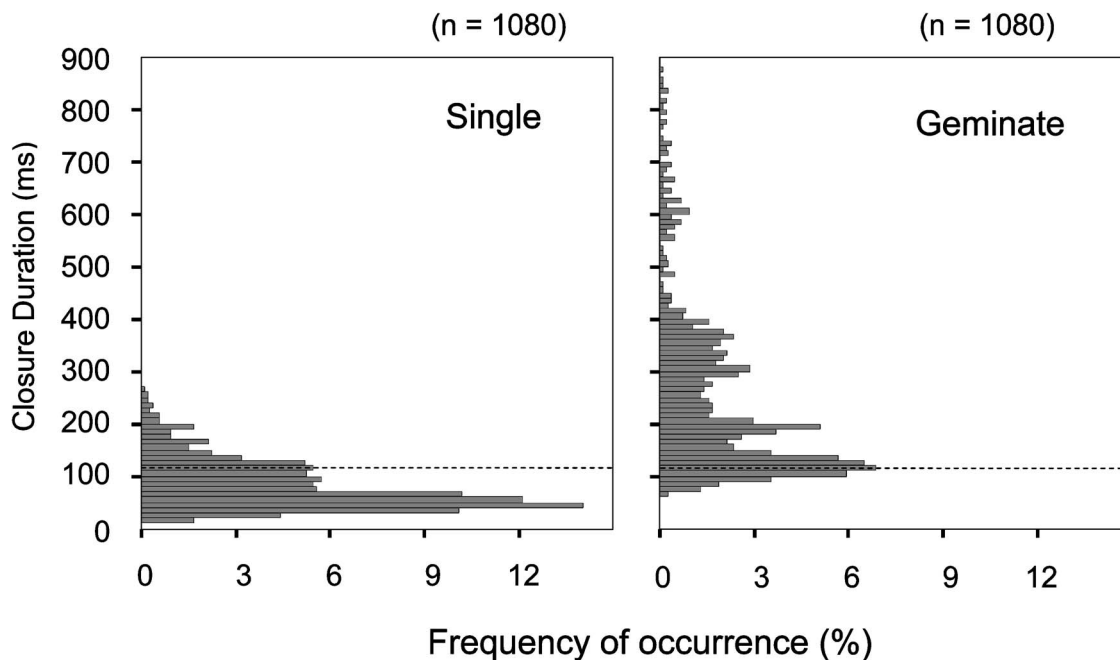
FIG. 1. Histogram of single versus geminate closure duration across all speakers and rates (set 2). Dashed line represents optimal boundary computed as 115 ms.

VOT began at the end of the closure. The end of VOT was defined as the beginning of the final vowel of the word, i.e., the beginning of the periodic waveform, which was the beginning of all formants for the vowel.

Some tokens proved to be impossible to label because of devoiced vowels, slurred pronunciation in fast tokens, or tape skips. In total, 20 such tokens (M1: 6; M2: 6; W1: 7; W2: 1) were removed out of 2880 total, or less than 1%.

The second author measured the entire data sets, and the first author confirmed in the 10% of the data that the measurement precisely followed the above criteria. The second author measured 5.6% of set 2 a second time to examine intrarater reliability, which was very high (Cronbach alpha =0.9992 for word duration; 0.9917 for vowels preceding closure; 0.9993 for closure duration).

2. Ratios

Besides the absolute duration measures, four ratios were compiled. Two ratios compared a singleton token with its corresponding geminate token. The first of this type was geminate closure duration divided by single closure duration (GC/SC). The second was the duration of three-mora words divided by the duration of two-mora words, with two-mora word duration standardized by two (3M/2M) (Han, 1994). For example, if word duration was 400 and 600 ms for two- and three-mora words, the value was calculated to be 600 $\times 2/400=3$; thus, the ratio would be 2:3. Two more ratios were measured within individual tokens: closure duration divided by preceding vowel duration (C/V) and closure duration divided by word duration (C/W).

3. Statistical analysis

In each stimulus set (sets 1 and 2), three-way repeated-measures analyses of variance (ANOVAs) were conducted separately for closure duration, word duration, VOT, C/V ratios, and C/W ratios as dependent variables. Rate (slow, normal, fast), quantity (single, geminate), and consonant (/k/, /t/) were within-subjects factors. In addition, two-way repeated-measures ANOVAs were conducted for ratios of single versus geminate closure durations (GC/SC) and for word duration ratios of two versus three mora words (3M/2M). Rate (slow, normal, fast) and consonant (/k/, /t/) were within-subjects factors. A separate ANOVA was conducted for sentence duration with rate (slow, normal, fast) and quantity (single, geminate) as within-subjects factors.

4. Optimal boundaries

Finally, optimal boundaries were determined separately for closure duration, C/V ratio, and C/W ratio in each stimulus set. Optimal boundaries for closure duration, for example, were determined by first creating histograms with 10-ms bins (as seen in Fig. 1) and determining the range of overlapping closure values between singleton and geminate categories (e.g., 60 through 270 ms in Fig. 1). Each bin was taken as a boundary, and the number of misclassified tokens was calculated. For example, if 270 ms is chosen, all singleton tokens are classified correctly but all geminate tokens less than 270 ms are misclassified. Using these results, the range of analysis was narrowed to where misclassified tokens were fewest in number (e.g., 110 and 120 ms in Fig. 1). This process was repeated within this range of analysis using smaller bins until a particular millisecond boundary with fewest misclassified tokens was found. The same procedure

Y. Hirata and J. Whiton: Japanese single/geminate distinction

TABLE I. ANOVA results for durations of sentence, stop closure, word, and VOT, and ratios of geminate to single closure (GC/SC), three- to two-mora words (3M/2M), closure to preceding vowel (C/V), and closure to word (C/W). R=Rate (slow, normal, fast); Q=quantity (single, geminate); Cons=consonant (/t/, /k/); R×Q =rate by quantity interaction; R×cons=rate by consonant interaction; Q×cons=quantity by consonant interaction. The measures, G/C and 3M/2M, were examined with only two factors (R and cons).

Measure	Set 1 (nonwords)				Set 2 (real words)			
	Effect	F	df	p	Effect	F	df	p
Sentence duration	R	56.52	(2,6)	0.000	R	25.34	(2,6)	0.001
	Q	190.66	(1,3)	0.001	Q	49.16	(1,3)	0.006
	R×Q	52.52	(2,6)	0.000	R×Q	33.69	(2,6)	0.001
Closure duration	R	35.81	(2,6)	0.000	R	20.27	(2,6)	0.002
	Q	130.85	(1,3)	0.001	Q	40.12	(1,3)	0.008
	R×Q	28.36	(2,6)	0.001	Cons	23.22	(1,3)	0.017
					R×Q	13.63	(2,6)	0.006
					R×cons	7.51	(2,6)	0.023
					Q×cons	33.76	(1,3)	0.010
GC/SC ratio					Cons	47.72	(1,3)	0.006
Word duration	R	54.68	(2,6)	0.000	R	24.59	(2,6)	0.001
	Q	135.05	(1,3)	0.001	Q	67.25	(1,3)	0.004
	R×Q	34.37	(2.6)	0.001	R×Q	24.93	(2,6)	0.001
3M/2M ratio	R	5.38	(2,6)	0.046	R	28.50	(2,6)	0.001
					Cons	17.71	(1,3)	0.025
VOT	R	7.93	(2,6)	0.021	Cons	20.51	(1,3)	0.020
	Cons	25.37	(1,3)	0.015	R×cons	8.92	(2,6)	0.016
	R×Q	12.84	(2,6)	0.007				
	R×cons	33.80	(2,6)	0.001				
C/V ratio	R	12.80	(2,6)	0.007	R	16.21	(2,6)	0.004
	Q	74.01	(1,3)	0.003	Q	56.76	(1,3)	0.005
	R×Q	7.54	(2,6)	0.023	R×Q	5.48	(2,6)	0.044
C/W ratio	R	17.46	(2,6)	0.003	R	71.33	(2,6)	0.000
	Q	254.28	(1,3)	0.001	Q	448.60	(1,3)	0.000
					Cons	27.77	(1,3)	0.013

was repeated for C/V and C/W ratios in both sets. The smallest bins were those of 1 millisecond for closure duration and of 0.01 for C/V and V/W ratios.

III. RESULTS

The results of ANOVAs for all durational measures are summarized in Table I, which shows significant main effects and interactions. Table II shows results of *post hoc* tests (Bonferroni) specifically for rate effects. Tables III, IV, and V summarize mean values of all measures for set 2.

A. Sentence duration

The purpose of examining sentence duration was to ensure that speakers spoke sentences at three distinct rates, as instructed. The results indicated that speakers produced three distinct speaking rates for both sets (see Table I). *Post hoc* tests indicated that the difference between any rates was significant except in set 2, where the difference between fast and normal rates was only marginally significant.

Sentence duration was affected by quantity, indicating that the duration of sentences including single stops was shorter than those including geminate stops (set 1: 2.022 versus 2.132 s; set 2: 1.939 versus 2.086 s). This is consistent with Port *et al.* (1987), who found that sentence duration correlates highly with the number of moras in the sentence. For both sets, a rate by quantity interaction was found. This interaction indicates that sentence duration difference between those including single and geminate stops was greater for slower speech (e.g., slow: 2.882 versus 3.127 s; normal: 1.757 versus 1.888 s; fast: 1.178 versus 1.243 s for set 2).

B. Closure duration

There was little overlap between single and geminate closure durations within each rate (set 1: fast: 1.7% overlap; normal and slow: no overlap; set 2: fast: 0.4% overlap; normal: 3.3% overlap; slow: 2.2% overlap). Our main interest here, however, is to examine whether there is a significant overlap across three rates (Fig. 1).[6]

TABLE II. Results of *post hoc* tests (Bonferroni) for rate effects. S, N, and F stand for slow, normal, and fast rates, respectively. Numbers in parentheses indicate the mean values.

$>^{**}$ $p \leqslant 0.01$
$>^{*}$ $p \leqslant 0.05$
$>$ $p < 0.1$
$,$ $p \geqslant 0.1$

	Set 1 (nonwords)	Set 2 (real words)
Sentence duration (s)	S(3.252)>*N(1.715)>*F(1.264) S>*F	S(3.005)>*N(1.822)>F(1.211) S>*F
Closure duration (ms)	S(269)>*N(108)>*F(71) S>*F	S(278)>N(140)>F(79) S>F
GC/SC ratio	No rate effect	No rate effect
Word duration (ms)	S(593)>*N(291)>*F(214) S>**F	S(625)>*N(363)>F(242) S>*F
3M/2M ratio	S (3.07), N (3.03), F (2.79) S,F	S(3.19)>*N(3.06)>F(2.82) S>*F
VOT (ms)	S(26)>*N(21), F (20) S,F	No rate effect
C/V ratio	S (2.18), N (1.55), F(1.34) S>*F	S (2.25), N (1.79), F (1.47) S>F
C/W ratio	S (0.43), N (0.35), F (0.31) S>F	S(0.41)>*N(0.36)>*F(0.31) S>**F

Figure 2 shows single and geminate closure durations for three rates for set 2, and similar results were obtained for set 1. Following the criterion used in Magen and Blumstein (1993), "a significant overlap" was said to exist when the mean geminate closure duration subtracting one standard deviation did not exceed the mean single closure duration adding one standard deviation. Single and geminate closure durations did not overlap within each rate. However, there was a significant overlap between single closures at a slow rate ($m=134$ ms; SD=34.7) and geminate closures at a fast rate ($m=116$ ms; SD=15.2). There was also a significant overlap between the slow single ($m=134$ ms; SD=34.7) and normal geminate closures ($m=210$ ms; SD=53.0). An identical overlapping pattern was found for set 1 (slow single: $m=140$ ms; SD=29.1; fast geminate: $m=105$ ms; SD=16.1; normal geminate: $m=165$ ms; SD=10.8).

For closure duration in both sets, a three-way ANOVA indicated a significant main effect for rate, indicating that closure duration was longer for slower rates (Table I). *Post hoc* tests indicated significant differences between any pair of rates (Table II) for set 1, and marginally significant differences for set 2. The weaker differences between the rates for set 2 are probably due to durational variations associated with a variety of segments in the word. The main effect of quantity was highly significant, indicating that geminate closure (set 1: 222 ms; set 2: 250 ms) was longer than single closure (set 1: 76 ms; set 2: 82 ms). The main effect of consonant was also significant, indicating that closure for /k/ (set 1: 148 ms; set 2: 164 ms) was shorter on average than closure for /t/ (set 1: 151 ms; set 2: 168 ms). The ANOVA also showed a significant rate by quantity interaction, indicating that rate changes affected geminate closure duration more than single closure duration. The difference between single and geminate closure durations was greater for slower speech in both sets (see Fig. 2 for set 2).

For only set 2, a rate by consonant interaction was significant, indicating that closure duration was greater for /t/ than for /k/, and that the difference (/t/ minus /k/ duration) was greater for slower speech (Table III). The values indicated that this difference was very slight (slow versus normal versus fast: 9 versus 2 versus 0 ms). A quantity by consonant interaction also reached significance. Single closure duration for /t/ was greater than /k/ (85 versus 78 ms), but geminate closure duration did not differ (250 versus 249 ms).
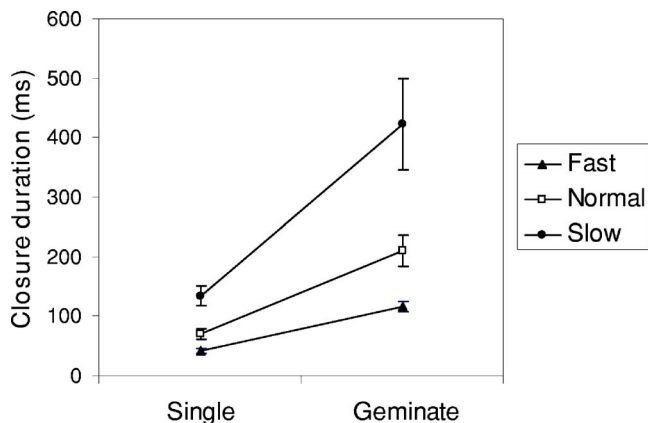
FIG. 2. Duration of single and geminate closures at three rates averaged across speakers (set 2). The error bars represent one standard error from the mean.

TABLE III. Mean duration (ms) of stop closure, word, and VOT for set 2. Values in parentheses indicate standard deviations.

			Slow		Normal		Fast	
			Mean	SD	Mean	SD	Mean	SD
Closure duration	Single	t	142	(33)	72	(20)	44	(10)
		k	127	(36)	67	(17)	40	(8)
	Geminate	t	426	(154)	210	(50)	115	(17)
		k	419	(151)	211	(56)	118	(14)
Word duration	Single	t	489	(119)	289	(52)	207	(16)
		k	480	(120)	289	(50)	198	(11)
	Geminate	t	762	(211)	437	(78)	282	(30)
		k	769	(201)	439	(79)	282	(23)
VOT	Single	t	19	(3)	19	(3)	23	(6)
		k	35	(6)	28	(3)	26	(4)
	Geminate	t	19	(4)	24	(16)	17	(3)
		k	36	(6)	26	(4)	22	(5)

C. Ratio of geminate to single closure duration (GC/SC)

Unlike absolute closure duration, the ratio of single versus geminate closure duration was not affected by rate. For example, the ratios for slow versus normal versus fast were 2.88 versus 3.36 versus 3.13 in set 1 and 3.18 versus 3.15 versus 3.06 in set 2. As shown in Table I, no effect of rate was found for either set. This shows that geminate closure was approximately three times as long as single closures across all rates. The only effect found was that of consonant in set 2 (Table IV). The closure ratio of /k/ was greater than that of /t/ (3.30 versus 2.96).

D. Word duration

The results of ANOVAs in sets 1 and 2 showed identical main effects and interactions. Rate affected word duration, and *post hoc* tests indicated significant differences between rates for sets 1 and 2 except in one comparison showing only marginal significance (set 2, normal versus fast). A significant main effect of quantity indicated that the word with single stops was shorter than those with geminate stops (Table III). The interaction of rate and quantity was also significant, similar to the previous (rate by quantity) interaction seen in closure duration. Rate affected geminate word durations more than singleton word durations, and the difference between singleton and geminate word durations was greater for slower speech in both sets (see Table III for set 2).

E. Ratio of three- to two-mora word duration (3M/2M)

The mean duration ratios of three- to two-mora words, with two-mora word duration standardized by two (see Sec. II C 2), were 2.96 for set 1 and 3.02 for set 2. This means that the durations of two-mora words versus three-mora words were close to 2:3, which is precisely the prediction made by Han (1994) and Port *et al.* (1987). In contrast to the closure duration ratio, the three- to two-mora word duration ratio was affected by rate. The ratio tended to be smaller (less than 2:3) when rate increased. The results of *post hoc* tests, however, suggest that the rate effect on the two- versus three-mora words was quite small in both sets (see Tables II and IV). Figure 3 shows the duration of two- and three-mora words plotted against each other across rates. The overall correlation between the two measures was high $[r=0.98, p<0.001]$. Most importantly, the regression line $(Y_{(2\ \text{mora})}=0.588 \times X_{(3\ \text{mora})}+33.7)$ (Fig. 3) showed a remarkable similarity to the hypothetical line based on the Port *et al.*'s (1987) mora theory: The word duration reflects the number of moras

TABLE IV. Mean ratios of geminate to single closure (GC/SC) and three- to two-mora words (3M/2M) for set 2. Values in parentheses indicate standard deviations.

		Slow		Normal		Fast	
		Mean	SD	Mean	SD	Mean	SD
GC/SC ratio	t	3.01	(0.35)	3.04	(0.31)	2.82	(0.59)
	k	3.36	(0.31)	3.25	(0.17)	3.30	(0.52)
3M/2M ratio	t	3.16	(0.09)	3.06	(0.03)	2.76	(0.13)
	k	3.23	(0.03)	3.07	(0.05)	2.87	(0.16)
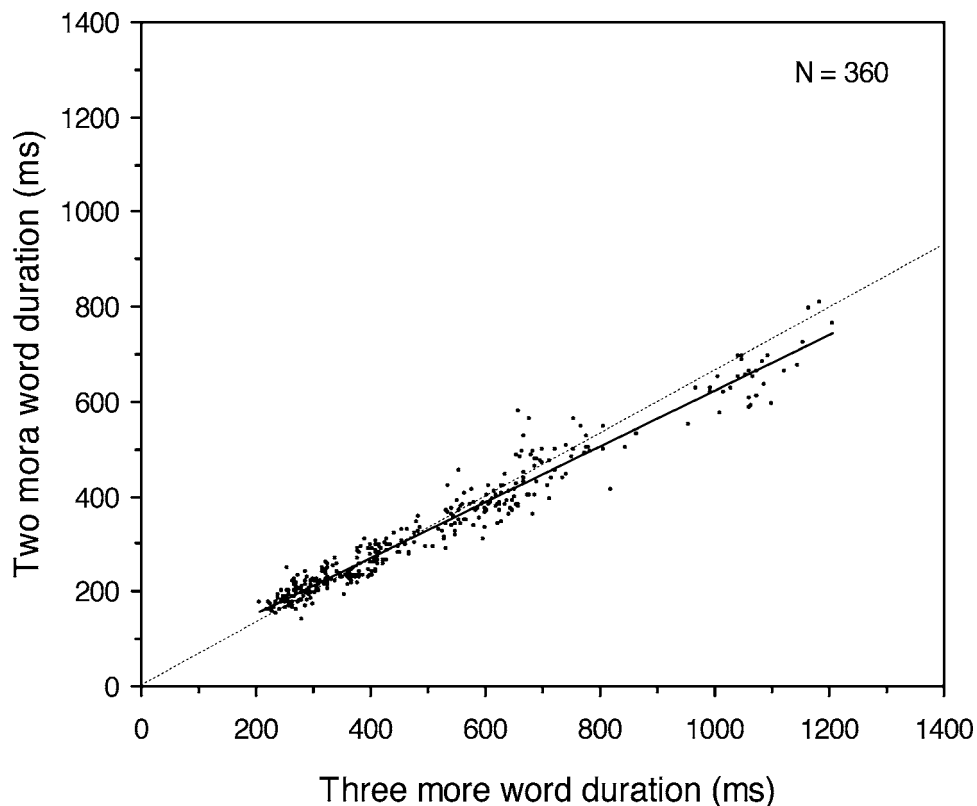
FIG. 3. Two- versus three-mora word duration across rates (set 2). Each data point represents the average of three repetitions produced by each speaker at each rate for each of 30 word pairs. The regression line (solid) ($Y_{(2\ mora)}=0.588\times X_{(3\ mora)}+33.7$) was close to the hypothetical line (dotted) ($Y_{(2\ mora)}=2/3 \times X_{(3\ mora)}$) predicted by Port *et al.* (1987).

it contains, i.e., two- versus three-mora words are 2:3 (thus, $Y_{(2\ mora)}=2/3\times X_{(3\ mora)}$).

F. VOT

A significant main effect was obtained for consonant in both sets. The mean VOT value was, on average, higher for /k/ than for /t/, consistent with Shimizu (1996) and Sagisaka and Tohkura (1984). For both sets VOT for /k/ was affected more by rate than VOT for /t/ (see Table III for set 2). An important finding in the measure of VOT was that no effect of quantity was found in both sets, indicating that VOTs for single and geminate stops did not differ. This result is consistent with Homma (1981).

The difference between the results of the two sets was that a main effect of rate was found significant for set 1. However, only the difference between VOTs at normal and slow rates was significant. Similar VOT values were found for set 2, but an effect of rate was not found. A significant rate by quantity interaction was also found for set 1 only. Single VOTs were greater than geminate VOTs for fast (22 versus 18 ms) and normal (22 versus 19 ms) rates, but they did not significantly differ at slow rate (26 versus 27 ms).

G. Ratio of closure to preceding vowel duration (C/V)

In order to examine the relationship between the durations of single and geminate closures and preceding vowels,

these two measures were plotted against each other. Figure 4 shows the results of set 2, and very similar results were found for set 1. Single closure duration increased linearly with the preceding vowel duration [set 1: $r=0.87$; $p<0.001$; set 2: $r=0.82$; $p<0.001$]. Geminate closure duration also increased linearly with the preceding vowel duration, but the correlation was weaker than that of the single closure [set 1: $r=0.85$; $p<0.001$; set 2: $r=0.76$; $p<0.001$]. Singleton and geminate tokens were separated by different regression lines: slopes were smaller for singleton (set 1: 1.27; set 2: 0.99) than for geminate (set 1: 3.39; set 2: 3.20), and intercepts were greater for singleton (set 1: −17.27; set 2: 3.75) than for geminate (set 1: −79.22; set 2: −47.60).

The ANOVAs revealed identical effects and interactions for the C/V ratio in sets 1 and 2. A significant main effect was obtained for rate, but *post hoc* tests showed that this effect was weak (Table II). A significant main effect for quantity indicated that the C/V ratio was smaller for singleton than geminate tokens (singleton versus geminate: 1.00 versus 2.38 for set 1; 1.06 versus 2.61 for set 2). As these values indicate, the ratios for singleton tokens were very similar in the two sets, while they differed slightly more for geminate tokens. Rate affected the C/V ratio in geminate words more than in singleton words (a significant rate by quantity interaction). The difference between singleton and geminate C/V ratios was greater for slower speech for both sets (e.g., Table V for Set 2).
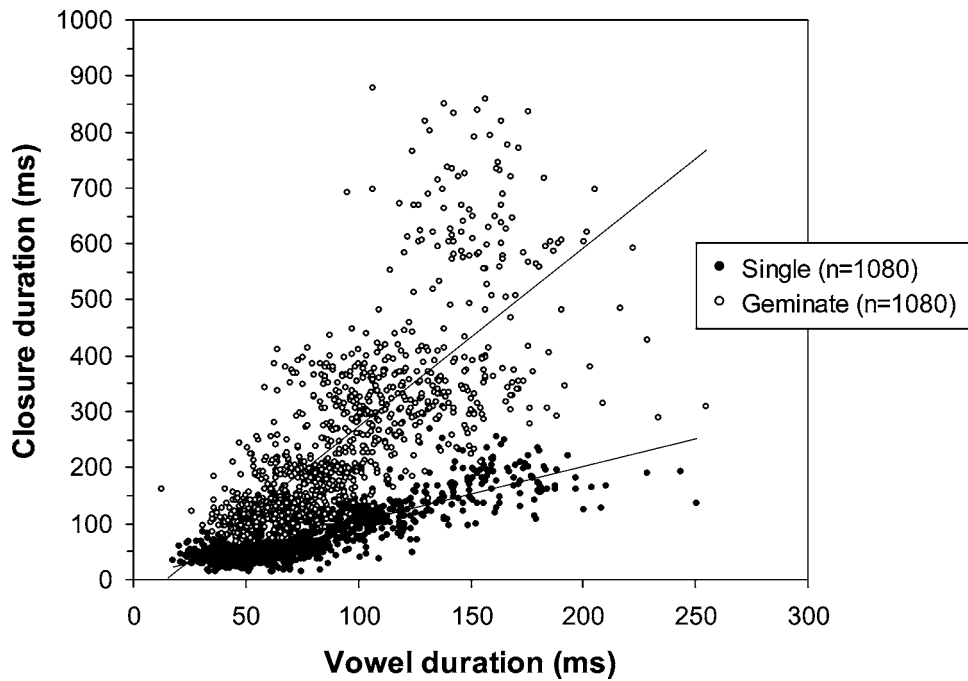
FIG. 4. Duration of single and geminate closures relative to preceding vowel duration (set 2).

H. Ratio of closure to word duration $(C/W)^{7}$

Figure 5 shows the duration of single and geminate closures relative to word duration across four speakers for set 2. The figure highlights two points. First, these two measures are linearly correlated across rates and speakers (although statistical analysis could not be performed for the linear relationship since the two measures are in a part–whole relationship). Second, even though the closure durations significantly overlapped between the singleton and geminate tokens (Fig. 1), they are separate when plotted against word duration.

A main effect of rate on the C/W ratio was significant in both sets. As the rate slowed, the C/W ratio became greater. This rate effect, however, did not result in substantial overlap between the two categories (Fig. 6; see the next section on optimal boundaries). A main effect of quantity was also significant in both sets. A notable result is that the mean C/W ratios for singleton and geminate tokens were identical in both sets: 0.24 versus 0.48. This indicates that the duration of the closure relative to the word duration is a stable measure in spite of the durational variations in segments comprising
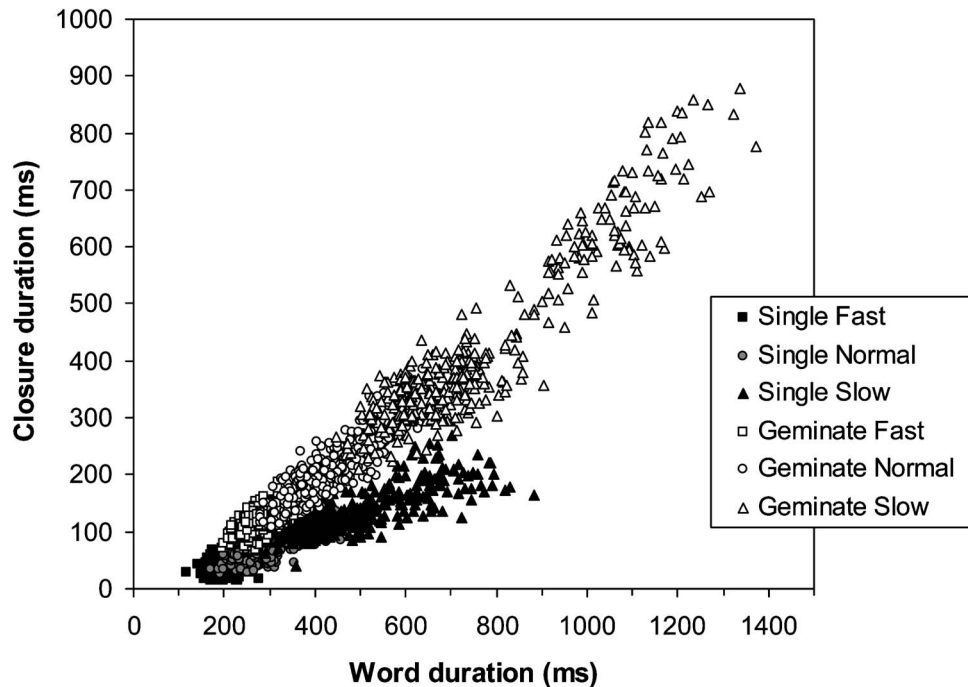
FIG. 5. Duration of single and geminate closures relative to word duration (set 2).

TABLE V. Ratios of closure to preceding vowel (C/V) and closure to word (C/W) for set 2. Values in parentheses indicate standard deviations.

			Slow		Normal		Fast	
			Mean	SD	Mean	SD	Mean	SD
C/V	Single	t	1.28	(0.19)	1.09	(0.23)	0.92	(0.13)
ratio		k	1.14	(0.07)	1.06	(0.12)	0.88	(0.13)
	Geminate	t	3.34	(0.83)	2.49	(0.34)	1.96	(0.14)
		k	3.25	(0.84)	2.53	(0.48)	2.10	(0.19)
C/W	Single	t	0.30	(0.01)	0.25	(0.03)	0.22	(0.04)
ratio		k	0.27	(0.01)	0.23	(0.03)	0.20	(0.03)
	Geminate	t	0.55	(0.05)	0.48	(0.03)	0.41	(0.02)
		k	0.54	(0.05)	0.48	(0.04)	0.42	(0.02)

words. A main effect of consonant was found only in set 2, although the difference was extremely small (/k/ versus /t/: 0.36 versus 0.37) (Table V).

I. Optimal boundaries

Optimal boundaries for closure duration, C/V ratios, and C/W ratios were calculated separately for sets 1 and 2 (Table VI). The purpose was to address the major question of the present study, i.e., to examine whether a single value can be found to classify singleton and geminate tokens accurately across rates and speakers. The optimal boundary for closure duration was found to be either 73 or 74 ms for set 1. With either of these boundaries, 17.8% of set 1 tokens were misclassified. For set 2, a different optimal boundary was found (115 ms) (see Fig. 1). With this boundary, 18.6% of set 2 tokens were misclassified.

Optimal boundaries for C/V ratios were found to accurately classify greater numbers of tokens than absolute closure duration. For set 1, 1.52 and 1.53 yielded the same number of misclassified tokens (7.9%). In set 2, 1.53 was found to be the optimal boundary, with 8.7% of tokens misclassified. It is interesting to note that an identical C/V ratio of 1.53 was found independently in the two sets, while different optimal boundaries were obtained from the two sets for closure durations.

The C/W ratios proved the most accurate at classifying singleton versus geminate tokens. An identical C/W ratio boundary of 0.35 misclassified only 2% and 4.3% of tokens in sets 1 and 2, respectively. All of the four speakers showed similar optimal boundaries, when calculated individually (For set 2, M1: 0.35; M2: 0.33; W1: 0.35; W2: 0.36). Figure 6 shows the ratio of closure to word duration plotted against
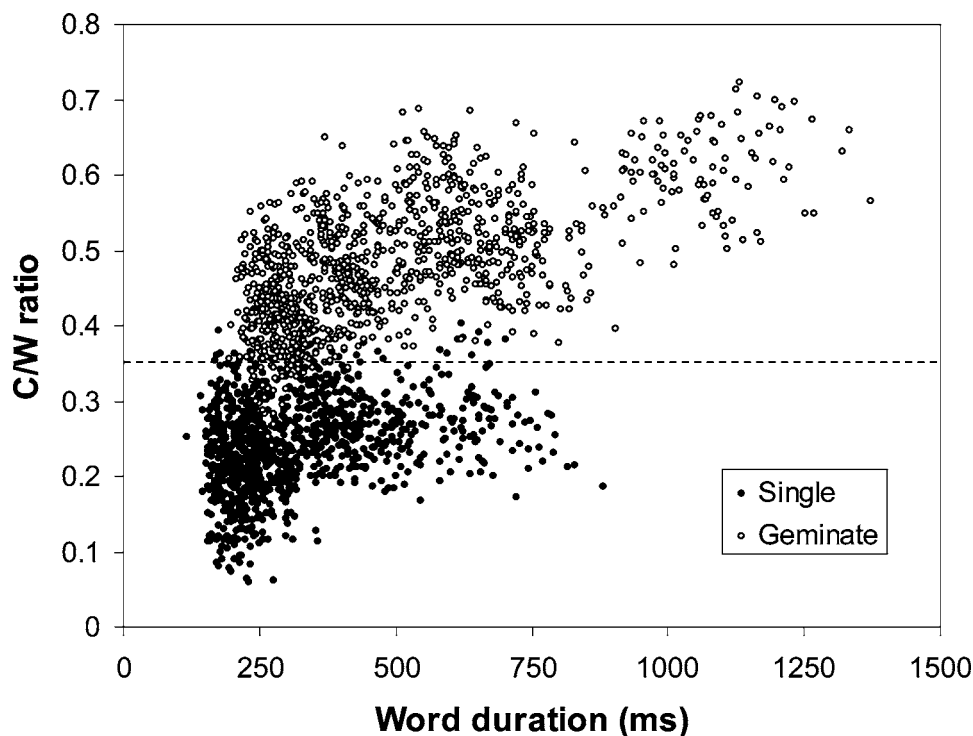
FIG. 6. Ratio of closure to word duration plotted against word duration (set 2). Dashed line represents optimal boundary computed as 0.35.

TABLE VI. Optimal boundaries for closure duration, C/V ratio, and C/W ratio. Set 1 includes 720 tokens, and set 2 includes 2160 tokens, both spoken by all four speakers.

		Optimal boundary	Number of misclassified tokens	Percentage misclassified	Classification accuracy
Closure duration	Set 1	73 or 74 ms	122 single 6 geminate	17.8%	82.2%
	Set 2	115 ms	235 single 167 geminate	18.6%	81.4%
C/V ratio	Set 1	1.52 or 1.53	25 single 32 geminate	7.9%	92.1%
	Set 2	1.53	114 single 74 geminate	8.7%	91.3%
C/W ratio	Set 1	0.35	8 single 7 geminate	2%	98%
	Set 2	0.35	46 single 47 geminate	4.3%	95.7%

word duration for set 2. The dashed line shows that the C/W optimal boundary of 0.35 clearly separates the singleton and geminate tokens in spite of the variations in speaking rate of the words. It is remarkable that the C/W boundary ratio of 0.35 was obtained separately from the two sets, classifying all tokens with very high accuracy (98% and 95.7%).

IV. DISCUSSION AND CONCLUSIONS

This study examined how rate affected various durational measures associated with single and geminate distinctions in Japanese. When singleton and geminate tokens were produced at three rates by four speakers, there was considerable overlap between the closure duration of the stops (Fig. 1). In particular, the duration of single stop closures spoken at the slow rate significantly overlapped with the duration of geminate stop closures spoken at both the fast and normal rates (Fig. 2). Rate effects were asymmetric (Port *et al.*, 1980; Pind, 1999), affecting geminate closure duration significantly more than single closure duration. As a result, the durational difference between single and geminate closures was greater for slower speech. In summary, the effect of rate on closure duration is to yield significant overlap between the two categories, and to lengthen geminate closures more than single closures as speaking rate decreases.

Similar findings were obtained for word duration. Word duration reflected the distinction between single and geminate stops: words including single stops were significantly shorter than words including geminate stops. This result is consistent with the previous findings by Port *et al.* (1987) and Han (1994). Rate affected the duration of both types of words, and the effect was greater for those including geminate stops. In contrast, VOT did not significantly differ between single and geminate stops, and this result is consistent with Homma (1981) and Beckman (1982). The effect of rate was very small, which contrasts with findings in English voiceless stops (Miller *et al.*, 1986; Kessinger and Blumstein, 1998; Boucher, 2002). This difference might be due to the fact that English has long-lag VOT as opposed to Japanese, which is said to have middle-lag VOT (Vance, 1987).

Small effects of rate were found on 3M/2M, C/V, and C/W ratios: the ratios tended to be greater for slower speech. However, no effect of rate was found for the GC/SC ratio. The ratio of single and geminate closure duration was 1: 2.88–3.36 (cf. 1: 2.6–3 from Homma, 1981; Beckman, 1982; Han, 1994), and this ratio was not systematically affected by variations in speaking rate. The correlation between the two- and three-mora word duration was 0.98 (Fig. 3), higher than the correlation between the preceding vowel and the stop closure (0.76–0.87, Fig. 4). Furthermore, the regression line for the two- and three-mora word duration was very close to the hypothetical line (assuming the ratio of 2:3), showing the close correspondence between the number of moras of the word and its duration, as predicted by Port *et al.* (1987).

The results of optimal boundaries (Table VI) addressed the question of whether a single value can be found to distinguish singleton and geminate tokens accurately. Different closure duration values were found in the two sets to optimally classify the two categories with the accuracy of 81%–82%. This accuracy was not as high as that for C/V and C/W ratios because of the significant overlap in closure duration between the rates, as discussed earlier. An identical C/V optimal boundary of 1.53 was found in both sets, with the classification accuracy higher than the closure duration optimal boundaries (91%–92%). The most reliable optimal boundary was that of the C/W ratio (Fig. 6). An identical optimal boundary of 0.35 was found in both sets with the accuracy of 95.7%–98%. This means that, in disyllabic words, if the closure duration is less than 35% of the word duration, the word almost certainly includes a single stop, and if the closure is greater than 35% of the word duration, then the word almost certainly includes a geminate stop. The materials included 6 pairs of words in set 1 and 30 pairs of words in set 2, both spoken by four speakers at three speaking rates ($n=720$ for set 1; 2160 for set 2). Set 1 consisted of words with controlled segments, but set 2 consisted of words with a variety of consonants differing in their duration. The result that the C/W boundary ratio accurately classified singleton and geminate tokens to such an extent across speakers, speaking rates, and word types strongly supports

the existence of durational invariance. The fact that word duration and closure duration each showed extreme durational regularity (Figs. 3 and 5) might explain why C/W was the best measure.

Throughout the results of the two sets of stimuli, there were more similarities than differences. For example, even though both absolute duration and ratio measures were affected by speaking rate, overlap between the singleton and geminate tokens was less for the ratio measures than the absolute duration measure, as suggested by the results of optimal boundaries (Table VI). Other striking similarities between the two sets were the identical optimal boundaries (C/V and C/W) as summarized earlier, and the mean C/W ratios for singleton and geminate tokens identical in the two sets (0.24 and 0.48). Taken together, durational variations due to different consonants and vowels were not an obstacle to finding a boundary that classifies the two consonant length categories correctly.

The results altogether support the view that the form of acoustic invariance is relational (Pickett *et al.*, 1999). It is *relational* in the sense that one part of an utterance relates to another part of the utterance in a predictable way that is not affected by different speaking rates and speakers. To clarify, C/W ratio does not have to be exactly 0.24 for all singleton tokens or 0.48 for all geminate tokens. The meaning of *invariance* here is that there is a clear upper boundary for single stops and lower boundary for geminate stops (0.35), beyond which the categories switch.

In Hirata (2004), in Japanese disyllabic words that included a long and short vowel (e.g., /kato/ versus /kato:/), the ratio of contrasting vowel to word duration (V/W ratio) of 0.38 optimally distinguished the two vowel length categories with highest accuracy. What would be the theoretical significance of this V/W ratio of 0.38 in Hirata (2004) and the C/W optimal boundary value of 0.35 found in the present study? The traditional mora theory for Japanese disyllables (Han, 1962; Kimura, 1985) claims that the word-medial single stop in CV.CV (2 moras) receives 0.5 mora, and that the geminate stop in CV.C.CV (3 moras) receives 1.5 moras because the first word-medial C is assigned one mora. This yields a closure duration ratio of 1:3, which the present data strongly support with no rate effect. The percentage of consonant closure duration in the total word duration is hypothetically 25% (0.5/2) and 50% (1.5/3) for two- and three-mora words, respectively, and the exact middle boundary can then be calculated as 37.5%. Our data provide values very similar to those in the model. Similar explanations can be given for two- and three-mora words including short and long vowels. There is a concern as to whether this mora counting works for words without initial consonants. The C/W ratio was greater for words without initial consonants in the present study (0.30 versus 0.23 for V.CV versus CV.CV; 0.54 versus 0.46 for V.C.CV versus CV.C.CV; both $p < 0.001$), but it is important to point out that this small difference does not invalidate the present optimal boundary of 0.35 (or the hypothetical boundary of 0.375).

Whether the C/W boundary ratio of 0.35 corresponds with Japanese native speakers' perception is an investigation that warrants further study. In Hirata (1990), the perceptual boundary was found to be 0.39 on average, but the stimuli were solely constructed from words that were originally spoken with single stops. It is therefore likely that this boundary was biased towards more responses for single stops (see also Fukui, 1978, for secondary cues to this distinction). Aoyama (2001) examined the native Japanese speakers' perception of single versus geminate nasals in /han(:)a/, and the perceptual boundary ratio of the nasals to the word fell between 0.34 and 0.37, which is consistent with the traditional mora theory. As mentioned in the Introduction, the acoustic invariance theory consists of two claims: one stating the presence of invariance in the acoustic signal, and the other stating that the perceptual system is sensitive to that invariance. Future perceptual study would be necessary to verify the second claim of this theory in relation to the present data. The major conclusion of this study, however, is that the present data support the first claim of the relational acoustic invariance: that there exists a relational acoustic invariance across speakers, rates, and word types in the distinction of single and geminate stops in Japanese disyllabic words.

Finally, one question that arises from the present study is why "word" (as in C/W) is a crucial unit for durational control in Japanese. According to Port *et al.*'s (1987) mora timing theory, sentence duration should also correlate highly with the number of moras it contains, implying the tendency for duration of any units (e.g., words, phrases, and sentences) to be correlated with the number of moras. This means that, for example, the present target words (two and three moras) plus one preceding and one following mora, i.e., a total of four and five moras, should also show durational regularity. That is, the ratio of stop closure to the duration of this unit should be 12.5% (0.5/4) and 30% (1.5/5), with the optimal boundary of 21.25%. Preliminary data (from five pairs of words by one speaker at three rates) suggest 11.9% and 25.0% for this unit including single and geminate stops, respectively, and thus, the boundary of 21.25% still seems to hold (though sentential rules, e.g., phrase-final lengthening, need to be incorporated into this). Similarly, when words contain more than two or three moras, the same timing model should apply. What all of these imply is that the present C/W ratio that clearly indicated the Japanese single/geminate distinction in the case of disyllables might be a microscopic aspect of a larger durational system such as the mora timing theory proposed by Port *et al.* (1987). There are studies showing that these model values are not acoustically real (Fukui, 1978; Beckman, 1982), and the issue of what constitutes a mora theory in Japanese is the subject of a continuing debate (Warner and Arai, 2001). Investigations of the durational structure of units larger than the present stimuli should shed more light on this issue.

ACKNOWLEDGMENTS

sion of this paper, and Anders Löfqvist, Robert Port, Keiichi Tajima, and an anonymous reviewer for their insightful reviews.

APPENDIX:

Set 1: Nonword stimuli

k	/kaka/	/kak:a/
	/keke/	/kek:e/
	/koko/	/kok:o/

t	/tata/	/tat:a/
	/tete/	/tet:e/
	/toto/	/tot:o/

Set 2: Real word stimuli

ka	/buka/	"subordinate"	/buk:a/	"price"
	/joka/	"free time"	/jok:a/	"four days"
	/dʒika/	"market price"	/dʒik:a/	"parents' house"
	/rika/	"science"	/rik:a/	"early summer"
	/nika/	"second chapter"	/nik:a/	"daily schedule"
	/haka/	"gravestone"	/hak:a/	"mint"
	/saka/	"hill"	/sak:a/	"writer"
	/ika/	"less than"	/ik:a/	"first chapter"
	/aka/	"red"	/ak:a/	"worsen"

ke	/jake/	"desperation"	/jak:e/	"windbreaker"
	/ake/	"open"	/ak:e/	"disbelief"
	/kake/	"bet"	/kak:e/	"beriberi"

ko	/hako/	"box"	/hak:o/	"eight (counter)"
	/neko/	"cat"	/nek:o/	"root"
	/kako/	"past"	/kak:o/	"parenthesis"

ta	/heta/	"unskilled"	/het:a/	"decreased"
	/buta/	"pig"	/but:a/	"hit (past tense)"
	/wata/	"cotton"	/wat:a/	"divided"
	/nita/	"boiled"	/nit:a/	(surname)
	/uta/	"song"	/ut:a/	"sold"

te	/mate/	"wait (imperative)"	/mat:e/	"wait"
	/sate/	"let me see…"	/sat:e/	"leave"
	/ute/	"hit (imperative)"	/ut:e/	"sell"
	/kate/	"win (imperative)"	/kat:e/	"selfish"
	/tate/	"vertical"	/tat:e/	"stand"

to	/soto/	"outside"	/sot:o/	"softly"
	/mito/	(place name)	/mit:o/	"catcher's mitt"
	/moto/	"formerly"	/mot:o/	"more"
	/oto/	"sound"	/ot:o/	"husband"
	/kato/	"transition"	/kat:o/	"cut"

[1]By calling the Japanese consonant length contrast "single" and "geminate" stops, we do not intend to imply any theoretical stance here. For example, the notion of geminate, meaning *double*, might be appropriate for languages that have the singleton/geminate duration ratio of 1:2, but might not be appropriate for languages that have the ratio of 1:3, as has been found for Japanese. This ratio of 1:3 has been taken as evidence for the existence of moras by Han (1962), and Port *et al.* (1987) have developed the mora timing theory in terms of word duration corresponding to the number of moras. The present study used the terms single and geminate [as in Han (1994) and Warner and Arai (2001)] in order to provide a point of comparison to other languages, and to follow the specific procedures taken to evaluate the acoustic invariance theory as in Pickett *et al.* (1999).

[2]The participants in the present study were not speakers of the standard Tokyo dialect in a strict sense as in many previous studies. Japanese has a pitch accent (Vance, 1987), and dialectal differences are known to yield variations in pitch accent of Japanese words (Sugito, 2000). However, Beckman (1982) reported that pitch accents do not play a major role in segmental duration in Japanese. Other studies indicate similar findings regarding effects of pitch accent on duration in Japanese (Hirata, 2004; Kori, 1997; Bradlow *et al.*, 1995). While we do not disagree that there are systematic effects of accent on duration (Mori, 2001; Hirata, 1990), the present study treated the accent of real words as a random variable. The object of the present study was to test whether there exists durational invariance despite dialectal differences that might yield some variations in accent of the words.

[3]The voiceless stop /p/ was not included in the stimulus sets because there are few disyllabic word pairs contrasting in the length of /p/ in Japanese.

[4]If a visible pause existed between /sokowa/ and the word, the following additional criteria were used. For words beginning with vowels, three speakers (M2 was the exception) inserted a glottal stop before the target word. In this case, the beginning of formants at all regions was marked as the beginning of the word. Occasionally separated striations (i.e., irregular voicing due to the glottal stop) were observed at the beginning of these vowels. If the striations appeared to be contiguous with the rest of the vowels in the spectrograms, they were included. For M2, when the preceding phrase and word were not separated by a glottal stop, the beginning of the abrupt change in formants for /a/ in /sokowa/ reaching to the target vowel (e.g., /i/ in /ika/) was determined to be the beginning of the word. For /aka/, the second formant ($F2$) remained steady and rose before the stop closure for /k/, and the boundary was hard to determine. Upon perceptual examination, the point where $F2$ began to ascend appeared to be the best location to separate the target from the preceding phrase. In words beginning with /h/, /s/, or /dʒ/, the beginning of the word was marked as the beginning of aperiodic noise on the waveform. If the word began with /m/, /n/, or /r/, the beginning of the low-frequency murmur with no other visible formants was marked as the beginning of the word. In words beginning with /w/, the first and second formants ($F1$ and $F2$) descended at the end of /sokowa/ and stayed in a low, steady state briefly before rising again. The word beginning was marked as the point where $F1$ and $F2$ began to rise. For /b/, many tokens had prevoicing, in which case the beginning of a low murmur was taken to be the start of the word. If the voicing and burst coincided, the beginning of the burst was marked as the beginning of the word. In the case of /j/, the third formant ($F3$) ascended from the preceding vowel /a/ and reached a steady state. The beginning was marked as the beginning of that steady state.

[5]In some words, the beginning of the vowel preceding the stop closure was difficult to determine and more detailed criteria proved necessary. In words beginning with /w/, the beginning of /a/ (which was the only vowel following /w/) was marked at the point where $F1$ and $F2$ became steady after rising at the beginning of the word. In words beginning with /j/, $F3$ moved from the higher frequency at the beginning of the word to lower frequency regions that became steady. The beginning of this steady state was marked as the beginning of the vowel. For words that started with aperiodic noise (those beginning with /h/, /s/, or /dʒ/), the beginning of a periodic waveform was marked as the beginning of the vowel.

[6]Some of the geminate tokens had extremely long closure duration, i.e., more than 500 ms. They were mostly from speaker W1 (88 tokens), only 3 tokens from speaker M2, and none from speakers M1 and W2. All of these tokens sounded natural to a native Japanese speaker, except that they did sound slow and sluggish.

[7]In the measurement of words beginning with voiceless stops (/t/ and /k/, 10 out of 60 words) (see the Appendix), we systematically excluded the word-initial closure duration, as in Homma (1981) with the same reason: random insertion of a pause before the target words. This was a limit of the present measurement method, and the /t/ and /k/ words might have been measured as slightly shorter than the rest of the words. As predicted, among words containing word-medial single stops (e.g., /kata/) in set 2, the /t/ and /k/ words were significantly shorter than the other words (298 versus 330 ms) [$p=0.004$]. However, among words containing word-medial geminate stops (e.g., /kat:o/), the duration of /t/ and /k/ words did not significantly differ from that of the other words (474 versus 500 ms) [$p=0.19$]. The consequence of having shorter duration for /t/ and /k/ words might be that the C/W ratio is greater for /t/ and /k/ words than the rest of the words in set 2. The mean C/W ratios were significantly greater for /t/ and /k/ words (single: 0.25; geminate: 0.50) than the rest (single: 0.24; geminate: 0.48) [single: $p=0.03$; geminate: $p<0.001$], but the differences were quite small.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Y. Hirata and J. Whiton: Japanese single/geminate distinction   1659

Along with other analyses we conducted, it was concluded that the effect of exclusion of the initial closure for /t/ and /k/ words (the limit of measurement method) was well within the range of durational variations associated with different segments (e.g., /s/, /h/, /j/, /w/, /m/, /n/, /b/, /dʒ/, and /ɾ/).

Aoyama, K. (**2001**). "Geminates and singletons: On 'unstretchability' of single segments," in *Proceedings of LP'2000*, edited by B. Palek and O. Fujimura (Karolinum, Prague, Czech), pp. 95–113.

Beckman, M. (**1982**). "Segment duration and the 'mora' in Japanese," Phonetica **39**, 113–135.

Boersma, P., and Weenink, D. (2002). "PRAAT: Doing Phonetics by Computer," Website: ⟨http://www.fon.hum.uva.nl/praat/⟩

Boucher, V. J. (**2002**). "Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates," Percept. Psychophys. **64**(1), 121–130.

Bradlow, A. R., Port, R. F., and Tajima, K. (**1995**). "The combined effects of prosodic variation on Japanese mora timing," Int. Congr. Phonetic Sci. **4**, 344–347.

Dunn, M. H. (1993). "The phonetics and phonology of geminate consonants: A production study," Doctoral dissertation (Yale University).

Fukui, S. (**1978**). "Perception for the Japanese stop consonants with reduced and extended durations," Bull. Phon. Soc. Japan **159**, 9–12.

Han, M. (**1962**). "The feature of duration in Japanese," Onsei no Kenkyuu **10**, 65–80.

Han, M. (**1992**). "The timing control of geminate and single stop consonants in Japanese: A challenge for non-native speakers," Phonetica **49**, 102–127.

Han, M. (**1994**). "Acoustic manifestations of mora timing in Japanese," J. Acoust. Soc. Am. **96**(1), 73–82.

Hankamer, J., Lahiri, A., and Koreman, J. (**1989**). "Perception of consonant length: Voiceless stops in Turkish and Bengal," J. Phonetics **17**, 283–298.

Hansen, B. B. (**2004**). "Production of Persian geminate stops: Effects of varying speaking rate," in *Proceedings of the 2003 Texas Linguistics Society Conference*, edited by A. Agwuele, W. Warren, and S. H. Park (Cascadilla Proceedings Project, Somerville, MA), pp. 86–95.

Hirata, Y. (**1990**). "Perception of geminated stops in Japanese word and sentence levels," Bull. Phon. Soc. Japan **194**, 23–28.

Hirata, Y. (**2004**). "Effects of speaking rate on the vowel length distinction in Japanese," J. Phonetics **32**, 565–589.

Homma, Y. (**1981**). "Durational relationship between Japanese stops and vowels," J. Phonetics **9**, 273–281.

Kessinger, R. H., and Blumstein, S. E. (**1998**). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phonetics **26**, 117–128.

Kimura, S. (**1985**). "*Nichi-eigo washa no choo-on mekanizumu: Nichi-eigo no heisaon no chooonjoo no sai to nihongo no rizumu*" (English translation: Stop consonants of English and Japanese, and the theoretical pattern of the typical rhythm in Japanese), Akita Daigaku Kyooiku Gakubu Kenkyuu Kiyoo Jinbun Kagaku Shakai Kagaku **35**, 17–32.

Klatt, D. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Kori, S. (**1997**). "*Nihongo no intonation*," in *Akusento, Intoneeshon, Rizumu to Poozu*, edited by M. Sugito (Sanseido, Tokyo).

Lahiri, A., and Hankamer, J. (**1988**). "The timing of geminate consonants," J. Phonetics **16**, 327–338.

Lahiri, A., Gewirth, L., and Blumstein, S. E. (**1984**). "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study," J. Acoust. Soc. Am. **76**(2), 391–404.

Magen, H. S., and Blumstein, S. E. (**1993**). "Effects of speaking rate on the vowel length distinction in Korean," J. Phonetics **21**, 387–409.

Miller, J. L., Green, K. P., and Reeves, A. (**1986**). "Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast," Phonetica **43**, 106–115.

Mori, Y. (**2001**). "Effects of accentual fall on final lengthening in Japanese," J. Phon. Soc. Japan **5**, 92–106.

Ohtsubo, K. (**1981**). "*Nihonjin no sokuon no umu no hanbetsu nooryoku ni tsuite*," Nagoya Daigaku Soogoo Gengo Sentaa Gengo Bunka Ronbunshuu **3**(1), 39–47.

Pickett, E. R., Blumstein, S. E., and Burton, M. W. (**1999**). "Effects of speaking rate on the singleton/geminate consonant contrast in Italian," Phonetica **56**, 135–157.

Pind, J. (**1995**). "Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues," Percept. Psychophys. **57**(3), 291–304.

Pind, J. (**1999**). "Speech segment durations and quantity in Icelandic," J. Acoust. Soc. Am. **106**(2), 1045–1053.

Port, R. F. (1977). "The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words," Doctoral dissertation (Indiana University Linguistics Club, Indiana University).

Port, R. F., Al-Ani, S., and Maeda, S. (**1980**). "Temporal compensation and universal phonetics," Phonetica **37**, 235–252.

Port, R. F., Dalby, J., and O'Dell, M. (**1987**). "Evidence for mora timing in Japanese," J. Acoust. Soc. Am. **81**(5), 1574–1585.

Sagisaka, Y., and Tohkura, Y. (**1984**). "Phoneme duration control for speech synthesis by rule," Trans. Inst. Electron. Commun. Eng. Jpn., Part A **J67-A**, 629–636.

Shimizu, K. (**1996**). *A Cross-Language Study of Voicing Contrasts of Stop Consonants in Six Asian Languages* (Seibido, Tokyo).

Stevens, K. H., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**(5), 1358–1368.

Sugito, M.(**2000**). *Onsei Hakei wa Kataru* (Izumi Shoin, Osaka, Japan).

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (**1991**). "An investigation of locus equations as a source of relational invariance for stop place categorization," J. Acoust. Soc. Am. **90**(3), 1309–1325.

Svastikula, M. L. K. (1986). "A Perceptual and Acoustic Study of the Effects of Speech Rate on Distinctive Vowel Length in Thai," Doctoral dissertation, University of Connecticut (ProQuest Information and Learning, Ann Arbor, MI).

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State Universtiy of New York Press, Albany, NY).

Warner, N., and Arai, T. (**2001**). "Japanese mora-timing: A review," Phonetica **58**, 1–25.

Watanabe, S., and Hirato, N. (**1985**). "*Nionsetsugo ni okeru musei haretsuon to sokuon no handan kyookai to senkoo boin no nagasa no kankei*," Onsei Gengo **1**, 1–8.

Acoustic characteristics of the vowel systems of six regional varieties of American English

Cynthia G. Clopper[a) and David B. Pisoni
Department of Psychology, Indiana University, Bloomington, Indiana 47405

Kenneth de Jong
Department of Linguistics, Indiana University, Bloomington, Indiana 47405

Previous research by speech scientists on the acoustic characteristics of American English vowel systems has typically focused on a single regional variety, despite decades of sociolinguistic research demonstrating the extent of regional phonological variation in the United States. In the present study, acoustic measures of duration and first and second formant frequencies were obtained from five repetitions of 11 different vowels produced by 48 talkers representing both genders and six regional varieties of American English. Results revealed consistent variation due to region of origin, particularly with respect to the production of low vowels and high back vowels. The Northern talkers produced shifted low vowels consistent with the Northern Cities Chain Shift, the Southern talkers produced fronted back vowels consistent with the Southern Vowel Shift, and the New England, Midland, and Western talkers produced the low back vowel merger. These findings indicate that the vowel systems of American English are better characterized in terms of the region of origin of the talkers than in terms of a single set of idealized acoustic-phonetic baselines of "General" American English and provide benchmark data for six regional varieties. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000774]

I. INTRODUCTION

[1]In what is now one of the most frequently cited papers in speech science and acoustic-phonetics, Peterson and Barney (1952) reported the results of an acoustic analysis of 10 vowels of American English produced by 33 men, 28 women, and 15 children in hVd utterances. When they plotted the first and second formant frequencies of the vowels in an $F1 \times F2$ space, they found a high degree of both within-speaker and cross-speaker variability in the production of vowel phonemes. While Peterson and Barney's data have since become important benchmarks in the characterization of American English vowels, the talkers in their study varied substantially in terms of native dialect and even native language. As a result, their data more closely approximate the speech of the American eastern seaboard than General American English. In addition, their data were collected over 50 years ago and cannot be expected to accurately reflect current pronunciation patterns, because American English, like all languages, continues to change over time (Bauer, 1985; Labov, 1994).

More recently, Hillenbrand, Getty, Clark, and Wheeler (1995) replicated the Peterson and Barney (1952) methods with male and female adult and child speakers from the northern Midwest in an attempt to control for some of the dialect-related problems in the original study. Hillenbrand *et al.* replicated Peterson and Barney's results with respect to cross-speaker variability, but they also observed a number of differences in the mean formant frequencies between their study and the earlier Peterson and Barney study. In particular, the authors found evidence of the Northern Cities Chain Shift (NCCS) in their speech samples. The NCCS is characterized by the clockwise rotation of the low and low-mid vowels; /æ/ is raised and fronted, /ɛ/ and /ʌ/ are backed, /ɔ/ is lowered, and /ɑ/ is lowered and fronted, as shown in Fig. 1 (Labov, 1998). The comparison of the Peterson and Barney mean formant frequencies with those from the Hillenbrand *et al.* talkers revealed that all but the most recent component of the NCCS, /ʌ/ backing, were present in both the male and female adult productions. The NCCS is the common speech pattern of residents in urban and suburban areas in upstate western New York and the northern Midwest, including Buffalo, Chicago, and Detroit (Labov, 1998). Because Hillenbrand *et al.* used native speakers from the northern Midwest, it is not surprising that their talkers would display this particular dialect variant of American English.

In a subsequent study, Hagiwara (1997) noted that the Hillenbrand *et al.* (1995) data were limited because they reflected only a single dialect of American English and he provided mean vowel formant frequency data for speakers from southern California as a supplement to the Hillenbrand *et al.* and Peterson and Barney (1952) datasets. The data Hagiwara described diverged from the earlier data reported by Hillenbrand *et al.*, particularly with respect to the high back vowels /u/ and /ʊ/, which were fronted, as is common in southern California speech (Thomas, 2001). Hagiwara concluded by making a plea for other researchers to collect and publish formant frequency data on vowels from other varieties of American English:

a)Now at the Department of Linguistics, Northwestern University, Evanston, Illinois 60208.

J. Acoust. Soc. Am. **118** (3), Pt. 1, September 2005   0001-4966/2005/118(3)/1661/16/$22.50   © 2005 Acoustical Society of America   1661
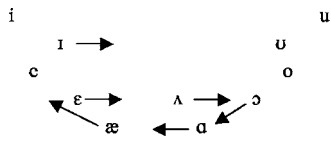
FIG. 1. Schematic of the Northern Cities Chain Shift.

Documentation of every American dialect on such a scale is obviously beyond the scope of any single researcher or research group… However, studies of a dozen or so speakers are well within the scope of most researchers, perhaps most students, and, in the absence of studies of large numbers of speakers, would be better than nothing. If the results of many such studies were combined, they would fill a significant void in objective descriptions of American English (p. 658).

American sociolinguists have in fact collected large corpora of vowel formant frequency data from talkers all over the United States and Canada (e.g., Labov, Ash, & Boberg, forthcoming; Thomas, 2001). However, their research interests have focused on descriptions of vowel variation in the United States (Thomas, 2001) or definitions of the regional varieties of American English (Labov *et al.*, forthcoming). The Labov *et al. Atlas of North American English*, for example, presents data in the form of maps showing where certain variants do and do not occur. Thomas, on the other hand, presented vowel spaces for individual talkers but made no claims about what we might expect the vowel system of the average speaker from a given location to look like. While answers to questions about the acoustic characteristics of American English vowel systems may be lurking in the work of sociolinguists, the results of their acoustic analyses are typically not presented with the familiar summary ellipses of Peterson and Barney (1952), but instead are presented on maps or in vowel spaces of individual talkers.

Thus, we are still lacking a comprehensive acoustic-phonetic description of the characteristics of the major regional varieties of American English. Hillenbrand *et al.* (1995) provided some benchmark vowel data for the northern Midwest and Hagiwara (1997) provided the same for southern California. However, the acoustic characteristics of "American English" vary based on where the speakers are from. The present study was designed to provide a summary and description of the acoustic characteristics of the vowel systems of six regional varieties of American English in an effort to fill this important gap in our understanding of how American English vowel systems are structured and how they differ with respect to one another. An underlying assumption of this research program is that synchronic differences between various regional dialects of American English can be defined relative to some standard or baseline variety. We will return to the issue of how to determine the baseline variety of American English in the General Discussion.

II. METHODS

A. Talkers

Forty-eight talkers between the ages of 18 and 25 were selected from the Nationwide Speech Project corpus (Clop-
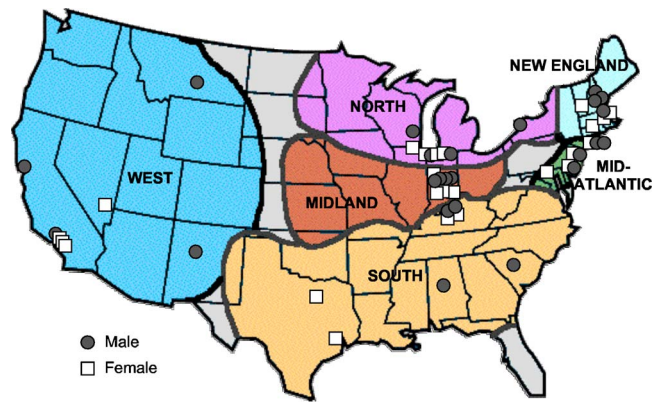
FIG. 2. (Color online) Map of the 6 dialect regions and the hometowns of the 48 talkers included in the acoustic analysis.

per, 2004) for use in the current study. All of the talkers were monolingual native speakers of American English with no history of hearing or speech disorders. Both parents of each talker were also native English speakers. The 48 talkers included four males and four females from each of six dialect regions of the United States: New England, Mid-Atlantic, North, Midland, South, and West. The six regions were based on the Labov *et al.* (forthcoming) descriptions of phonological variation in the United States. The map in Fig. 2 shows the geographic boundaries of the six dialect regions, as well as the hometown of each of the 48 talkers. Female talkers are indicated by light squares and male talkers are indicated by dark circles. Each talker had lived in a single dialect region for his or her entire life and both of his or her parents were also raised in that same dialect region. While all of the talkers were recorded at Indiana University in Bloomington, each talker had lived in Bloomington, Indiana for less than two years at the time of recording to reduce the effects of dialect leveling. Additional demographic information about the talkers is provided in the Appendix.

Based on previous research by sociolinguists (Labov, 1998; Labov *et al.*, forthcoming, Thomas, 2001), we predicted that the talkers in this study would produce dialect-specific variants of a number of the vowels we examined. As noted above and shown in Fig. 1, the vowel system of the Northern dialect of American English is characterized by the Northern Cities Chain Shift (Labov, 1998). The Southern dialect of American English, on the other hand, is characterized by the Southern Vowel Shift (Labov, 1998). The primary attribute of this shift is the fronting of the back vowels /u/ and /o/. In addition, the front lax vowels /ɪ/ and /ɛ/ are moving to the periphery of the vowel space through raising and fronting and the front tense vowels /i/ and /e/ are being centralized through lowering and backing. A schematic of the Southern Vowel Shift is shown in Fig. 3.
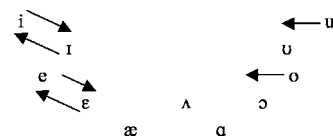
FIG. 3. Schematic of the Southern Vowel Shift.

TABLE I. Vowel tokens for acoustic analysis. Tokens followed by (5) were taken from the hVd utterances, for which five repetitions of each token were available. Tokens followed by (3) were taken from the sentence-length utterances, for which three repetitions of each token were available in sentence-final position.

Vowel	Tokens
i	heed (5)
ɪ	hid (5)
e	hayed (5)
ɛ	head (5)
æ	had (5)
ɑ	hod (5)
ɔ	frogs (3), logs (3)
ʌ	hud (5)
o	hoed (5)
ʊ	hood (5)
u	who'd (5)

The common feature of the Midland, Western, and New England dialects is the merger of the low-back vowels /ɑ/ and /ɔ/, creating homophones of such pairs of words as *caught* and *cot* or *Dawn* and *Don* (Labov, 1998). Other features of Western New England (eastern New York State, Vermont, and western Massachusetts) reflect several components of the Northern Cities Chain Shift with some raising of /æ/, fronting of /ɑ/, and backing of /ɛ/ (Boberg, 2001; Thomas, 2001). Western speech is characterized by the low back merger and by /u/ fronting (Labov *et al.*, forthcoming; Thomas, 2001). Unlike Southern back-vowel fronting, however, the Western pattern is typically limited to fronting of /u/. The Midland dialect is the least marked of the regional American English varieties, exhibiting no distinct features other than the ɑ∼ɔ merger. The Mid-Atlantic dialect does not exhibit the ɑ∼ɔ merger and, in fact, the two vowels are found to be more distinct due to /ɔ/ raising (Labov, 1994; Thomas, 2001). /æ/ also exhibits raising in some words, but not others, in the Mid-Atlantic region, due to a maintenance of a historical contrast between long and short /æ/ (Labov, 1994; Thomas, 2001).

B. Stimulus materials

A subset of the materials collected from each talker was used in an acoustic analysis of 11 vowels of American English: /i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u/, as shown in Table I. Five repetitions of each of the vowels except /ɔ/ were obtained from each talker in hVd utterances, for a total of 50 tokens per talker. In addition, six tokens of /ɔ/ were obtained in sentence-final position, three in the word *frogs* and three in the word *logs*, from each talker.[1] The resulting set of stimulus materials included 56 vowel tokens per talker with 5–6 tokens per vowel.

The stimulus materials were digitally recorded on a Macintosh Powerbook G3 laptop using a Shure SM10A head-mounted microphone, a microphone tube preamplifier, and a Roland UA-30 USB audio interface in a sound-attenuated chamber (IAC Audiometric Testing Room). Each utterance was recorded in an individual .aiff 16-bit sound file at a sampling rate of 44.1 kHz. Additional information about the collection of the stimulus materials can be found in Clopper (2004).

C. Procedures

Five acoustic measures were obtained from each of the 56 vowel tokens from each of the 48 talkers: vowel duration, first and second formant frequencies at the one-third temporal point in the vowel, and first and second formant frequencies at the two-thirds temporal point in the vowel, for a total of 13 440 measurements. All of the measurements were made using the Speech Analysis tool in WaveSurfer 1.6.2 (Sjölander and Beskow, 2004). The speech analysis tool included a time-aligned waveform, f0 trace, and wide-band spectrogram with formant tracks for F1, F2, F3, and F4. The automatic formant-tracking procedure was computed using a 12th-order LPC analysis over a 49 ms window with a 10 ms frame interval (Sjölander and Beskow, 2004).

For each token, the duration measurements were made first. The onset of the vowel was marked by the onset of voicing for those vowels preceded by a voiceless consonant and by a sudden change in intensity or formant frequency for those vowels preceded by a voiced consonant. The offset of the vowel was marked by the offset of voicing or a sudden drop in intensity, indicating closure. Particularly for those vowels following a liquid consonant, vowel onsets were determined by visual inspection of the waveform and spectrogram as well as by ear. Vowel duration was calculated as the difference between offset and onset of the vowel in milliseconds.

Formant values were automatically extracted from the formant traces at time stamps determined from the duration measures. The first-third and second-third time stamps were located at one-third and two-thirds of the vowel duration plus the time stamp of the onset. Formant measures were hand corrected by the first author by visual inspection of the wide-band spectrogram using the cursor tool as necessary.

A total of 28 vowel tokens (1%) were excluded because the talker misread the word (25 tokens) or a recording error occurred (3 tokens). All of the excluded tokens were from the hVd materials set in which trials with disfluencies or mispronunciations were not repeated.

The measurements were hand-checked for outliers prior to any further analysis. A total of 37 measurements out of the total 13 440 (0.3%) were rechecked as potential outliers. 1 was an error in measurement, 5 were typographic errors in data recording, and 14 were formant tracking errors not hand-corrected at the time of the original measurement. These measurements were all corrected prior to the analysis of the data. The remaining 17 potential outliers were found to be due to natural variation in the corpus and these data points were not altered.

A small subset (1%) of the tokens was also remeasured by the first author to assess reliability. The reliability subset included tokens from all six dialects and all 11 vowels. The absolute difference between the first and second duration measures ranged from 0 to 11 ms, with a mean difference of 3 ms. The absolute difference for the formant frequency
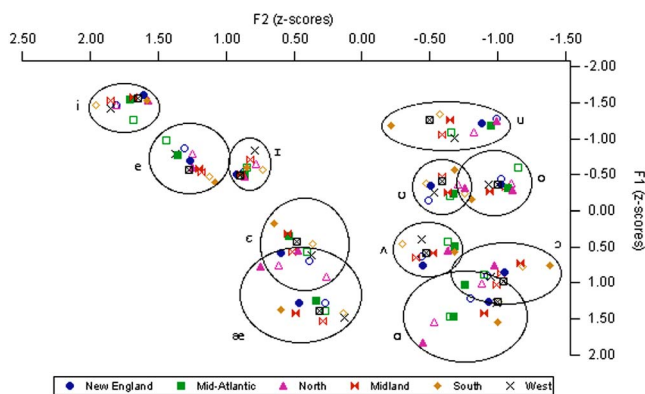
FIG. 4. (Color online) Mean *z*-score normalized formant frequency values for 11 American English vowels for each of the six dialects for the male talkers (filled symbols) and female talkers (open symbols). For the West, plain X's indicate female talkers and boxed X's indicate male talkers. The ellipses were hand-drawn to include every token for each vowel: /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/.

measures ranged from 0 to 44 Hz, with a mode of 0 Hz and a mean of 0.3 Hz. These differences are within the acceptable range of error for acoustic measures (Hillenbrand *et al.*, 1995). The reliability analysis and the outlier check suggest that the measurements were highly reliable.

III. RESULTS

A summary of the mean formant frequencies for the 11 American English vowels /i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u/ for each dialect is shown in Fig. 4. The filled symbols represent the means for the male talkers and the open symbols represent the means for the female talkers. For the Western talkers, the plain X's represent the female talkers and the boxed X's represent the male talkers. The means are based on the formant frequency measures taken at the first-third temporal point and were normalized using Lobanov's (1971) *z*-score transformation to reduce formant frequency variation due to anatomical differences between males and females.

Figure 4 clearly shows the Northern Cities Chain Shift for both the male and female Northern talkers (triangles), with the fronting and lowering of /ɑ/ in *hod* and the raising and fronting of /æ/ in *had* relative to the other dialects. In addition, the male and female Southern talkers (diamonds) show fronting of /u/ in *who'd* and /o/ in *hoed*. The Midland and Western talkers also show some fronting of /u/ in *who'd*. Finally, the New England males and females, Midland females, and Western males and females appear to have a merger or partial merger of /ɑ/ and /ɔ/ in *hod* and *frogs/logs*. In order to quantitatively assess differences in formant frequency due to talker dialect, a series of statistical analyses was conducted on the acoustic measurements.

A. Statistical analysis

Prior to the analysis, all of the formant frequency measures were normalized using Lobanov's (1971) *z*-score metric for each talker. The *z*-score method was selected because a recent comparison of vowel normalization metrics by

Adank, Smits, and van Hout (2004) showed that Lobanov's metric was the best transformation for reducing anatomical variation while still maintaining phonological and sociolinguistic variation. The *z*-score transform is a vowel-extrinsic, talker-intrinsic normalization procedure that centers each talker's vowel space on the origin in an F1×F2 plane. The equation in (1) was used to calculate the *z*-score normalization for each talker for each vowel token. The normalization was computed separately for F1 and F2. In the equation, *z* is the transformed normalized frequency value of F1 or F2 for a given vowel token, *f* is the raw frequency in Hz of that formant value, μ is the overall mean frequency of the relevant formant frequency (F1 or F2), and σ is the standard deviation of the overall mean of the same formant. The means and standard deviations are calculated individually for each talker based on all of the tokens produced by that talker. Given this talker-intrinsic nature of the transform, we expected that vowel-specific within-dialect gender differences would be maintained while overall gender differences in formant frequency would be greatly reduced. The duration measures were not normalized prior to statistical analysis.

$$z = (f - \mu)/\sigma. \tag{1}$$

A repeated measures ANOVA was calculated using vowel category (i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, or u) as a within-subjects factor and dialect (New England, Mid-Atlantic, North, Midland, South, or West) and gender (male, female) as between-subjects factors for each of three measures: vowel duration, F1 *z* score (at the first-third temporal point), and F2 *z* score (at the first-third temporal point). Because three analyses were computed, the *p* value was set to 0.01 for each ANOVA.

1. Duration

The repeated measures ANOVA on vowel duration revealed a significant main effect of vowel [$F(10,2020)$ =363.3, $p<0.001$], a significant main effect of dialect [$F(5,2020)=3.8$, $p=0.003$], a significant vowel X dialect interaction [$F(50,2020)=3.0$, $p<0.001$], and a significant vowel X gender interaction [$F(10,2020)=3.5$, $p<0.001$]. The main effect of gender, the dialect X gender interaction, and the vowel X dialect X gender interaction were not significant. The significant main effect of vowel category merely confirms that American English vowels differ in their inherent length and no further analyses on that factor were conducted. The significant main effect of dialect suggests that some dialects have longer or shorter overall vowels than others. Post-hoc Tukey tests on dialect revealed significant differences between the Southern talkers and the New England, Mid-Atlantic, and Western talkers based on vowel duration (all $p<0.01$). Overall, the Southerners had significantly longer vowels than the talkers from New England, the Mid-Atlantic, and the West.

The vowel X dialect interaction suggests that the effects of dialect differences on vowel duration are not consistent across all vowels. To explore this interaction, a one-way ANOVA on vowel duration with dialect as the factor was computed for each of the 11 vowels. To correct for the large number of analyses, the *p* value for the ANOVAs and the

Clopper *et al.*: Regional American English vowel systems

post-hoc Tukey tests was set to 0.001. Significant main effects of dialect were found for /ɪ/ $[F(5,234)=7.6, p<0.001]$, /ɛ/ $[F(5,228)=8.2, p<0.001]$, /ʌ/ $[F(5,229)=9.3, p<0.001]$, and /ʊ/ $[F(5,229)=8.3, p<0.001]$. In all four cases, post-hoc Tukey tests revealed that the vowels produced by the Southern talkers were significantly longer than the vowels produced by the New England, Mid-Atlantic, and Western talkers (all $p<0.001$). The vowel /ɛ/ was also longer for Southerners than Northerners and the vowel /ʌ/ was longer for Southern talkers than for Northern and Midland talkers (all $p<0.001$). These results suggest that Southerners did not produce generally longer vowels or have an overall slower speaking rate (as indicated by longer vowels), but that the vowel duration differences based on dialect were selective in nature and were due to longer lax vowels for Southern talkers than for the other dialect groups. That is, the durational distinction between lax and tense vowels was reduced for the Southern talkers.

The vowel X gender interaction suggests that while there was no overall effect on vowel duration due to gender, the male and female talkers did produce significant duration differences for some of the vowels. A series of t tests was conducted to compare the male and female talkers, collapsed across dialect, for each of the 11 vowels. To correct for the large number of analyses, the p value for the t tests was set to 0.001. Significant differences due to gender were found for /ɪ/, /ɛ/, /ɑ/, /ʌ/, and /ʊ/ (all $p<0.001$). In all five cases, the females produced longer vowels than the males. These results parallel those for the Southern talkers and reveal that the female talkers produced longer lax vowels than the male talkers, although they did not produce longer vowels overall.

2. Nucleus formant frequencies

The repeated measures ANOVA on the F1 and F2 z scores revealed a significant main effect of vowel $[F(10,2020)=3039.5, p<0.001$ for F1 and $F(10,2020)=4575.9, p<0.001$ for F2]. These results merely confirm the existence of significant differences in formant frequencies due to vowel category, and will not be analyzed further. Both the F1 and the F2 analyses also revealed a significant vowel X dialect interaction $[F(50,2020)=10.1, p<0.001$ for F1 and $F(50,2020)=15.5, p<0.001$ for F2], a significant vowel X gender interaction $[F(10,2020)=12.6, p<0.001$ for F1 and $F(10,2020)=16.6, p<0.001$ for F2], and a significant vowel X dialect X gender interaction $[F(50,2020)=2.5, p<0.001$ for F1 and $F(50,2020)=3.3, p<0.001$ for F2]. The main effects of dialect and gender and the dialect X gender interaction were not significant for either the F1 or F2 z scores. These findings suggest that the vowel spaces of the different dialects were not globally shifted along either F1 or F2, but that individual vowels were affected differentially by the six dialects. In addition, the z-score transform was successful in eliminating overall gender-related differences in formant frequency, but some vowel-specific differences in gender were retained, leading to the significant three-way interaction.

To explore the three-way interaction in more detail, post-hoc repeated measures ANOVAs were computed for the males and the females separately on the F1 and F2 z scores

with vowel category (i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, or u) as a within-subjects factor and dialect (New England, Mid-Atlantic, North, Midland, South, or West) as a between-subjects factor. The p value was again set at 0.001 to correct for the large number of post-hoc analyses. The repeated measures ANOVAs on male and female F1 and F2 z scores revealed significant main effects of vowel $[F(10,970)=2486.4, p<0.001$ for the male F1, $F(10,970)=2416.1, p<0.001$ for the male F2, $F(10,1050)=1143.3, p<0.001$ for the female F1, and $F(10,1050)=2219.3, p<0.001$ for the female F2]. The results again confirm the existence of significant differences in formant frequencies due to vowel category and will not be analyzed further.

Significant vowel X dialect interactions were also observed for both the male and female F1 and F2 z scores $[F(50,970)=8.3, p<0.001$ for male F1, $F(50,970)=11.5, p<0.001$ for male F2, $F(50,1050)=5.6, p<0.001$ for female F1, and $F(50,1050)=7.7, p<0.001$ for female F2]. The main effect of dialect was not significant in any of the post-hoc analyses. These results again suggest that the vowel spaces of the different dialects were not globally shifted along either F1 or F2, but that individual vowels are affected differentially by the six dialects. Post-hoc one-way ANOVAs on F1 and F2 for each gender with dialect as the factor were computed for each of the 11 vowels in order to explore the vowel X dialect interactions. The p value was again set at 0.001 to correct for the large number of post-hoc analyses. A summary of the results of the statistical analysis is provided in Table II.

For males for F1, significant effects of dialect were found for /e/ $[F(5,114)=8.9, p<0.001]$, /ɛ/ $[F(5,109)=14.4, p<0.001]$, /æ/ $[F(5,113)=16.3, p<0.001]$, /ɑ/ $[F(5,108)=14.7, p<0.001]$, and /ʊ/ $[F(5,113)=12.0, p<0.001]$. Post-hoc Tukey tests revealed that Southern male talkers produced significantly lower /e/s (higher F1) than the New England and Mid-Atlantic male talkers and significantly higher /ɛ/s (lower F1) than the New England, Northern, and Western male talkers (all $p<0.001$). These results are consistent with the Southern Vowel Shift. The Midland male talkers also produced significantly higher /ɛ/s than the New England and Northern talkers (both $p<0.001$), suggesting that the Southern pattern may be spreading to the Midland dialect region. The post-hoc Tukey tests also revealed significant /æ/ raising for the Northern male talkers compared to all of the other male talkers and significant /ɑ/ lowering for the Northern male talkers compared to the New England, Mid-Atlantic, Midland, and Western male talkers (all $p<0.001$). These findings reflect the Northern Cities Chain Shift. Finally, /ʊ/ was produced significantly higher by the Southern male talkers than the Mid-Atlantic, Northern, and Midland talkers (all $p<0.001$). This result was not anticipated based on the previous research on the Southern Vowel Shift, and appears to reflect back-vowel raising among Southern talkers.

For F2 for the males, the one-way ANOVAs revealed significant main effects of dialect for /e/ $[F(5,114)=7.6, p<0.001]$, /æ/ $[F(5,113)=16.1, p<0.001]$, /ɑ/ $[F(5,108)=38.7, p<0.001]$, /ɔ/ $[F(5,138)=24.8, p<0.001]$, /ʌ/ $[F(5,110)=10.9, p<0.001]$, /o/ $[F(5,113)$

TABLE II. Summary of the results of the acoustic analysis. Key: New England (E), Mid-Atlantic (A), North (N), Midland (M), South (S), West (W), Southern Vowel Shift (SVS), Northern Cities Chain Shift (NCCS), z-score transformation artifact (z-score), new Midland results (Midland), new Mid-Atlantic results (Mid-Atlantic), new Southern results (Southern).

Vowel	F1 Male	F2 Male	F1 Female	F2 Female	Results
i	n.s.	n.s.	n.s.	n.s.	
ɪ	n.s.	n.s.	n.s.	n.s.	
e	S > E, A	S < A	S > A	S < A, W	SVS
				M < A	Midland
ɛ	S < E, N, W	n.s.			SVS
			N > A, M, S, W	N < M	NCCS
	M < E, N				Midland
æ	N < ALL	N > ALL	N < ALL	N > ALL	NCCS
		S > A, W			Southern
ɑ	N > E, A, M, W	N > ALL	N > M	N > M, S, W	NCCS
		A > E, M, S, W		A > M, S, W	Mid-Atlantic
ɔ	n.s.	A > E, M, S, W	n.s.	n.s.	Mid-Atlantic
		S < E, A, M, W			z score
ʌ	n.s.		n.s.	N < E, M, S, W	NCCS
		A < E		A < M, S	Mid-Atlantic
		S < E, M, W			z score
o	n.s.	S > N	n.s.	S > A, N	SVS
ʊ	S < A, N, M	n.s.	n.s.	n.s.	Southern
u	n.s.	S > E, A, N		n.s.	SVS
		W > N			West
			S < A, N, M, W		Southern
			E < W		Spurious

=5.1, $p < 0.001$], and /u/ [$F(5,114) = 16.7$, $p < 0.001$]. Post-hoc Tukey tests revealed significant differences between the Southern and Mid-Atlantic male talkers in /e/ fronting ($p < 0.001$), consistent with the centralization of /e/ in the Southern Vowel Shift. The Northern male talkers produced significantly fronted /æ/s and /ɑ/s compared to all of the other male talkers (all $p < 0.001$), consistent with the Northern Cities Chain Shift. The Mid-Atlantic male talkers also produced significantly fronted /ɑ/s and /ɔ/s compared to the New England, Midland, Southern, and Western male talkers and significantly backed /ʌ/s compared to the New England male talkers (all $p < 0.001$). These results were not expected based on the previous literature, but seem to suggest an alignment of the low back vowels with respect to F2 in the Mid-Atlantic dialect, as shown in Fig. 4. The post-hoc Tukey tests also revealed significant /o/ and /u/ fronting for the Southern male talkers compared to the Northern male talkers. The Southern male talkers also produced significantly fronted /u/s compared to the New England and Mid-Atlantic male talkers (all $p < 0.001$). These results are consistent with the Southern Vowel Shift. The Western male talkers also produced significantly more fronted /u/s than the Northern talkers ($p < 0.001$), as predicted.

Finally, the Southern male talkers produced significantly fronted /æ/s compared to the Mid-Atlantic and Western male talkers and significantly backed /ʌ/s and /ɔ/s compared to the New England, Mid-Atlantic (/ɔ/ only), Midland, and Western male talkers (all $p < 0.001$). The fronting of /æ/ among Southern talkers was unexpected and may reflect a new shift

in Southern American English. However, the apparent backing of /ʌ/ and /ɔ/ for the Southern male talkers is probably an artifact of the z-score transform. Due to the significant back-vowel fronting produced by the Southern talkers, the top part of the vowel space for these talkers is narrower than for talkers with less extreme back-vowel fronting. In terms of the z-score transformation, the overall mean second formant frequency is higher and the second formant frequency standard deviation is smaller for talkers with fronted back vowels than for talkers without. The low back vowels normalized relative to this higher mean and smaller standard deviation appear artificially backed as the result of a larger numerator and smaller denominator in the z-score calculation. The result of the transform is therefore artificially backed low back vowels for the Southern male talkers.

Taken together, the results for the male talkers revealed components of the Northern Cities Chain Shift among the Northern male talkers, components of the Southern Vowel Shift among the Southern male talkers, Western /u/ fronting, some aspects of the Southern Vowel Shift among the Midland male talkers, and a new alignment of the low back vowels among the Mid-Atlantic males.

For the female talkers for F1, significant effects of dialect were revealed in the one-way ANOVAs for /e/ [$F(5,114) = 7.1$, $p < 0.001$], /ɛ/ [$F(5,113) = 10.0$, $p < 0.001$], /æ/ [$F(5,114) = 12.7$, $p < 0.001$], /ɑ/ [$F(5,110) = 4.9$, $p < 0.001$], and /u/ [$F(5,114) = 10.2$, $p < 0.001$]. Post-hoc Tukey tests revealed overall patterns similar to those found for the male talkers. In particular, the Southern female

talkers produced significantly lower /e/s than the Mid-Atlantic females ($p < 0.001$), consistent with the Southern Vowel Shift. The Northern females also produced lower /ɛ/s than the Mid-Atlantic, Midland, Southern, and Western females, as well as significantly higher /æ/s than all of the other female talkers, and significantly lower /ɑ/s than the Midland female talkers (all $p < 0.001$). These results are all consistent with the Northern Cities Chain Shift. The post-hoc Tukey tests also revealed significantly higher /u/s for the Southern females than the Mid-Atlantic, Northern, Midland, and Western females (all $p < 0.001$). While this finding was not predicted based on previous descriptions of Southern American English, it is consistent with the results of the analysis of the male talkers, which also revealed back-vowel raising. Finally, the New England females were found to produce significantly higher /u/s than the Western females ($p < 0.001$). This result was not anticipated based on earlier work and further research is needed to explore the effects of dialect on back vowel raising.

For F2 for the female talkers, the ANOVAs revealed significant main effects of dialect for /e/ [$F(5,114) = 8.5$, $p < 0.001$], /ɛ/ [$F(5,113) = 5.2$, $p < 0.001$], /æ/ [$F(5,114) = 17.6$, $p < 0.001$], /ɑ/ [$F(5,110) = 18.4$, $p < 0.001$], /ʌ/ [$F(5,113) = 16.8$, $p < 0.001$], and /o/ [$F(5,113) = 7.4$, $p < 0.001$]. Post-hoc Tukey tests revealed significant /e/ backing for the Southern female talkers compared to the Mid-Atlantic and Western female talkers (both $p < 0.001$), consistent with the Southern Vowel Shift. The Midland females also produced significantly backed /e/s compared to the Mid-Atlantic female talkers, suggesting that this pattern of the Southern Vowel Shift may be spreading to the Midland dialect region. The Northern female talkers produced significantly backed /ɛ/s compared to the Midland talkers, significantly fronted /æ/s compared to all of the other female talkers, significantly fronted /ɑ/s compared to the Midland, Southern, and Western talkers, and significantly backed /ʌ/s compared to the New England, Midland, Southern, and Western talkers (all $p < 0.001$). These results are all consistent with the Northern Cities Chain Shift. Similar to the Mid-Atlantic male talkers, the Mid-Atlantic female talkers produced significantly fronted /ɑ/s compared to the Midland, Southern, and Western female talkers and significantly backed /ʌ/s compared to the Midland and Southern female talkers (all $p < 0.001$). These results are consistent with our earlier claim of an increasing alignment of the low back vowels with respect to F2 among Mid-Atlantic talkers. Finally, the Southern females produced significantly greater /o/ fronting than the Mid-Atlantic and Northern females (both $p < 0.001$).

Taken together, the analysis of the female data revealed components of the Northern Cities Chain Shift among the Northern females, components of the Southern Vowel Shift among the Southern females, some aspects of the Southern Vowel Shift among Midland females, and a new shift in the low back vowels for the Mid-Atlantic female talkers.

3. Merger of /ɑ/ and /ɔ/

To assess the degree of merger of /ɑ/ and /ɔ/, a series of paired-sample t tests was calculated. For each of the six dia-

lect regions, one paired-sample t test was computed for F1 z scores and one was computed for F2 z scores. Significant differences in this analysis suggest distinct vowels, whereas nonsignificant differences in both F1 and F2 suggest a merger or partial merger. Due to the large number of comparisons, the p value was set at 0.005 for this analysis. A partial merger of /ɑ/ and /ɔ/ was found for the New England talkers [$t(7) = 5.4$, $p = 0.001$ for F1 and $t(7) = 3.5$, $p = 0.009$ for F2], the Mid-Atlantic talkers [$t(7) = 3.5$, $p = 0.01$ for F1 and $t(7) = 2.1$, $p = 0.08$ for F2], the Midland talkers [$t(7) = 2.7$, $p = 0.03$ for F1 and $t(7) = 2.0$, $p = 0.08$ for F2], and the Western talkers [$t(7) = 3.6$, $p = 0.01$ for F1 and $t(7) = 0.03$, $p = 0.98$ for F2]. However, /ɑ/ and /ɔ/ were clearly distinct for the Northern talkers [$t(7) = 6.3$, $p < 0.001$ for F1 and $t(7) = 6.5$, $p < 0.001$ for F2] and the Southern talkers [$t(7) = 6.3$, $p < 0.001$ for F1 and $t(7) = 4.4$, $p = 0.003$ for F2].

4. Summary

Overall, the results were highly consistent across both the male and female talkers. However, the three-way interaction between dialect, vowel, and gender was significant, suggesting that gender-specific differences are important. The primary differences between the two genders involve the extent of the Northern Cities Chain Shift and the magnitude of back-vowel fronting. In particular, we found evidence of the later stages of the Northern Cities Chain Shift, particularly /ɛ/ and /ʌ/ backing, among the females but not the males. With respect to back-vowel fronting, we found significant effects of /u/ and /o/ fronting for the males, but only /o/ fronting for the females. An inspection of the raw data suggests that /u/ fronting is more advanced for the female talkers across all of the dialect regions, while /o/ fronting remains predominantly a Southern phenomenon. For the males, however, /u/ fronting is still mostly a Southern and Western feature and has not spread to the northern and eastern dialects. Finally, a merger or partial merger of /ɑ/ and /ɔ/ was found for the New England, Mid-Atlantic, Midland, and Western talkers. An inspection of the individual vowel productions, however, provides additional insights into the range of talker variation observed within and across dialect regions.

B. Descriptive summary

Figures 5–10 show all of the tokens for each of the 11 American English vowels for each dialect. In these figures, the tokens have been plotted in raw frequency in Hertz and by gender for clarity. The ellipses in these figures were drawn by hand to include all of the tokens for each vowel.

1. New England

The top panel of Fig. 5 shows the data for the New England male talkers. The most striking aspect of this figure is the split between the different talkers in their production of /æ/. While three of the talkers showed raised /æ/s similar to those found in the Northern dialect region, one talker (NE1) retained lowered /æ/s.[2] In addition, the New England males produced a merger or partial merger of /ɑ/ and /ɔ/.
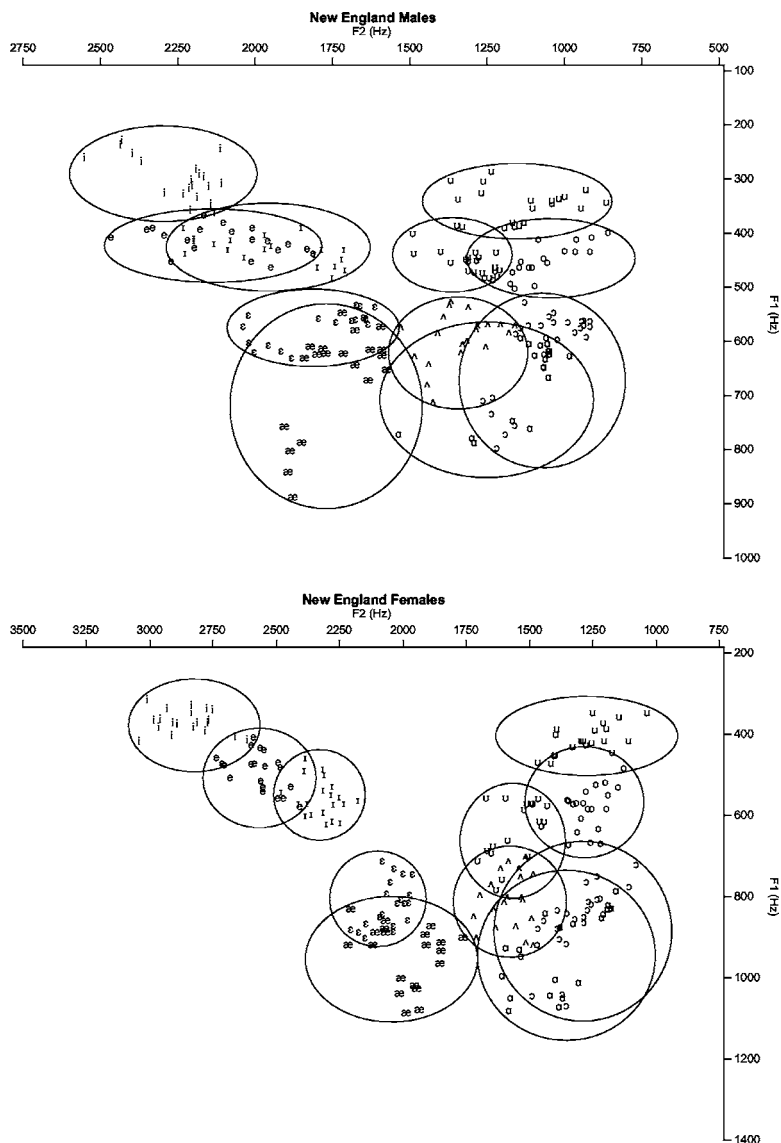
J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Clopper *et al.*: Regional American English vowel systems   1667

The data for the New England females are shown in the bottom panel of Fig. 5. Like the New England males, the New England females showed two different patterns of productions of /æ/. One talker (NE8) had a raised /æ/, consistent with the Northern Cities Chain Shift, while the other three talkers maintained a distinction between the nuclei of /æ/ and /ɛ/. The New England females also produced a merger or partial merger of /ɑ/ and /ɔ/, like their male counterparts.

2. Mid-Atlantic

The data for the Mid-Atlantic males are shown in the top panel of Fig. 6. The Mid-Atlantic males seem to be the least variable subgroup of talkers in the NSP corpus. In general, the ellipses in Fig. 6 are smaller than the ellipses found in the other figures. We also see evidence of the merger of /ɑ/ and /ɔ/ among the Mid-Atlantic males. An inspection of the individual vowel spaces suggests that although this may be the case for two of the talkers (AT1 and AT3), the other two talkers maintained distinct low back vowels.

The bottom panel of Fig. 6 shows the vowel formant frequency data for the Mid-Atlantic female talkers. Like the Mid-Atlantic males, the Mid-Atlantic females were inconsis-

tent in producing the low-back merger. One talker (AT9) showed a clear merger of /ɑ/ and /ɔ/, while the other three talkers maintained a greater distinction between these two vowels. Only one talker (A18) produced the expected raised /ɔ/. The Mid-Atlantic female talkers were also variable in their production of /u/. One talker (A18) produced fronted /u/s, while the other three females produced more backed /u/s. Finally, Fig. 6 shows a great deal of overlap between /ʊ/ and /ʌ/ for the Mid-Atlantic females. An inspection of the individual vowel spaces suggests that this overlap is not due to mergers at the individual talker level, but is due to variation across talkers in the production of these vowels, particularly in F1.

3. North

The top panel of Fig. 7 shows the data for the Northern males. The Northern Cities Chain Shift appears to be present in all four talkers in this sample. /ɑ/ is clearly distinct from /ɔ/, due to the lowering and fronting of /ɑ/. All four talkers also produced raised and/or fronted /æ/s. In addition, /ɛ/ shows some backing to reduce overlap with raised /æ/. An inspection of the individual vowel spaces indicates that the
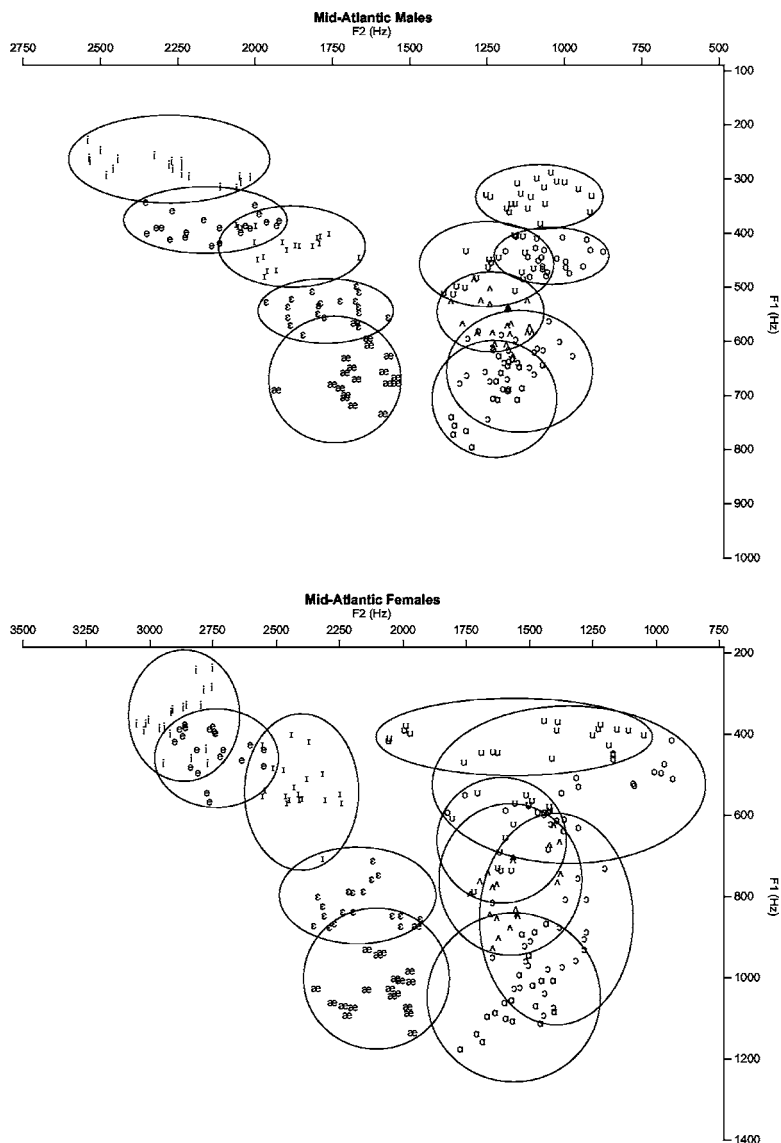
FIG. 6. All tokens produced by the Mid-Atlantic male (top) and female (bottom) talkers for the 11 vowels /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/. The ellipses were hand-drawn to include every token for each vowel.

talkers from upstate New York and Wisconsin produced backed /ɛ/s, while the two talkers from northern Indiana maintained a more fronted production of /ɛ/. /ʌ/ was also backed and shows some overlap with /ɔ/.

The data for the Northern females are shown in the bottom panel of Fig. 7. The vowels produced by the Northern females also reflect the Northern Cities Chain Shift. Like the Northern males, these female talkers produced lowered and fronted /ɑ/s, raised and fronted /æ/s, and backed /ɛ/s and /ʌ/s. One of the Northern female talkers (NO9) also produced fronted /u/s, while the other three retained backed /u/s.

Despite the overlap of /æ/ and /ɛ/ in the vowel nuclei, the trajectories of these vowels are distinct for both the male and female talkers. While /ɛ/ moves up over the middle third of the vowel, /æ/ moves back and down, so that the offsets of the two vowels are well separated in the F1 × F2 space.

4. Midland

The vowel tokens for the Midland male talkers are shown in the top panel of Fig. 8. This figure reveals two distinctive splits between the talkers. First, two talkers (MI3 and MI4) showed fronted /u/s, whereas the other two talkers

had more backed /u/ productions. In addition, two of the talkers (M12 and MI3) had raised /æ/s, whereas the other two talkers retained the lower /æ/ production. These results are particularly interesting because /u/ fronting is associated with the Southern Vowel Shift while /æ/ raising is associated with the Northern Cities Chain Shift, but they were both present in a single Midland talker. It is somewhat surprising that a single talker (MI3) would exhibit both Southern /u/ fronting and Northern /æ/ raising. It is also interesting to note in this figure that the vowel /ɪ/ is completely encompassed by the vowel /e/. This result is due to an apparent merger of these two vowels (at least in terms of nucleus formant frequencies) for one of the Midland male talkers (MI4). An inspection of the trajectories of these vowels for MI4, however, suggests that the phonemic distinction is maintained through differences in the spectral change from the onset to the offset of the vowel. While /e/ moves forward and up over the middle third of the vowel, /ɪ/ moves backward, and the offsets of the two vowels do not overlap in the F1 × F2 space.

The vowel tokens for the Midland females are plotted in the bottom panel of Fig. 8. Unlike their male counterparts,
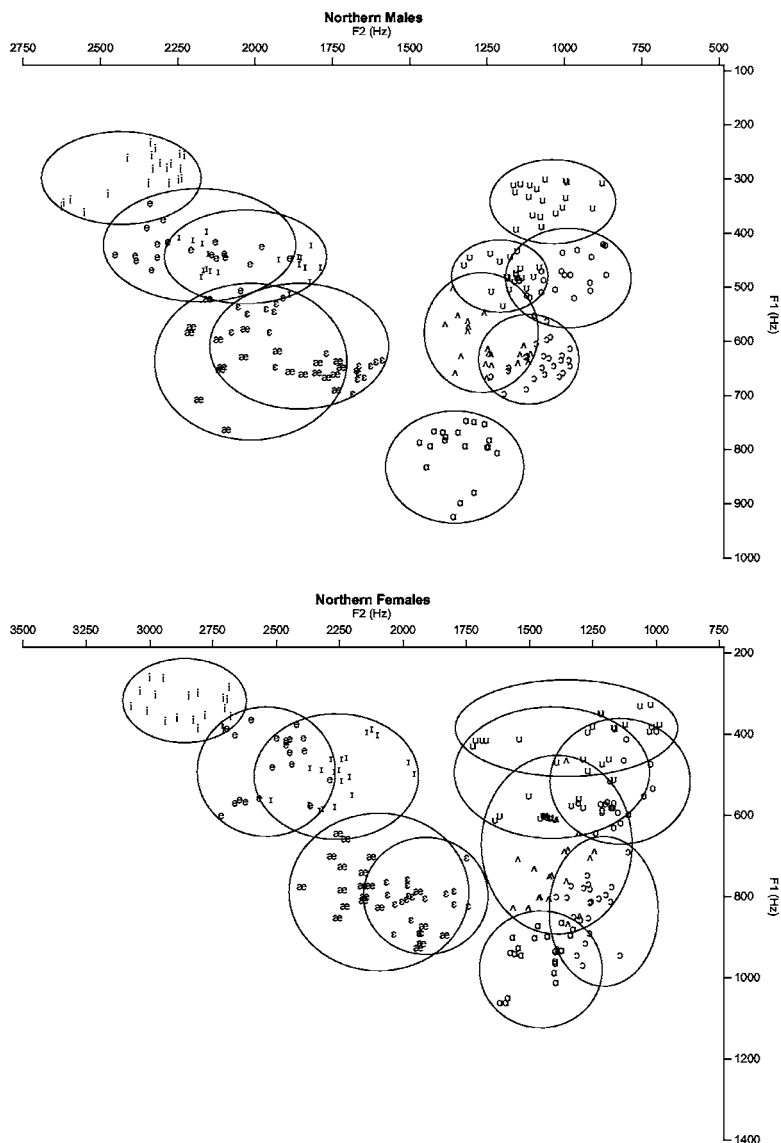
FIG. 7. All tokens produced by the Northern male (top) and female (bottom) talkers for the 11 vowels /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/. The ellipses were hand-drawn to include every token for each vowel.

who individually showed evidence of Southern and Northern features, the Midland females showed very few shifted vowels. An inspection of the individual vowel spaces suggests that the variation in Fig. 8 is due to overall differences between talkers and not to individual differences in vowel shifts or mergers, with the exception of one talker (MI6) who produced slightly raised /æ/s. The Midland female talkers as a group also exhibited the merger of /ɑ/ and /ɔ/ and this merger is evident in the individual vowel spaces for three (MI8, MI9, and MI0) of the four talkers.

5. South

Vowel production data for the Southern males are shown in the top panel of Fig. 9. /u/ and /o/ fronting are consistently present in all of the tokens. In the case of /o/ fronting, this results in a near-complete overlap of /o/ and /ʊ/ across talkers. An inspection of the individual vowel spaces, however, suggests that each individual talker maintained a distinction between /o/ and /ʊ/ in both the nucleus position and in the formant trajectory over the course of the vowel. Like the Midland talkers, the Southern talkers showed highly similar /e/s and /ɪ/s with respect to formant frequencies at the first-

third temporal point, but in all four talkers, the trajectories for /e/ and /ɪ/ were clearly distinct. /e/ moved up and front over the course of the vowel, while /ɪ/ moved back and down. Finally, /æ/ was also fronted consistently in all of the tokens.

The bottom panel of Fig. 9 is a plot of the vowel tokens for the Southern females. The Southern females were by far the most variable subgroup of talkers in the NSP corpus. The overlap across different vowels is quite large for the entire space, except for /i/. As a group, the Southern females showed a general trend for /u/ and /o/ fronting like the Southern males, but individually, the female talkers ranged from virtually no fronting of these vowels by one of the talkers from Kentucky (SO6) to very fronted back vowels by one of the talkers from Texas (SO7).

6. West

The top panel of Fig. 10 shows the data for the last group of male talkers, the Westerners. The large variance in F2 for /u/, /ʊ/, and /o/ is due to the talker from Montana (WE4), who produced back vowels with very low second formants. The merger of /ɑ/ and /ɔ/ is also visible in this

Midland Males
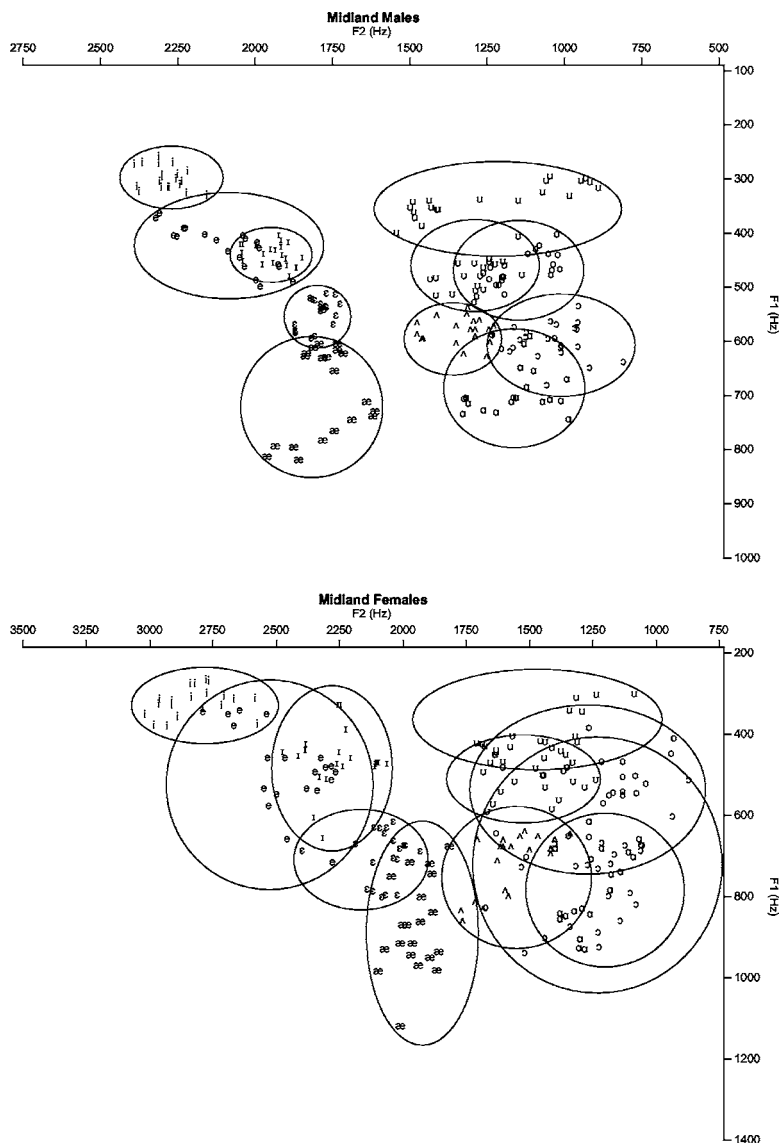
Midland Females

FIG. 8. All tokens produced by the Midland male (top) and female (bottom) talkers for the 11 vowels /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/. The ellipses were hand-drawn to include every token for each vowel.

figure, and an inspection of the individual vowel spaces suggests that all four talkers produced merged or nearly merged low back vowels.

Finally, the data for the Western female talkers are shown in the bottom panel of Fig. 10. Like their male counterparts, the Western females showed a merger of /ɑ/ and /ɔ/ as a group and individually. In addition, /u/ fronting was found in two of the Western females (WE7 and WE9), but not in the others.

In general, the female talkers were more variable within a given dialect than the male talkers. A comparison of the individual vowel spaces, however, suggests that individual females were not generally more variable in their productions of specific vowels than individual males, but that across different female talkers within a given dialect, there was simply more variation in the formant frequencies. Impressionistically, the female talkers in the NSP corpus sounded more heterogeneous than the male talkers in terms of speaking style, which may explain the observed differences in the raw acoustic measures.

The raw data presented in Figs. 5–10 reveal substantial variability within each dialect region, despite the relative ho-mogeneity of the talkers with respect to age, ethnicity, and level of education. As noted above, /æ/ raising was variable in New England, /u/ fronting was variable in the Mid-Atlantic, Midland, and West, and the merger of /ɑ/ and /ɔ/ was also variable in the Mid-Atlantic and West. This within-dialect variability reflects the different geographic locations and influences that the talkers have been exposed to. Each of the six geographic dialect regions described in the current study is composed of smaller speech communities, which may have their own unique dialect features. In addition, the talkers had different residential and travel histories, which provided them with different opportunities to hear and acquire regional and social variants. The results of the statistical analyses, however, confirm that these talkers also produce consistent between-dialect differences that reflect previous descriptive sociolinguistic research on the primary regional varieties of American English.

IV. DISCUSSION

The statistical analysis of the vowel duration and formant frequency measures confirmed the presence of the

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Clopper *et al.*: Regional American English vowel systems 1671
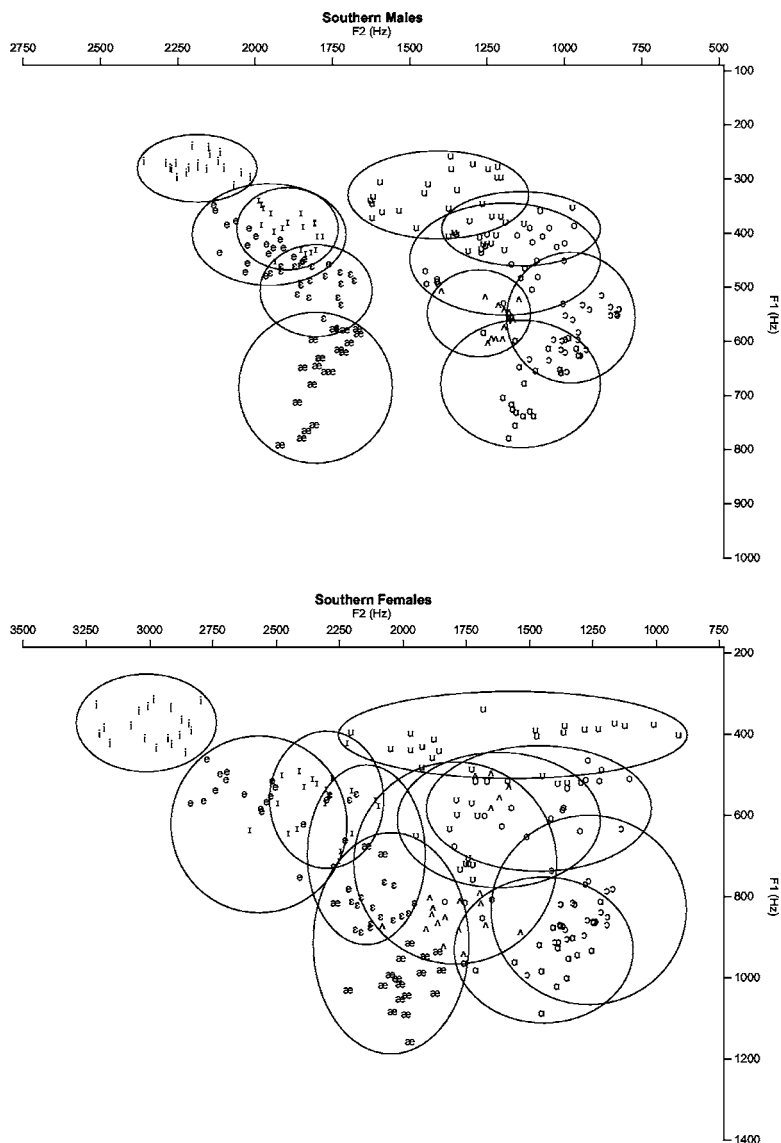
FIG. 9. All tokens produced by the Southern male (top) and female (bottom) talkers for the 11 vowels /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/. The ellipses were hand-drawn to include every token for each vowel.

Northern Cities Chain Shift in the Northern talkers and the Southern Vowel Shift in the Southern talkers. In particular, Northerners produced lowered and fronted /ɑ/s and fronted and raised /æ/s. The Northern females also produced backed /ɛ/s and /ʌ/s. The Southern talkers exhibited fronting of /o/ and /e/ centralization. The Southern males also produced fronted /u/s and raised /ɛ/s. We also found evidence of several new features of Southern speech: the fronting of /æ/ (male talkers only) and the raising of /ʊ/ and /u/. Finally, the Southerners produced longer lax vowels than any of the other talkers. This increase in lax vowel duration could lead to greater perceptual confusability between the tense and lax vowel pairs (such as /i/ and /ɪ/ or /u/ and /ʊ/), but an inspection of the vowel trajectories suggests that Southern talkers maintain the tense/lax distinction through greater spectral change in the lax vowels than the other talker groups. Taken together, the most robust dialect differences in the present dataset were the Northern Cities Chain Shift among the Northern talkers and the Southern Vowel Shift among the Southern talkers.

In addition, the Westerners produced a merger of /ɑ/ and /ɔ/, as predicted, and the Western males exhibited /u/ fronting. The low-back merger was also reliable for the Midland talkers. In addition, both the male and female Midland talkers exhibited some aspects of the Southern Vowel Shift with the males producing raised /ɛ/s and the females producing centralized /e/s. The individual data plotted for each dialect region in Figs. 5–10 also suggest that other dialect-specific features are present in at least some talkers, although they might not be consistent enough across all of the talkers to be significant in the statistical analyses conducted in this study. For example, some Midland talkers showed /u/ fronting, a typically Southern feature, while others showed Northern features, such as /æ/ raising. All three of the talkers who exhibited these Northern and/or Southern features in their speech were from the Indianapolis metropolitan area. These results are consistent with earlier claims in the literature that the Midland dialect region is not a unique dialect, but instead may be a transition area between the North and the South (Davis and Houck, 1992; but also see Frazer, 1994; Johnson, 1994).
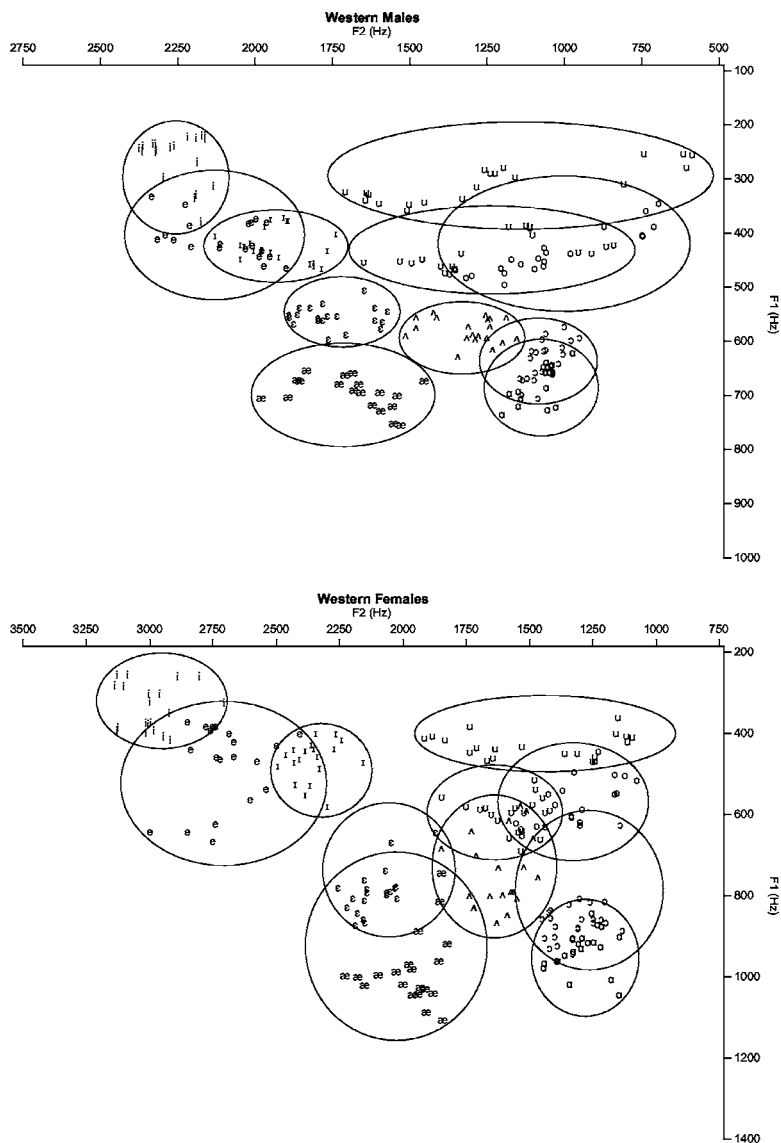
FIG. 10. All tokens produced by the Western male (top) and female (bottom) talkers for the 11 vowels /i ɪ e ɛ æ ɑ ɔ ʌ o ʊ u/. The ellipses were hand-drawn to include every token for each vowel.

Boberg (2001) suggested that Western New England can also be treated as a transition area between Eastern New England and the North. At first glance, the data presented here support this interpretation: some of the New England talkers showed the /æ/ raising found in the North and some did not. However, /æ/ raising among the male talkers was found in both Eastern and Western New England, while the females from Western New England were split on /æ/ raising with one producing the raised variant and one not. In addition, all of the New England talkers, both Easterners and Westerners, showed the merger of /ɑ/ and /ɔ/ reportedly found in Eastern New England. A partial merger of /ɑ/ and /ɔ/ in the New England talkers was confirmed by the statistical analysis. These results suggest that Eastern and Western New England are perhaps more homogeneous than previously suggested (Boberg, 2001; Labov *et al.*, forthcoming).

The distinctive features of the Mid-Atlantic talkers were less clear. We did not find evidence of raised /ɔ/ in most of the speakers, and, instead, found evidence of an unexpected merger of /ɑ/ and /ɔ/. In addition, the acoustic analysis revealed significant differences between the Mid-Atlantic talkers and the other talker groups in the F2 of /ɑ/, /ɔ/, and /ʌ/. The result of these shifts is particularly apparent in Fig. 4, which shows the alignment of these three low vowels with respect to F2 for the Mid-Atlantic males (filled squares). The pattern is less clear for the Mid-Atlantic females, although significant effects were found for F2 for both /ɑ/ and /ʌ/ for these talkers. Given the reported tendency for [ɔ] raising in the Mid-Atlantic (Labov, 1994; Thomas, 2001), these findings were somewhat surprising. However, the lexical items used to obtain the measure of [ɔ] (*frogs* and *logs*) were among a small set of words that have been reported to show large variation between regional varieties of American English and do not always pattern with other words containing the same vowel (Trager, 1930; Wells, 1982). Thus, this effect may be lexically specific. An additional exploration of this result is needed to determine the precise nature of the low back vowels in Mid-Atlantic speech.

In addition to these potential lexical effects, several other aspects of the recording conditions and stimulus materials should be considered in interpreting the results. First, while the recordings were made using high quality digital equipment in a sound-attenuated booth, the participants were asked to read the stimulus materials that were displayed on a computer monitor, which may have resulted in somewhat unnatural utterances. Both speech scientists and sociolinguists have shown that more informal speech can be obtained when participants are allowed to talk spontaneously (Labov, 1972; Ladefoged, Kameny, and Brackenridge, 1976). Second, the stimulus materials that were measured in this study were restricted to the hVd context, and we did not explore vowel tokens in pre-/r/, pre-/l/, or pre-nasal environments. However, the effects of the following consonant, particularly nasals and liquids, on vowel production have been well documented by sociolinguists (Labov *et al.*, forthcoming; Thomas, 2001). Third, most of our discussion of the vowel productions were based on acoustic measurements made at a single temporal point in each vowel token, but a preliminary inspection of the trajectories suggests that additional variation may also be present in how talkers from different regions manipulate spectral change to maintain vowel category contrasts. Finally, while the NSP corpus included nearly an hour of speech from each talker, we examined only a small number of tokens from each talker and the statistical analyses relied on differences between four talkers of each gender of each dialect. In addition, our talkers were relatively homogeneous with respect to age, level of education, and ethnicity. Despite these limitations, however, we hope to have laid the foundation for future research that will examine in more detail the production of American English vowels in different speaking styles and different phonetic and lexical contexts, the role of spectral movement in maintaining vowel category contrast, and talkers representing a wider range of ages, ethnicities, and socioeconomic groups.

The *z*-score transformation that we used to normalize the vowel formant frequency data in this study was generally successful. The statistical analyses revealed no significant effect of gender on F1 or F2, but significant effects of vowel category and significant vowel X dialect and vowel X gender interactions were maintained. The transformation did produce artifacts in the statistical analysis, however, particularly for the Southern talkers. As discussed above, the back-vowel fronting that is found in the speech of Southern talkers led to a higher mean F2 and a smaller F2 standard deviation. The *z*-score transformation then produced artificially backed low back vowels as a result of the larger numerator and smaller denominator. Similarly, the *z*-score transformation is not adequate for the normalization of formant frequency measures obtained at multiple temporal points in the vowel. When vowels that exhibit spectral movement are included in an analysis, the overall shape of the vowel space will differ at different temporal slices, and we predict that artifacts of the transformation will be produced. We can conclude that in cases where vowel systems are being compared that differ in their overall shape, the *z*-score transform should be used with caution.

As mentioned in the Introduction, one important assumption underlies the present discussion of the different vowel systems of regional varieties of American English: The characteristics of each dialect are defined relative to an unspecified baseline. This baseline could be defined historically in terms of earlier vowel systems in the United States. For example, Labov (1994) characterized the Northern Cities Chain Shift in historical terms, by describing the vowel system of the Northern dialect as the result of a series of phonological changes that can be traced through both real-time and apparent-time data[3] over the course of the second half of the 20th century. An alternative to this historical perspective would be to identify baseline pronunciations based on the current vowel systems of the different regional varieties. For example, Fig. 4 showed relatively consistent productions of /i/ across the six different dialects, but much greater variation in /æ/, /ɑ/, and /o/. In the case of /æ/, the Northern talkers were the only ones who produced the raised and fronted variant, suggesting that the lowered and backed production should be treated as the baseline. Similarly, the Southern /o/ was fronted, while a backed /o/ was found in the other five dialects, suggesting that the backed variant should be treated as the baseline. By considering both historical developments and synchronic idiosyncrasies, speech scientists can develop implicit baseline productions to which many other possible variants are compared.

The present results demonstrate that the acoustic characteristics of the vowel systems of American English differ based on the region of origin of the talker. In the current study, talkers from the Northern dialect region reliably produced the Northern Cities Chain Shift, whereas talkers from the Southern dialect region reliably produced some features of the Southern Vowel Shift. Finally, the merger of /ɑ/ and /ɔ/ was robustly present in New England, Mid-Atlantic, Midland, and Western talkers. Our analysis also revealed four new results: the spreading of the high front vowel shifts from the Southern to the Midland dialect, the raising of the high back vowels for the Southern talkers, the fronting of /æ/ for the Southern male talkers, and the alignment of the low back vowels /ɑ/, /ɔ/, and /ʌ/ with respect to F2 in the Mid-Atlantic dialect. These data provide new benchmarks for the acoustic characteristics of the vowel systems of six regional varieties of American English as well as implicit baselines for "General" American English.

APPENDIX: TALKER DEMOGRAPHIC INFORMATION

Key: New England (NE), Mid-Atlantic (AT), North (NO), Midland (MI), South (SO), West (WE).

Talker ID	Sex	Age	Hometown (city, state)	Education	Occupational field (or major)
NE1	m	24	Chichester, NH	BA/BS	Psychology
NE2	m	20	Marblehead, MA	Undergraduate[4]	Law
NE3	m	20	Sandown, NH	Undergraduate	Business
NE4	m	18	Londonderry, NH	Undergraduate	Finance
NE6	f	24	Winchester, MA	BA/BS	English
NE7	f	18	Newfane, VT	Undergraduate	Business
NE8	f	18	Enfield, CT	Undergraduate	Translation
NE0	f	18	Sharon, MA	Undergraduate	Psychology
AT1	m	18	Long Island, NY	Undergraduate	Business
AT2	m	18	Manalapan, NJ	Undergraduate	Law
AT3	m	18	Marlboro, NJ	Undergraduate	Accounting
AT5	m	18	Plainview, NY	Undergraduate	Television
AT6	f	18	Middletown, NJ	Undergraduate	Musical Theater
AT7	f	20	Great Neck, NY	Undergraduate	Public Relations
A18	f	18	Owings Mills, MD	Undergraduate	Education
AT9	f	18	Oradell, NJ	Undergraduate	Apparel Merchandising
NO2	m	20	Buffalo, NY	Undergraduate	Law
NO3	m	18	St. John, IN	Undergraduate	Undecided
NO4	m	19	New Berlin, WI	Undergraduate	Business
NO5	m	18	Hobart, IN	Undergraduate	Education
NO6	f	19	South Bend, IN	Undergraduate	Education
NO8	f	18	Wilmette, IL	Undergraduate	Advertising
NO9	f	20	Munster, IN	Undergraduate	Business
NO0	f	18	South Bend, IN	Undergraduate	Journalism
MI1	m	19	Centerville, IN	Undergraduate	Broadcasting
M12	m	19	Greencastle, IN	Undergraduate	Film
MI3	m	19	Carmel, IN	Undergraduate	Film
MI4	m	18	Anderson, IN	Undergraduate	Investment Banking
MI6	f	19	Grandview, IN	Undergraduate	Nutrition Science
MI8	f	19	Seymour, IN	Undergraduate	Dentistry
MI9	f	18	Selvin, IN	Undergraduate	Psychology
MI0	f	20	Lafayette, IN	Undergraduate	Law
SO1	m	18	New Albany, IN	Undergraduate	National Security
SO2	m	22	Statesville, SC	BA/BS	Music Performance
S22	m	20	Georgetown, IN	Undergraduate	Education
SO5	m	18	Birmingham, AL	Undergraduate	Investment Banking
SO6	f	18	Louisville, KY	Undergraduate	Journalism
SO7	f	18	Dallas, TX	Undergraduate	Retail
SO8	f	19	League City, TX	Undergraduate	Interior Design
S10	f	19	Owensboro, KY	Undergraduate	Physical Therapy
WE2	m	20	Albuquerque, NM	Undergraduate	Law
WE3	m	19	Oakland, CA	Undergraduate	Journalism
WE4	m	23	Billings, MT	MA/MS	Library Science
WE5	m	18	Santa Clarita, CA	Undergraduate	Music Performance
WE6	f	20	West Covina, CA	Undergraduate	Broadcast News
WE7	f	19	Los Angeles, CA	Undergraduate	Music Performance
WE8	f	21	Henderson, NV	Undergraduate	Biology
WE9	f	20	Orange, CA	Undergraduate	Fashion

[1]Due to an oversight in the design and collection of the NSP corpus, /ɔ/ was not available in hVd context. The use of the *frogs* and *logs* tokens, while not ideal, allowed us to examine the vowel systems of our talkers more completely and to discuss the regional extent of the merger of /ɑ/ and /ɔ/. More details about the sentences can be found in Clopper (2004).

[2]Plots of the individual vowel spaces, including formant trajectories, for the 48 talkers are presented in Clopper (2004).

[3]"Realtime" refers to data collected longitudinally to study language change over time within individuals and within a speech community. "Apparent-time" refers to data collected at a single point in time from participants with a range of ages, from which inferences about language change can be tentatively drawn (Labov, 1994).

[4]"Undergraduate" refers to talkers who were enrolled as undergraduate students at Indiana University at the time of recording.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Bauer, L. (**1985**). "Tracing phonetic change in the received pronunciation of British English," J. Phonetics **13**, 61–81.

Boberg, C. (**2001**). "The phonological status of Western New England," Am. Speech **76**, 3–29.

Clopper, C. G. (2004). "Linguistic experience and the perceptual classification of regional varieties of American English," Indiana University Ph.D. dissertation.

Davis, L. M., and Houck, C. L. (**1992**). "Is there a Midland dialect?—Again," Am. Speech **67**, 61–70.

Frazer, T. C. (**1994**). "On transition areas and the 'Midland' dialect: A reply to Davis and Houck," Eur. J. Morphol. **69**, 430–435.

Hagiwara, R. (**1997**). "Dialect variation and formant frequency: The American English vowels revisited," J. Acoust. Soc. Am. **102**, 655–658.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Johnson, E. (**1994**). "Yet again: The Midland dialect," Am. Speech **69**, 419–430.

Labov, W. (**1972**). "Some principles of linguistic methodology," Lang. Soc. **1**, 97–120.

Labov, W. (**1994**). *Principles of Linguistic Change: Internal Factors* (Blackwell, Malden, MA).

Labov, W. (**1998**). "The three dialects of English," in *Handbook of Dialects and Language Variation*, edited by M. D. Linn (Academic Press, San Diego), pp. 39–81.

Labov, W., Ash, S., and Boberg, C. (**to be published**). *Atlas of North American English* (Mouton de Gruyter, New York).

Ladefoged, P., Kameny, I., and Brackenridge, W. (**1976**). "Acoustic effects of style of speech," J. Acoust. Soc. Am. **59**, 228-231.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.

Sjölander, K., and Beskow, J. (2004). *WaveSurfer*, Version 1.6.2 (computer software), Centre for Speech Technology, Royal Institute of Technology, Stockholm.

Thomas, E. R. (**2001**). *An Acoustic Analysis of Vowel Variation in New World English* (Duke University Press, Durham, NC).

Trager, G. L. (**1930**). "The pronunciation of 'short a' in American Standard English," Am. Speech **5**, 396–400.

Wells, J. C. (**1982**). *Accents of English 3: Beyond the British Isles* (Cambridge University Press, Cambridge).

Production and perception of clear speech in Croatian and English[a)]

Rajka Smiljanić[b)] and Ann R. Bradlow[c)]

Department of Linguistics, Northwestern University, Evanston, Illinois 60208

Previous research has established that naturally produced English clear speech is more intelligible than English conversational speech. The major goal of this paper was to establish the presence of the clear speech effect in production and perception of a language other than English, namely Croatian. A systematic investigation of the conversational-to-clear speech transformations across languages with different phonological properties (e.g., large versus small vowel inventory) can provide a window into the interaction of general auditory-perceptual and phonological, structural factors that contribute to the high intelligibility of clear speech. The results of this study showed that naturally produced clear speech is a distinct, listener-oriented, intelligibility-enhancing mode of speech production in both languages. Furthermore, the acoustic-phonetic features of the conversational-to-clear speech transformation revealed cross-language similarities in clear speech production strategies. In both languages, talkers exhibited a decrease in speaking rate and an increase in pitch range, as well as an expansion of the vowel space. Notably, the findings of this study showed equivalent vowel space expansion in English and Croatian clear speech, despite the difference in vowel inventory size across the two languages, suggesting that the extent of vowel contrast enhancement in hyperarticulated clear speech is independent of vowel inventory size. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000788]

I. INTRODUCTION

Talkers naturally and spontaneously adopt a distinct intelligibility-enhancing mode of speech production called "clear speech" when they are aware of a speech perception difficulty on the part of the listener due to background noise, a hearing impairment, or a different native language. It seems rather obvious that, in an attempt to make themselves more intelligible, most talkers will speak more slowly, more loudly, and in a more "exaggerated" manner, regardless of their language background. What is not so obvious is the extent to which the intelligibility-enhancing modifications that talkers adopt are driven by phonological, structural properties. In this paper, we report on a cross-linguistic study in which we test the hypothesis that clear speech production reflects the interaction of universal, auditory-perceptual factors, which serve to enhance the overall acoustic salience of the speech signal, and phonological, structural factors, which serve to enhance the acoustic "distance" between contrasting phonological categories.

A substantial body of previous work has provided us with important insights into the nature of high-intelligibility clear speech in English; however, there is a paucity of data on clear speech production and perception in any language other than English. Previous data on English clear speech perception have established that clear speech significantly,

although to different degrees, enhances intelligibility for various listener populations, including normal-hearing and hearing-impaired adults, children with and without learning disabilities as well as non-native listeners, and under a variety of degraded listening conditions, including varying levels of noise and reverberation (Picheny *et al.*, 1986; Payton *et al.*, 1994; Uchanski *et al.*, 1996; Bradlow and Bent, 2002; Bradlow *et al.*, 2003; Ferguson, 2004). Regarding the articulatory modifications of naturally produced English clear speech, the accumulated results show that clear speech involves a wide range of acoustic/articulatory adjustments, including a decrease in speaking rate, which involves longer segments as well as more frequent and longer pauses, an increase in pitch range, greater sound-pressure levels, more salient stop releases, greater obstruent rms intensity, increased energy in the 1000–3000 Hz range of long-term spectra, and an expanded vowel space (Picheny *et al.*, 1986, 1989; Krause and Braida, 2004; Bradlow *et al.*, 2003; Liu *et al.*, 2004; Moon and Lindblom, 1994; Ferguson and Kewley-Port, 2002; Johnson *et al.*, 1993). However, the generalizability of these findings to clear speech production in other languages is unknown due to the lack of cross-linguistic clear speech research.

To our knowledge, there are only two studies of clear speech in languages other than English. Gagne and colleagues (2002) investigated the intelligibility of CV and VCV syllables in Canadian French clear and conversational speech in auditory, visual, and audiovisual modalities. Their results showed a clear speech intelligibility benefit in all three modalities; however, acoustic data for the two perceptually distinct speaking styles were not presented. Bradlow

(2002) examined vowel production and CV coarticulation in clear and conversational speech in English and Spanish; however, the materials used in that study were rather limited (just high vowels, /i/ and /u/) and there were no accompanying intelligibility data that would show whether the clear speech intelligibility benefit was of similar magnitude in the two languages.

Given this almost exclusively monolingual focus of previous clear speech research, there is no direct evidence for the influence of phonological structure on clear speech production. Nevertheless, there are two independent sources of circumstantial evidence for the influence of phonological features on clear speech production. First, some of the acoustic-phonetic features of English clear speech are directly related to the sound structure of English. For example, Uchanski (1988, 1992) found that the duration contrast between tense/long and lax/short vowels was enhanced in English clear speech (by lengthening the tense/long vowels to a greater extent than the lax/short vowels). This finding suggests that the nonuniform increase in segment durations for clear speech reflects the temporal structure of the language at the segmental level. Similarly, Cutler and Butterfield (1990) found that preboundary syllable lengthening was exaggerated in clear speech relative to conversational speech, especially in cases where the preboundary syllable occurred before a word that began with a weak syllable. This suggests that clear speech production also reflects the temporal structure of the language at the suprasegmental level where the basic rhythmic structure of the language comes into play. The extra syllable lengthening before a word that begins with a weak/unstressed syllable makes the word boundary particularly salient in exactly the situation where the language-specific stress-group-based segmentation strategy will fail (Cutler *et al.*, 1986; Cutler and Otake, 1994).

The second source of circumstantial evidence for a language-specific influence on clear speech production comes from studies of language-specific influences on slow-to-fast rate modifications, on adult-directed-to-infant-directed style modifications, and on broad versus narrow focus conditions. While these rate, style, and focus modifications differ from the conversational-to-clear speech transformation in that they are typically elicited in the laboratory with no explicit instruction to enhance intelligibility (and these studies usually do not include corresponding measures of intelligibility), they are similar to clear speech in that they involve a change from a relatively hypoarticulation style to a relatively hyperarticulation style (Lindblom, 1990). Solé (1992, 1995) found that in Spanish, vowels preceding nasal consonants are nasalized for approximately the same amount of time in an absolute sense across slow, normal, and fast speaking rates. In contrast, English vowels preceding nasal consonants are nasalized for approximately the same proportion of their total duration (virtually 100% of the vowel duration) across speaking rates. This difference is taken to reflect the different status of anticipatory vowel nasalization in the two languages: in Spanish, it is an unintended effect of constraints on vocal tract dynamics (the minimum amount of time for the velum to lower is constant regardless of speaking rate), whereas in English, vowels pre-

ceding nasals are intentionally nasalized by a process that adjusts to speaking rate variations. Similarly, a cross-linguistic comparison of infant- versus adult-directed speech demonstrated equivalent amounts of vowel space expansion for American English, Swedish, and Russian infant-directed speech relative to adult-directed speech (Kuhl *et al.*, 1997). In contrast, Andruski, Kuhl, and Hayashi (1999) found dramatically reduced vowel space expansion for Japanese infant-directed speech. While it is difficult to determine whether this difference in vowel space expansion across Japanese and the other three languages reflects a property of the Japanese language, such as the relatively uncrowded vowel inventory (although note that Russian also has a relatively uncrowded vowel space and yet exhibited significant vowel space expansion for infant- versus adult-directed speech), or a property of the culture, such as a reduced tendency to use a distinct infant-directed style of speech, this finding suggests that the amount of vowel space expansion for hyperarticulated speaking styles may vary across languages.

Further evidence for a language-specific effect on vowel space expansion comes from a pair of studies aimed at testing predictions of the Theory of Adaptive Dispersion, which states that talkers seek to provide a sufficient degree of distinctiveness among contrasting categories while minimizing articulatory effort needed to achieve this distinctiveness (Lindblom 1986, 1990; Diehl and Lindblom 2002). Hay *et al.*, (2003) and Coren and Heckmann (2004) tested the prediction that languages with large vowel inventories, such as English, French, and German, will exhibit greater vowel space expansion for words in narrow focus relative to the same words in broad focus than languages with small vowel inventories, such as, Japanese.[1] The results supported their prediction, thereby providing additional evidence in favor of the claim that hyperarticulation in general is responsive to language-specific, phonological, structural properties. It is important to note, however, that none of these studies of hyperarticulation involves a mode of speech that is for the express purpose of enhancing intelligibility, nor do they present any accompanying intelligibility data which could help establish whether these hyperarticulations have any bearing on intelligibility.

Taken together then, the work on English clear speech production and cross-language studies of rate, speaking style, and focus variation suggest that the acoustic-phonetic features that characterize the conversational-to-clear speech transformation may vary across languages in a way that is related to language-specific phonological structure and patterns of phonetic implementation. However, only systematic cross-language comparisons of the conversational-to-clear speech transformation and its consequences for intelligibility would allow us to determine conclusively whether clear speech (i.e., global hyperarticulation for the express purpose of enhancing intelligibility) is or is not an English-specific phenomenon, and whether phonological contrast enhancement is likely to be a significant mechanism for the clear speech intelligibility benefit. To this end, we conducted a comparative study of Croatian and English clear speech. These two languages are well-suited for this comparison be-

cause of the structural differences between their phonologies. Although they are both Indo-European languages, they come from different language families (Slavic versus Germanic) and are typologically very different along multiple sound-based parameters. For example, in terms of rhythmic structure and phonotactics, English is stress timed, with complex consonant clusters in both onset and coda positions and extensive vowel reduction in unstressed syllables. In contrast, Croatian remains unclassified in terms of the three most common rhythm classes (stress-, syllable-, or mora-timed), since it allows complex consonant clusters in both onset and coda positions (like canonical stress-timed language), yet does not exhibit vowel quality reduction (i.e., centralization towards schwa) in unstressed syllables (like canonical syllable-timed languages). Most relevant for our investigation is that English has a large vowel inventory with 14 contrasting vowel quality categories, while Croatian has a relatively small vowel inventory with just 5 contrasting vowel quality categories.

In this paper, we set out to establish first, whether the clear speech effect is present in both production and perception of a language other than English, namely Croatian. That is, do talkers from both languages respond similarly to the instruction to "speak clearly for the sake of a listener with speech perception difficulties," and if so, does the conversational versus clear speech mode difference in the two languages correlate with an intelligibility difference in both languages? The second goal of the present study is to perform a systematic acoustic-phonetic comparison of the conversational-to-clear speech mode transformation in the two languages. This comparison is based on productions of comparable materials by several talkers in each language who were recorded under comparable conditions, and focuses on global, signal enhancing modifications, including speaking rate and pitch range, and on phonological contrast enhancement as reflected in vowel space expansion. In this regard, we ask specifically whether the vowel space is expanded (equally) in both English and Croatian, languages with large (14) and small (5) vowel inventories, respectively. The results will allow us to identify the talker characteristics that likely contribute to the characteristically high intelligibility of clear speech in two unrelated languages, as well as to investigate the interaction of general auditory-perceptual and phonological factors in promoting the clear speech intelligibility benefit.

II. METHODS

A. Participants

1. Production

Five native talkers of Croatian (two female and three male) and five native talkers of English (three female and two male) served as participants in the production study. Age range was between 18 and 25 for Croatian talkers and between 28 and 48 for English talkers. Croatian talkers came to the United States from Croatia within the last 5 years to pursue undergraduate degrees at Northwestern University. They were all from the same region on the coast of Croatia. English talkers were graduate students in the Linguistics De-

partment at Northwestern University. They were native talkers of general American English. None of the talkers had any known speech or hearing impairment at the time of recording. They were not aware of the purpose of the recordings. All participants were paid at the end of the recording session.

2. Perception

Twenty Croatian and 30 English listeners participated in Croatian and English sentence-in-noise perception tests, respectively. The Croatian listeners were undergraduate students in the English Department at the University of Zagreb, Croatia. The age of Croatian listeners ranged between 18 and 28 years. They were paid for their participation. Undergraduate students at Northwestern University received class credit for their participation in the English listening test. The English listeners' ages ranged between 18 and 22 years. None of the listeners had any known speech or hearing impairment at the time of the test.

B. Stimuli

Twenty sentences were designed in each language to investigate the effect of clear speech production and perception in Croatian and English. In order to minimize the signal-independent contextual cues available to listeners in the perception tests, we constructed semantically anomalous sentences. The particular words used in these sentences were selected to allow for measurement of various specific phonological features of each language, such as duration of long versus short vowels in Croatian and of tense versus lax vowels in English, voice onset time, and vowel quality, etc., in both languages. In this paper, we focus on speaking rate, pitch range, and the vowel space characteristics. We will explore the effect of clear speech on other language-specific phonological contrasts in a future paper. Example sentences are given in (1) for each language. Keywords used for identification scores in the listening test are underlined.

(1)
    (a)   Croatian: *Nada će dobiti tri dokaza i puni mjesec.*
                "Nada will get three proofs and a full moon."
    (b)   English: *Peter and his chief ticket were hooded by their bed.*

Croatian and English sentences were of similar length: the mean number of syllables was 12.8 (range 10–16) and 11.7 (range 9–14) in Croatian and English, respectively. Each sentence contained 4 keywords, giving a total of 80 keywords per set for scoring in the listening test. In order to ensure that the perception was not confounded by the listeners' lack of familiarity with some words, the average familiarity rating of the English keywords used in the perception experiment was 6.87/7, ranging from 5.5 to 7. These familiarity ratings were taken from the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984). No parallel familiarity ratings were available for Croatian. However, all five native talkers were asked after the recording sessions if they thought any of the words were unusual or unfamiliar to them. They reported no such words. Therefore, we were confident that any possible differences in

the results between the two languages could not be attributed to differences in the familiarity of the test words.

C. Procedure

1. Production

All English and Croatian talkers were recorded producing all 20 semantically anomalous sentences in their native language in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. The participants read the sentences, which were written on index cards, into a microphone directly to disk at 24-bit accuracy using an Apogee PSX-100 A/D D/A converter at a sampling rate of 16 kHz. Participants read 20 sentences in their native language once in conversational and once in clear speech. For the conversational style, the talkers were instructed to read as if they were talking to someone familiar with their voice and speech patterns. For the clear speaking style, the talkers were instructed to read as if they were talking to a listener with a hearing loss or a non-native speaker.[2] Sentences were randomized for each reading. This yielded a total of 40 sentences per speaker and 200 per language. The acoustic analyses of the recorded sentences were done using PRAAT software for speech analysis (Boersma, 1996).

2. Perception

After the recordings were made, the digital speech files were segmented into sentence-length files. In order to obtain equivalent overall amplitude levels, all speech files were equated for rms amplitude and then mixed with broadband white noise at a 0 dB signal-to-noise ratio. The signal-to-noise ratio used in this study was chosen based on pilot testing. Each sentence was preceded by a 400 ms leading silence and a 500 ms noise interval, and followed by a 500 ms noise interval.

Each participant in the perception experiment heard a total of 20 sentences in their native language produced by only one of the talkers. Half of the sentences heard were in conversational style and half in clear style for each talker condition. The listeners never heard the same sentence twice. In each talker condition, clear speech sentences preceded conversational sentences so that any clear speech benefit obtained could not be explained by the subject's adaptation to the task or to the talker's speech patterns. Furthermore, the sentences were counterbalanced for style, i.e., for each talker the ten sentences that were presented in the conversational style in one condition (in which half of the listeners participated) were presented in the clear speaking style in another condition (in which the other half of the listeners participated). Four Croatian listeners per talker (20 altogether) and six English listeners per talker (30 altogether) participated in the experiment. English subjects were seated in front of a computer in a sound-attenuated booth in the phonetics laboratory in the Department of Linguistics at Northwestern University. Croatian subjects were tested in a quiet room in the English Department at the University of Zagreb. Stimulus presentation was controlled by SUPERLAB PRO 2.01. The audio files were played through the computer sound card over headphones at a comfortable listening level set by the experimenter before the start of the experiment. Three practice sentences (from a different talker) were presented first so that the subjects could get used to the nature of the stimuli mixed with noise and the procedure of advancing to the next trial. After each trial, the subject pressed the space bar on the keyboard to initiate the next trial. Each trial was presented only once but the duration of the pause between two trials was controlled by the subjects themselves. They could take as long as they needed to record their responses. The listeners were instructed to write down every word they heard. The experimenter left the room/booth after the practice sentences.

D. Data analysis

1. Production

In order to investigate what articulatory modifications talkers adopted in clear speech production, we performed a series of comparable acoustic analyses in both languages. Previous research has established that in English, clear speech involves a wide range of acoustic/articulatory adjustments, such as a decrease in speaking rate, an increase in pitch range, as well as an increase in the acoustic distance between vowels (Picheny et al., 1986; Moon and Lindblom, 1994; Johnson et al., 1993; Ferguson and Kewley-Port, 2002; Bradlow et al., 2003). Following these findings, the specific acoustic-phonetic parameters that we targeted in this analysis were speech rate (overall sentence duration and number and duration of pauses), pitch range (difference between the highest and lowest tonal targets in the sentence), and vowel space expansion. All the acoustic measurements were performed on the exact same sentences that were used in the sentence-in-noise perception tests, i.e., the comparisons were made between the conversational and clear speech styles for each talker.

2. Perception

Each participant in the sentence-in-noise perception test received a keyword-correct score out of 40 for the 10 sentences they heard in each style (conversational versus clear). We adopted a strict scoring criterion for both languages. A keyword was counted as correct only if all morphemes of the target word were present and transcribed correctly, e.g., if the target word was "keeping," "keep, keeps, or kept" were scored as incorrect. Percentage-correct scores were calculated and then converted to rationalized arcsine transform units (RAU) (Studebaker, 1985). The transformed scores where then coded as RAU scores for 0 dB signal-to-noise ratio conversational style and for 0 dB signal-to-noise ratio clear style for each talker in each language. It has to be noted that the strict scoring criterion that we adopted might have penalized Croatian listeners more, since there are seven nominal cases and three grammatical genders in Croatian (in addition to two numbers) that often differ only in the word-final vowel. Therefore, the opportunity for transcribing a wrong affix was larger in Croatian. Nevertheless, we adhered to the strict scoring criterion as it was easy to apply consistently and objectively across both languages. Furthermore,

Talker	Intelligibility (RAU)			
	Conv.	Clear	Cl-Conv.	Diff (proportion conv.)
EF3	37.38	72.65	35.27	0.94
EF2	41.72	59.83	18.11	0.43
EF1	47.62	66.69	19.07	0.40
EM2	46.09	59.43	13.35	0.29
EM1	58.68	52.42	−6.26	−0.11
Average	46.30	62.20	15.91	0.39
CM1	39.38	65.19	25.81	0.66
CM2	52.34	77.57	25.23	0.48
CM3	60.79	71.21	10.42	0.17
CF2	53.50	61.23	7.73	0.14
CF1	42.37	47.08	4.71	0.11
Average	49.68	64.46	14.78	0.31

our primary interest here was in clear versus conversational speech intelligibility within each language, rather than the absolute levels of intelligibility across the two languages.

III. RESULTS

A. Perception

The average sentence perception scores (in RAU) as well as the average intelligibility gain (as a difference between clear and conversational scores and as a proportional increase relative to the conversational score) for all English and Croatian talkers are given in Table I. The talkers in all tables are ordered by the amount of conversational-to-clear speech intelligibility gain as a proportion of the conversational intelligibility score (clear-conversational/conversational). Letters E and C in talker labels stand for English and Croatian languages, respectively. The following F or M designates a female or male talker. The numbers refer to the recording order.

As seen in the table, the effect of style is quite robust overall. That is, in both languages listeners performed better in the clear speech condition than in the conversational speech condition. This pattern was consistent for all talkers except for EM1. Talker EM1 already received the highest intelligibility score of all English talkers in conversational speech, and the clear speech modifications did not result in a further intelligibility gain. The results, furthermore, showed that there was substantial variability across the talkers, both in the level of intelligibility in conversational style and in the amount of benefit afforded by the clear speaking style. ANOVA results for the effect of language (English versus Croatian) and style (conversational versus clear) on RAU scores supported these impressions. There was a main effect of style $[F(1,8)=14.611, p<0.01]$. The effect of language

and the language by style interaction were not significant. In summary, these results show that listeners recognized words more accurately in clear than in conversational speech in their native language. These findings expand our knowledge about clear speech by showing that the clear speech intelligibility effect is not specific to English.

B. Acoustic analyses: Global characteristics

1. Speaking rate

An increase in sentence duration typically accompanies the change in speaking style from conversational to clear (e.g., Picheny et al., 1986; Bradlow et al., 2003). It is not clear, however, how much or even whether speaking rate correlates with intelligibility. In studies of inter-talker variability in English conversational speech intelligibility, overall speaking rate either showed no correlation with overall intelligibility (Bradlow et al., 1996) or correlated with overall intelligibility for some but not all talkers (Hazan and Markham, 2004). Furthermore, Krause and Braida (2002) demonstrated that English clear speech can be produced at normal/conversational speaking rates with the concomitant intelligibility benefit. However, only a slight intelligibility benefit for hearing-impaired listeners was obtained from English clear speech at conversational speaking rates when compared with clear speech at slow speaking rates (Krause, 2001). These findings suggest that, while not crucial, the decrease in speaking rate typical of English clear speech production could be an important contributing factor to the intelligibility of clear speech. We also expect that any intelligibility benefit associated with a decrease in overall speaking rate for clear speech should be independent of the phonological structure of the language, and therefore should be similar across languages.

In order to assess the contribution of pause insertion and of individual segment lengthening to the difference in speaking rate for clear versus conversational speech, we counted the number of pauses and measured their duration. Next, we calculated the number of syllables produced per second after the pauses were excluded. A pause was defined as a period of silence of at least 5 ms in duration excluding silent periods before word-initial stop consonants where it would be impossible to determine the end of a pause and the beginning of the stop closure (similar to Bradlow et al., 2003). Table II shows speaking rate and pause results for all talkers in each speaking style.

All talkers but one (CF1) increased the number of pauses in clear speech as well as their duration. Most talkers, in fact, made no pauses in conversational speech. The average increase in the total number of pauses in clear speech (excluding CF1) ranged from 3 for EF2 to 18 for EF1 and CM1, with the average across all talkers being 9.5. Additionally, for all talkers the average pause duration was longer in clear speech than in conversational speech: the average pause duration increase in clear speech was 0.12 s, ranging from 0.052 s for CM2 and EF2 to 0.205 s for EF1. As an exception, CF1 had fewer pauses (3) in clear speech when compared to conversational speech (5). However, even for this talker the average duration of pauses in clear speech was

TABLE II. The total number of pauses, average pause duration (s), speaking rate (syll/s), and pitch range (semitones) results for each talker in two speaking styles in English and Croatian.

Talker	# of pauses			Pause duration (s)			Speaking rate (syll/s)				Pitch range (semitones)			
	Conv.	Cl.	Diff.	Conv.	Cl.	Diff.	Conv.	Cl.	Cl-Con.	Diff. (prop. conv.)	Conv.	Cl.	Cl-Con.	Diff. (prop. conv.)
EF3	0	8	8	0	0.11	0.11	5.98	4.22	−1.76	0.29	8.89	11.35	2.46	0.28
EF2	0	3	3	0	0.05	0.05	5.43	4.21	−1.22	0.22	12.06	13.34	1.28	0.11
EF1	0	18	18	0	0.21	0.21	5.64	4.07	−1.57	0.28	13.96	14.86	0.90	0.06
EM2	0	8	8	0	0.18	0.18	5.13	3.59	−1.54	0.30	11.44	11.34	−0.10	−0.01
EM1	0	5	5	0	0.12	0.12	4.90	3.71	−1.19	0.24	5.77	6.18	0.42	0.07
Average	0	8.4	8.4	0	0.13	0.13	5.42	3.96	−1.46	0.27	10.42	11.41	0.99	0.10
CM1	0	18	18	0	0.17	0.17	5.85	3.56	−2.28	0.39	9.57	11.82	2.25	0.24
CM2	2	19	17	0.14	0.19	0.05	5.59	4.10	−1.48	0.27	9.27	12.90	3.63	0.39
CM3	0	8	8	0	0.12	0.12	5.45	4.43	−1.02	0.19	7.72	8.91	1.19	0.15
CF2	5	17	12	0.14	0.22	0.08	4.68	4.00	−0.68	0.15	10.01	11.53	1.52	0.15
CF1	5	3	−2	0.07	0.20	0.12	5.80	4.14	−1.66	0.29	6.90	7.05	0.14	0.02
Average	2.4	13	10.6	0.07	0.18	0.11	5.47	4.05	−1.43	0.25	8.69	10.44	1.75	0.19

longer than in conversational speech. The average increase in pause duration in clear speech is 0.12 s for this talker. Furthermore, a review of the recordings suggested that the pauses in the conversational style, for this talker, were due to hesitations caused by unfamiliarity with the read sentences rather than by deliberate pausing due to the speaking style.

The speaking rate results showed that all talkers produced fewer syllables in clear speech when compared to conversational speech, indicating that the change in speaking rate was not due entirely to the insertion of pauses. On average, the talkers produced 1.44 syllables/s less in clear than in conversational speech. The decrease in the number of syllables produced in the clear speaking style when compared to the conversational speaking style ranged between 0.68 syll/s for CF2 and 2.28 syll/s for CM1. ANOVA results for the effect of language (English versus Croatian) and style (conversational versus clear) on speaking rate showed a significant main effect of style [$F(1,8)=94.713$, $p<0.0001$], but not of language. The language by style interaction was not significant either. Overall, the present results support previous findings that clear speech production is characterized by longer segmental durations as well as by insertion of more and longer pauses. Furthermore, the comparison between the languages shows that the change in overall speaking rate for clear speech relative to conversational speech is similar in both languages.

2. Pitch range

Fundamental frequency ($F0$) is another global acoustic-phonetic parameter that differs across talkers, genders, and speaking styles. Bradlow et al. (2003) and Picheny et al. (1986) have shown that $F0$ range is increased in clear speech for most talkers. However, similar to the speaking rate findings, it is not clear that $F0$ range directly affects intelligibility. Nevertheless, an increase in pitch range is a hyperarticulation feature that appears to be one of many English clear speech characteristics. We measured the range between the highest and lowest $F0$ points in hertz for each sentence in the two speaking styles. The hertz values were converted into semitones for ease of comparison across different pitch ranges. Mean pitch range results along with the pitch range increase/decrease in clear speech for all talkers in both speaking styles are given in Table II.

The average pitch range expansion was 21.22 Hz or 1.26 semitones for the female talkers, and 15.03 Hz or 1.48 semitones for the male talkers. However, there was considerable across-talker variability in the amount of pitch range expansion. Three out of ten talkers, CF1, EM1, and EM2, expanded pitch range less than 5 Hz (for EM2 the pitch range is even slightly reduced in clear speech, although the amount is negligible). Seven talkers, however, showed larger pitch range increases in clear speech. The average pitch range expansion for these talkers ranged between 9.151 and 40 Hz (for CM3 and CM2, respectively) or between 0.899 and 3.63 semitones (for EF1 and CM2, respectively). Closer examination of low and high $F0$ targets revealed that the strategy for achieving the pitch range expansion was similar across talkers in both languages, i.e., $F0$ high targets were raised while low $F0$ targets (the bottom of the talker's pitch range) remained relatively fixed. ANOVA results for the effect of language (English versus Croatian) and style (conversational versus clear) on pitch range showed a significant main effect of style [$F(1,8)=14.292$, $p<0.005$], but not of language. The style by language interaction was not significant. The results showed that, in the clear speaking style, talkers tended to increase their pitch range, and they did so to a similar degree in both languages.

In summary, the results of this examination of clear speech in terms of global acoustic measures demonstrated that a decrease in overall speaking rate, as reflected by both the number and duration of interword pauses and syllables-per-second production rate, and an increase in pitch range were consistent features of the conversational-to-clear speech transformation across talkers and across languages.
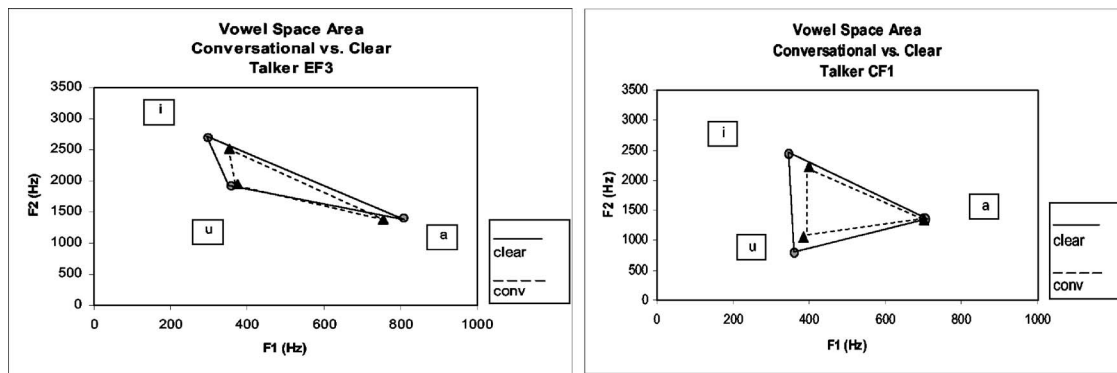
FIG. 1. Vowel space area measured as the Euclidean area covered by the triangle defined by the mean of each vowel category for the talker who exhibited the largest effects of vowel space expansion in each language. Solid lines connect the clear speech vowels (circles) and dashed lines connect the conversational speech vowels (triangles).

C. Acoustic analyses: Vowel space characteristics

In addition to the global acoustic changes for clear speech relative to conversational speech, we wanted to investigate fine-grained, acoustic-phonetic characteristics of clear speech as well. The present vowel space analyses aim to assess the contributions of language-specific and general/universal principles in the acoustic realization of vowel categories. In particular, we explore whether the language-specific phonological property of the number of vowel categories in a language influences the extent of clear speech vowel space expansion.

Vowel hyperarticulation in English has been associated with an intelligibility advantage on the basis of intertalker differences in overall intelligibility within normal, conversational speech (Byrd, 1994; Bond and Moore, 1994; Bradlow, Torretta, and Pisoni, 1996; Hazan and Markham, 2004) as well as on the basis of clear versus conversational style comparisons (Picheny *et al.*, 1986; Moon and Lindblom, 1994; Bradlow *et al.*, 2003; Krause and Braida, 2004). Thus, there is strong support for the claim that vowel space expansion is an intelligibility-enhancing strategy available to English talkers. However, in light of results concerning language-specific effects on vowel space expansion in infant- versus adult-directed speech and in narrow versus broad focus conditions discussed above (Andruski *et al.*, 1999; Hay *et al.*, 2003; Coren and Heckmann, 2004), we were not certain whether Croatian talkers would adopt this same vowel expansion strategy under clear speech production conditions since Croatian has a small (5) vowel inventory in contrast to the large (14) vowel inventory of English. In keeping with the infant-directed and narrow focus findings, and consistent with the Theory of Adaptive Dispersion (Lindblom, 1986; Diehl and Lindblom, 2002), one might predict that English clear speech will show greater vowel space expansion than Croatian. However, Bradlow (2002) found that the high vowels /i/ and /u/ in English and Spanish both showed similar peripheralization towards the vowel space extremes in the two languages, suggesting that clear speech production may involve hyperarticulation for all vowels regardless of the vowel inventory size. In the present comparison, we selected three peripheral vowels (i, a, u) to evaluate the vowel space characteristics in English and Croatian. These vowels can

potentially show the largest amount of articulatory modification since making them more peripheral, i.e., expanding the vowel space, is not limited by encroaching on other vowel categories but only by the limits of the vowel space itself.

F1 and F2 frequencies were taken from the midpoint of each vowel. All formant measurements were made automatically using an LPC formant tracking algorithm in PRAAT. Values that differed by more than 200 Hz from the mean for the category were hand checked and corrected if necessary. Three measures of the relationship between the speaking style and vowel space were used: vowel space area, vowel space dispersion (both following Bradlow *et al.*, 1996) and vowel peripheralization. Vowel space area was measured as the Euclidean area covered by the triangle defined by the mean of each vowel category. Vowel space dispersion was measured as the distance of each vowel from the central point in the talker's $F1 \times F2$ space. An overall vowel space dispersion value for each talker was then calculated as the mean of these distances. Finally, the extent of peripheralization in clear speech relative to conversational speech was measured for each vowel category separately. Here, the Euclidean distance in the $F1 \times F2$ space between the average location of each vowel in conversational and in clear speech was calculated separately for each talker and each vowel category. Figures 1–3 illustrate these three measurements for the talker who exhibited the largest effects of vowel space expansion in each language.

We included these three measures for several reasons. First, we wanted to see which measure would best capture vowel space expansion in the two languages. Conversational-to-clear vowel space modification could be of smaller magnitude in Croatian than in English. It is possible that this difference will only be evident in some but not all measures. Different measurements could also reveal if the two languages use similar or different strategies for vowel space expansion. For instance, both languages could overall cover a larger vowel space area in clear speech, but it could be the case that this is achieved through making /a/ more open in one language versus fronting /i/ and/or backing /u/ in another. Second, Bradlow *et al.* (1996) found that intelligibility scores correlated better with the vowel space dispersion measure than with the vowel space area measure, pos-
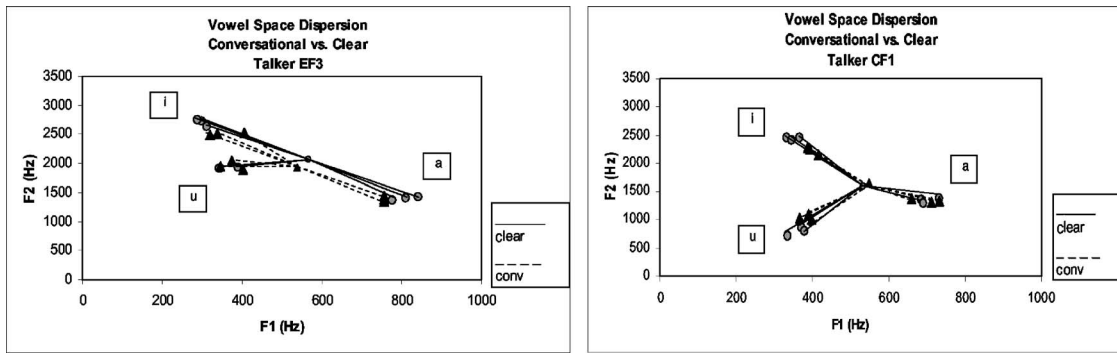
FIG. 2. Vowel space dispersion measured as the distance of each vowel from the central point in the talker's $F1 \times F2$ space for the talker who exhibited the largest effects of vowel space expansion in each language. Solid lines connect the clear speech vowels (circles) and dashed lines connect the conversational speech vowels (triangles).

sibly due to the fact that the latter measure is based on category averages and not the individual vowel tokens produced, as is the case with the vowel dispersion measure. Finally, vowel space area and dispersion both capture the overall vowel space covered by the three point vowels in the two speaking styles. In order to assess the magnitude of pe-

ripheralization of each individual vowel category, we included the third measure as well. Since each of these measurements addresses a slightly different aspect of a talker's clear speech production characteristics, we hoped to gain better insight into cross-language vowel space production strategies in different speaking styles.
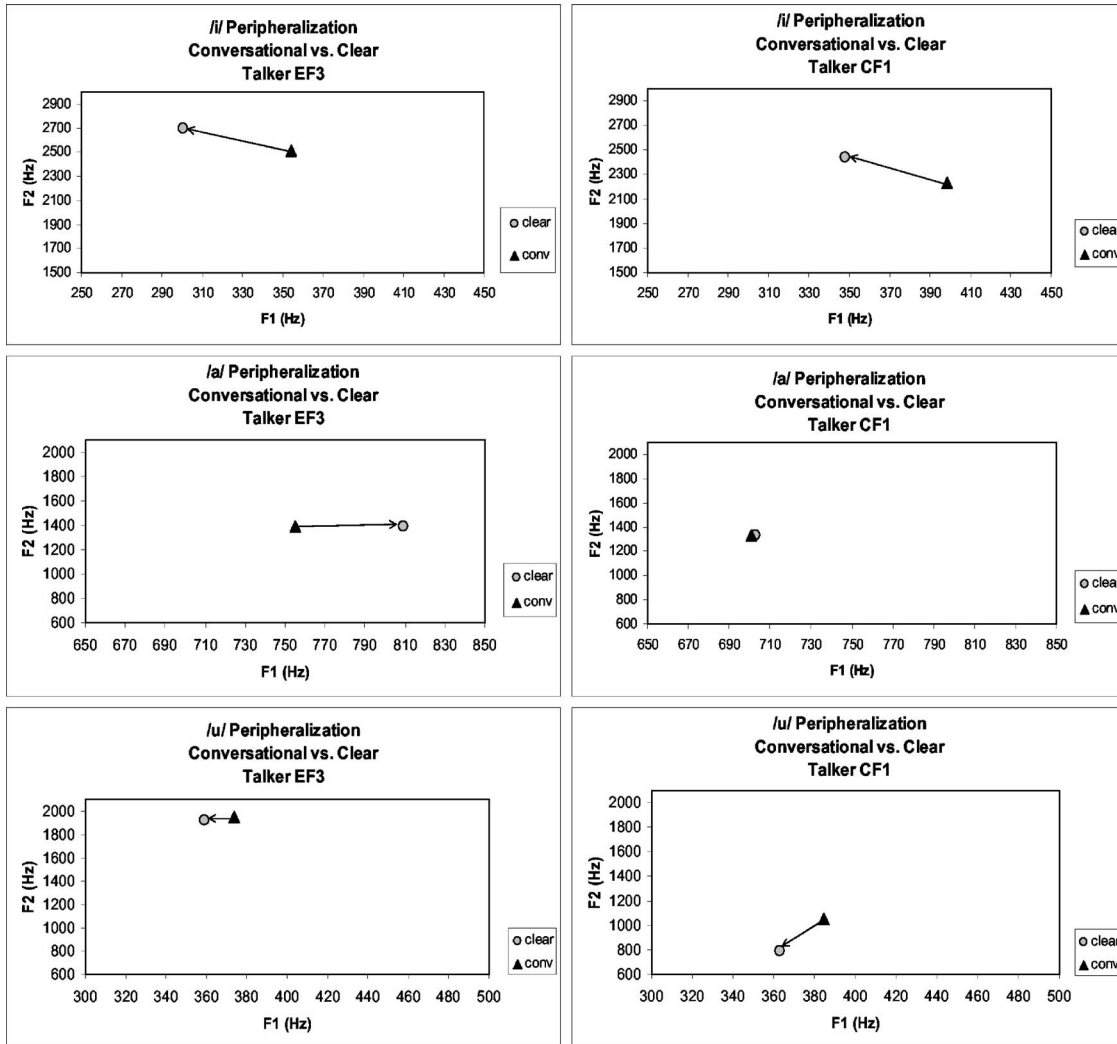
FIG. 3. Vowel peripheralization measured as the Euclidean distance in the $F1 \times F2$ space between the average location of each vowel in conversational and in clear speech for the talker who exhibited the largest effects of vowel space expansion in each language. Peripheralization is given for each vowel separately so that the scale (in hertz) could be extended.

TABLE III. Vowel space area (Hz²), vowel space dispersion (Hz), and vowel peripheralization (Hz) results for each English and Croatian talker in two speaking styles. Negative sign indicates vowel space area reduction and a move towards the vowel space center rather than towards the periphery for talker EF1.

Talker	Vowel space area (Hz²)				Vowel space dispersion (Hz)				Vowel space peripheralization (Hz) in clear speech (relative to conv.)			
	Conv.	Clear	Cl-Conv.	Diff. (prop. conv.)	Conv.	Clear	Cl-Conv.	Diff. (prop. conv.)	a	i	u	Ave.
EF3	101 048.06	159 701.50	58 653.44	0.58	457.17	549.52	92.35	0.20	54.39	191.50	37.80	94.57
EF2	116 190.42	167 536.05	51 345.63	0.44	513.62	595.70	82.09	0.16	67.01	144.98	35.80	82.60
EF1	235 006.41	221 917.33	−13 089.09	−0.06	633.96	648.83	14.87	0.02	32.36	64.92	−68.10	9.73
EM2	157 295.10	203 727.19	46 432.09	0.30	444.15	508.21	64.06	0.14	21.00	86.42	85.15	64.19
EM1	83 843.63	141 049.68	57 206.05	0.68	436.50	509.62	73.12	0.17	53.09	67.75	185.31	102.05
Avg.	138 676.72	178 786.35	40 109.62	0.39	497.08	562.38	65.30	0.14	45.57	111.12	55.19	70.63
CM1	121 897.66	197 783.65	75 886.00	0.62	420.21	508.72	88.51	0.21	39.86	75.71	186.29	100.62
CM2	164 867.28	246 607.29	81 740.00	0.50	489.70	549.09	59.38	0.12	67.94	65.14	91.64	74.91
CM3	137 553.06	178 709.15	41 156.10	0.30	493.70	522.40	28.70	0.06	28.56	53.52	46.39	42.82
CF2	488 133.66	585 893.24	97 759.58	0.20	791.85	815.35	23.50	0.03	81.52	54.71	110.12	82.12
CF1	184 795.81	284 005.05	99 209.24	0.54	522.64	663.28	140.64	0.27	4.63	214.11	259.18	159.30
Avg.	219 449.49	298 599.68	79 150.18	0.43	543.62	611.77	68.15	0.14	44.50	92.64	138.72	91.95

The results for vowel space area, dispersion, and peripheralization for all talkers are given in Table III. We will discuss the results of each measurement beginning with the vowel space area. Although there is variability in the average vowel space area across the talkers in both speaking styles, talkers in both languages did indeed expand the vowel space in clear speech. ANOVA results for the effect of language (English versus Croatian) and style (conversational versus clear) on vowel space area showed a significant main effect of style $[F(1,8)=48.691, p<0.0001]$, but not of language. The style by language interaction was almost significant $[F(1,8)=5.218, p=0.052]$. This almost-significant interaction was due to the one negative change, i.e., a decrease in the clear speech vowel area, in English, which was for the talker with the biggest conversational vowel space area (EF1). The statistical analysis thus supports previous findings that vowel space expansion is a correlate of the conversational-to-clear speech transformation. clear speech modifications, therefore, involve both global enhancements such as a decrease in speaking rate and pitch range expansion as well as the enhancement of phonological properties of a language, i.e., making the contrastive vowel categories more distinct from each other. In addition, these results show that talkers expanded the vowel space area equally in clear speech in both languages despite their different vowel inventories. In other words, in hyperarticulated clear speech the peripheral vowel categories were made more extreme, thereby utilizing a larger vowel space area even in a language with only five vowel categories.

Next, we turn to the second measurement of vowel space expansion, i.e., dispersion of individual vowels in the vowel space. The results of each individual talker's vowel space dispersion in the two speaking styles are given in Table III. As expected, given the vowel space area results, for most talkers vowel space dispersion was greater in clear than in conversational speaking style. That is, vowels were more peripheral from the central point of the talker's vowel space.

Moreover, this greater vowel space dispersion for clear speech appears to be similar in both languages. ANOVA results for the effect of language (English versus Croatian) and style (conversational versus clear) on vowel space dispersion showed a significant main effect of style $[F(1,8)=27.577, p<0.001]$, but not of language. The style by language interaction was not significant. These results corroborate the findings for vowel space area in that clear speech was characterized by greater vowel dispersion than conversational speech, and the effect was similar in both languages.

Finally, we looked at the amount of vowel peripheralization for each category separately in clear speech. This measure differs from the previous two in that it calculates the Euclidean distance in the $F1 \times F2$ space between the average token of a single vowel in conversational and clear speech styles. The results are shown in Table III. ANOVA showed no significant effect of language (English versus Croatian) or vowel (/a/ versus /i/ versus /u/) on the amount of vowel peripheralization in clear speech. The language by vowel interaction was also not significant. Overall, all three point vowels were peripheralized to a similar degree in both languages, despite the fact that the potential for perceptual confusion between separate vowel categories is smaller in Croatian with 5 vowels than in English with 14 vowels. This measurement, therefore, shows that talkers in both English and Croatian made all of their vowel productions more extreme in hyperarticulated clear speech. Furthermore, all of the vowel space results combined demonstrate that the three measurements adequately and similarly captured the vowel space expansion patterns in both languages. However, it is not clear that any of the three measurements was a better indicator of talker intelligibility (see the discussion below).

If we look more carefully at the individual talker's vowel spaces, we notice that talkers adopted somewhat different strategies in achieving the expansion. For instance, both EF3 and CF1, talkers with the largest amount of vowel space expansion in each language (as reflected in the clear-

conversational difference score for the vowel space area measure in Table III), produced higher and more front /i/ vowels in clear speech (as reflected by the value for /i/ peripheralization in Table III). Only EF3, though, lowered her jaw more and produced a lower, more open vowel /a/. CF1, on the other hand, produced a more retracted /u/ in clear speech. In general, there was a tendency for English talkers to retract /u/ in clear speech less than Croatian talkers. This is possibly due to the fact that /u/ in general American English is fairly fronted and making it more back would not necessarily make it a "better" realization of the vowel category. However, at least one English talker, EM1, retracted /u/ to a large degree (comparable to the amount of /u/ retraction in Croatian). Other talkers peripheralized vowels to different degrees. Unlike CF1, Croatian talkers CF2 and CM2, for example, produced a much more open vowel /a/. Similarly, English talkers EF3 and EF2 produced higher and more front /i/, but EF1 and EM1 did this to a lesser degree, etc. Although the articulatory strategies were different for different talkers, they all added up to the overall expansion of the vowel space with a larger distance between the contrastive categories. Most importantly for our present purposes, the overall expansion for clear speech relative to conversational speech and, in general, the strategies adopted to achieve this expansion, were language independent and seemed to be consistent regardless of the number of vowel categories in the language.

IV. DISCUSSION

In this paper, we set out to perform a cross-language comparison of clear speech perception and production in two typologically unrelated languages, namely Croatian and English. The overall goal was to identify acoustic-phonetic features that characterize the conversational-to-clear speech modifications in the two languages, and to establish whether these articulatory/acoustic adjustments are associated with an intelligibility gain.

The results showed that, although some talkers were more successful than others in improving their intelligibility, most talkers modified their speech production in clear speech in such a way that benefits the listener. Importantly, conversational-to-clear speech transformations in Croatian and English resulted in equal intelligibility benefits for Croatian and English listeners, respectively. These findings suggest that naturally produced clear speech is an effective way of enhancing speech perception under adverse speaking conditions regardless of the language involved.

We also found evidence that clear speech production is guided by both auditory-perceptual factors as well as by contrast enhancement mechanisms. Talkers of both languages enhanced the global salience of the speech signal by decreasing their speaking rate (both by producing longer segments and by inserting more frequent and longer pauses) and by expanding the pitch range. In addition, by producing more peripheral vowels in clear speech, the talkers enlarged the distance between the contrastive vowel categories which made them more distinct and perceptually less confusable. The results demonstrated that vowel space expansion was applied equally by talkers of both languages, suggesting that vowel contrast enhancement in hyperarticulated clear speech is independent of vowel inventory size.

These findings are consistent with the finding of Bradlow *et al.* (2003) that English and Spanish high vowels are more peripheral in clear than in conversational speech. However, these results differ from those obtained by Hay *et al.* (2003) and Coren and Heckmann (2004), who found no significant vowel space expansion in Japanese and substantial vowel space expansion in German for vowels in words in narrow focus versus in broad focus. These authors interpreted this cross-language difference in vowel space expansion as reflecting the cross-language difference in vowel inventory size, which is small for Japanese (5) and large for German (16). Their results were taken as evidence that hyperarticulation is sensitive to language-specific phonological properties. The question that the present data raise, then, is how can we explain the difference between Japanese and Croatian, two languages with similar and relatively small vowel inventories, in terms of their propensity to expand the vowel space under conditions of hyperarticulation?

It is possible that clear speech and narrow focus are two distinct speaking modes representing different types and/or degrees of hyperarticulation. Hyperarticulation in narrow focus is limited to one word, while clear speech affects the entire discourse. Furthermore, clear speech is specifically intended to enhance intelligibility, while narrow focus marks new or contrastive information. It is conceivable, then, that in narrow focus those cues that are already sufficiently distinct, such as $F1 \times F2$ vowel space characteristics in Japanese, are not enhanced. However, narrow focus and clear speech share numerous other enhancement characteristics, such as larger pitch excursions, vowel lengthening, vowel length contrast enhancement, etc. It therefore seems somewhat unlikely that vowel space expansion, if available as an enhancement strategy, would not accompany these other transformations (Smiljanić, 2004, 2005; Uchanski 1988, 1992; present study). If the difference between the two speaking styles is one of degree, i.e., if clear speech involves "stronger" hyperarticulation due to its intelligibility-enhancing nature, Japanese talkers may expand the vowel space in clear speech despite the absence of expansion in narrow focus. Conversely, Croatian talkers may show no vowel space expansion in narrow focus, in contrast to the vowel expansion of their clear speech productions. Such varied contrast enhancement strategies within a language also do not seem likely to us. We conducted a preliminary analysis of vowel space expansion in narrow focus in Croatian on data used in Smiljanić (2004). Although these data were designed and collected primarily for investigation of focus effects on vowel duration and pitch accents and are, therefore, not completely comparable to the data discussed here, the results indicated that talkers expanded the vowel space in narrow focus in much the same way as they expanded the vowel space in clear speech. Furthermore, based on the reported absence of vowel space expansion in Japanese infant-directed speech (Andruski *et al.*, 1999), it is likely that Japanese talkers would *not* expand their vowel space in clear speech. If, indeed, we do find such consistent behavior

within a language with regard to vowel space expansion, a cross-language comparison of Japanese, on the one hand, and Croatian (and possibly other languages with a five-vowel system such as Russian or Spanish) on the other, could provide further insight into the role played by inventory size in determining specific hyperarticulation strategies.

Although we have identified some acoustic-phonetic conversational-to-clear style transformations, in this study we have not attempted to relate these articulatory modifications directly to intelligibility. Our database of spoken sentences and intelligibility scores simply does not include enough talkers to provide for a meaningful assessment of the relationship between acoustic-phonetic variation and variability in intelligibility. Nevertheless, for the sake of future database development, we mention some tendencies that we observed in our database.

The two talkers who exhibited the highest intelligibility gain in clear speech were EF3 and CM1. For these talkers, conversational-to-clear speech modifications included a rather large speaking rate decrease and pitch range expansion compared to other talkers. Furthermore, both of these talkers exhibited a large vowel space area expansion. These modifications seem to conspire together to award these talkers the largest increase in intelligibility. It is equally informative to look at the two talkers with the smallest amount of intelligibility gain, namely EM1 and CF1. One would expect that the absence of the articulatory enhancements found for EF3 and CM1 will characterize EM1's and CF1's speech patterns. However, this is not quite what we found. EM1 did show a very small decrease in speaking rate as well as a rather small pitch range increase in clear speech. However, this talker showed a fairly large vowel space expansion. Moreover, this talker's vowel space in conversational speech was the smallest of all English talkers and, despite the large amount of expansion, his vowel space remained rather small in clear speech in comparison to the other talkers. Similarly, CF1 showed a very small pitch range increase with a narrow pitch range in both speaking styles. This talker decreased the speaking rate quite a bit, but was still the second fastest talker in both speaking styles. Finally, despite showing the largest vowel space expansion, CF1 had very small vowel space areas in both speaking styles when compared to the other Croatian female talker. As expected, her conversational intelligibility score was second lowest and her clear speech intelligibility score was the lowest. For this set of talkers, then, the largest expansion of the vowel space does not result in an equivalently large improvement in intelligibility since all vowels remain more centralized, which seems to be detrimental for the listeners. Finally, a converse pattern can be seen for talker EF1, who showed the least amount of vowel space expansion in clear speech. In fact, for this talker, there was a slight reduction in the vowel space area for clear speech relative to conversational speech. However, this talker had the largest vowel space areas in both speaking styles and the second highest intelligibility scores in both speaking styles. For this talker, the intelligibility gain was not the largest but overall intelligibility was very high in both speaking styles. It is possible that this talker produced some-

what hyperarticulated speech in her conversational readings compared to other talkers, leaving very little "room for improvement" in clear speech.

These overall patterns provide further support for the claim that variation along multiple articulatory parameters, such as speaking rate, pitch range, and vowel space area, all contribute to variability in overall intelligibility (e.g., Bradlow *et al.*, 1996; Hazan and Markham, 2004). Furthermore, within-talker enhancements of these various acoustic cues generally seem to result in improved intelligibility. However, the magnitude of the transformation is closely linked to the individual's speech patterns within a speaking style. That is, talkers who are highly intelligible in the conversational speaking style might not modify their speech much, and therefore do not have a large intelligibility gain in clear speech. This suggests that there is a limit to the benefit afforded by these enhancement strategies. In other words, enhancing the acoustic-phonetic cues beyond a certain amount, such as slowing down further or expanding vowel space even more, might result in speech that sounds unnatural and is therefore less intelligible. Finally, the fact that talker EM1 shows the smallest vowel space area and the narrowest pitch range in the conversational speaking style and still receives the highest intelligibility score for this speaking style demonstrates that the acoustic cues discussed above are not the only cues relevant to high intelligibility. Indeed, other studies have pointed to various other cues not examined here (Picheny *et al.*, 1986; Krause and Braida, 2004; Bradlow *et al.*, 2003; Liu *et al.*, 2004; Moon and Lindblom, 1994; Hazan and Markham, 2004).

V. CONCLUSIONS

The overall goal of this study was to confirm that the clear speech intelligibility effect was not an English-specific phenomenon, and in so doing to test the hypothesis that clear speech production is guided by both general, universal auditory-perceptual factors and phonological contrast enhancement factors. The data provided strong evidence that the general phenomenon of clear speech as a distinct, listener-oriented, intelligibility-enhancing mode of speech production exists in another language besides English. For the acoustic-phonetic features of the conversational-to-clear speech mode transformation examined here, we found that English and Croatian look remarkably similar. In both languages, talkers generally showed a decrease in speaking rate, an increase in pitch range, and an expansion of the vowel space in going from conversational to clear speech. Furthermore, despite the vastly different vowel inventory sizes (>10 for English, 5 for Croatian), the extent of vowel space expansion in the two languages was the same. These data have therefore revealed cross-language *similarities* in clear speech production. What remains for future research is the discovery of systematic, phonologically motivated cross-language *differences* in clear speech productions. To that end, we are currently conducting a series of further analyses of the materials in the present study to determine whether language-specific phonological contrast enhancement (such as enhancement of the phonemic vowel duration contrast of

Croatian, the tense versus lax vowel duration contrast of English, and the two-way voicing category distinction of both Croatian and English) is indeed a guiding principle of clear speech production and a significant source of the substantial intelligibility gain for naturally produced clear speech.

ACKNOWLEDGMENTS

[1]The difference between broad- and narrow-focused words reflects the difference in information content roughly corresponding to old (known to the listener) versus new (not previously known to the listener) information. Narrow focus can also refer to the added emphasis (contrastive information).

[2]It is important to note that both "conversational" and "clear" speech terms refer to read laboratory speech elicited by specific instructions given to talkers rather than in a more naturalistic situation. Although, ultimately, we would like to investigate spontaneously produced clear and conversational speaking styles, we believe that data obtained in this study approximate this goal sufficiently since talkers implemented conversational-to-clear speech articulatory modifications which resulted in significant intelligibility benefits.

Andruski, J. E., Kuhl, P. K., and Hayashi, A. (**1999**). "Point vowels in Japanese mothers' speech to infants and adults," J. Acoust. Soc. Am. **105**, 1095–1096.

Boersma, P. (1996). Available at: ⟨http://fonsg3.let.uva.nl/praat/praat.html⟩.

Bond, Z. S., and Moore, T. J. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Bradlow, A. R. (**2002**). "Confluent talker- and listener-related forces in clear speech production," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin) pp. 241–273.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for learning-impaired children: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech. I. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **15**, 39–54.

Coren, A. E., and Heckmann, C. L. (**2004**). "Contrast enhancement of vowels in Modern Standard German," J. Acoust. Soc. Am. **114**, 2363.

Cutler, A., and Butterfield, S. (**1990**). "Durational cues to word boundaries in clear speech," Speech Commun. **9**, 485–495.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Memory Lang. **33**, 824–844.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Memory Lang. **25**, 385–400.

Diehl, R. L., and Lindblom, B. (**2002**). "Explaining the structure of feature and phoneme inventories," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. Ainsworth, A. Popper, and R. Fay (Springer, New York).

Ferguson, S. H. (**2004**). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," J. Acoust. Soc. Am. **116**, 2365–2373.

Ferguson, S. H., and Kewley-Port, D., (**2002**). "Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **112**, 259–271.

Gagne, J.-P., Rochette, A.-J., and Charest, M. (**2002**). "Auditory, visual, and audiovisual clear speech," Speech Commun. **37**, 213–230.

Hay, J. F., Sato, M., Coren, A. E., and Diehl, R. L. (**2003**). "A cross-language study of vowel sounds produced with and without emphasis: Testing the theory of adaptive dispersion," J. Acoust. Soc. Am. **111**, 2367.

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: Phonetic targets are hyperarticulated," Language **69**, 505–528.

Krause, J. C. (**2001**). "Properties of naturally produced clear speech at normal rates and implications for intelligibility enhancement," unpublished Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.

Krause, J. C., and Braida, L. D. (**2004**). "Acoustic properties of naturally produced clear speech at normal speaking rates," J. Acoust. Soc. Am. **115**, 362–378.

Kuhl, P. K., Andruski, J. E., Chistovich, L., Chistovich, I., Kozhevnikova, E., Sundberg, U., and Lacerda, F. (**1997**). "Cross language analysis of phonetic units in language addressed to infants," Science **227**, 684–686.

Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 13–44.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Netherlands), pp. 403–439.

Liu, S., Del Rio, E., Bradlow, A. R., and Zeng, F. G., (**2004**). "Clear speech perception in acoustic and electrical hearing," J. Acoust. Soc. Am. **116**, 2374–2383.

Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (**1984**). "Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20 000 words," Research on Speech Perception: Progress Report No. 10, Indiana University, pp. 357–376.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1989**). "Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech," J. Speech Hear. Res. **32**, 600–603.

Smiljanić, R. (**2004**). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study*, Outstanding Dissertations in Linguistics, (Routledge, New York).

Smiljanić, R. (**2005**). "Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian," Papers in Laboratory Phonology 8, Yale University and Haskins Laboratories.

Solé, M-J. (**1992**). "Phonetic and phonological processes: The case of nasalization," Lang Speech **35**, 29–43.

Solé, M-J. (**1995**). "Spatio-temporal patterns of velopharyngeal action in phonetic and phonological nasalization," Lang Speech **38**, 1–23.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Uchanski, R. M. (**1988**). "Spectral and temporal contributions to speech clarity for hearing impaired listeners," unpublished Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Uchanski, R. M. (**1992**). "Segment durations in conversational and clear speech," unpublished manuscript.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., and Durlach, N. I. (**1996**). "Speaking clearly for the hard of hearing. IV. Further studies of the role of speaking rate," J. Speech Hear. Res. **39**, 494–509.

Aerodynamic transfer of energy to the vocal folds

Scott L. Thomson,[a] Luc Mongeau, and Steven H. Frankel

Ray W. Herrick Laboratories, Purdue University, West Lafayette, Indiana 47907-1077

The aerodynamic transfer of energy from glottal airflow to vocal fold tissue during phonation was explored using complementary synthetic and numerical vocal fold models. The synthetic model was fabricated using a flexible polyurethane rubber compound. The model size, shape, and material properties were generally similar to corresponding human vocal fold characteristics. Regular, self-sustained oscillations were achieved at a frequency of approximately 120 Hz. The onset pressure was approximately 1.2 kPa. A corresponding two-dimensional finite element model was developed using geometry definitions and material properties based on the synthetic model. The finite element model upstream and downstream pressure boundary conditions were based on experimental values acquired using the synthetic model. An analysis of the fully coupled fluid and solid numerical domains included flow separation and unsteady effects. The numerical results provided detailed flow data that was used to investigate aerodynamic energy transfer mechanisms. The results support the hypothesis that a cyclic variation of the orifice profile from a convergent to a divergent shape leads to a temporal asymmetry in the average wall pressure, which is the key factor for the achievement of self-sustained vocal fold oscillations. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000787]

I. INTRODUCTION

Voice production is a direct consequence of the flow-induced vibrations of the vocal folds (e.g., Flanagan, 1972). With the vocal folds at rest, the lung pressure increases, forcing the vocal folds apart. As the glottis opens, the effect of viscosity becomes negligible and the Bernoulli equation can be used to relate pressure and velocity. In the convergent part of the glottis, the velocity increases and the pressure decreases. In the divergent part of the glottis, the pressure increases up to the point of flow separation, after which it is more or less equal to the supraglottal pressure. As the folds approach each other, viscous effects become dominant, resulting in a decrease in flow velocity. As the flow stops, the pressure in the upstream region of the glottis increases, opposing closure. This pressure again forces the folds apart, and the cycle repeats itself.

Titze (1988) provided a further explanation of this mechanism by indicating that the intraglottal pressure does not necessarily need to "pull" the folds together, but, rather, that the intraglottal pressure must simply be greater during the opening phase than during the closing phase. It was explained that this is necessary in order for there to be a positive net transfer of energy from the airflow to the vocal folds; a positive value is required to overcome damping within the vocal fold tissue (in the absence of damping, the oscillations could be self-sustained with zero net energy transfer). The energy flow from the air stream to the vocal folds is roughly equivalent to the force exerted on the vocal folds by the normal pressure, multiplied by the normal component of velocity. If the average intraglottal pressure is positive during the entire cycle, then the pressure is in phase with the tissue velocity during opening, and out of phase during closing. Therefore, for a positive airflow-to-tissue energy exchange, the average intraglottal pressure must be slightly lower during closing than during opening. Titze postulated that this temporal asymmetry in intraglottal pressure is an essential feature of vocal fold vibration.

Titze further explained that the pressure-velocity phase relationship is primarily governed by two factors: vocal tract inertance and the cyclically alternating converging-diverging medial surface profile. Since excised larynges and other physical models readily self-oscillate without the need for any supraglottal acoustic loading, the converging-diverging motion seems a more plausible cause. An early observation of the contribution of the medial surface vertical phase difference to vocal fold vibration can be found in Ishizaka and Matsuidara (1968). Generally speaking, the glottis profile is convergent during opening and divergent during closing. This is illustrated in Fig. 1. Note that any possible effect of supraglottal loading has been neglected in the following discussion for simplicity. The local air velocity through the convergent orifice increases along the streamwise direction to maintain a constant mass flow rate, i.e., at any given time the velocity must increase in regions where the area decreases. The pressure consequently decreases along the streamwise direction. On the other hand, flow velocity decreases and the pressure increases along the divergent section of the orifice passage. For no supraglottal loading, the pressure at the glottis exit is zero for both geometries, and the net intraglottal pressure is positive in the case of the convergent orifice (since it has to decrease to a value of zero), and negative in the case of the divergent orifice (it has to increase to a value of zero). This creates the necessary pressure asymmetry for self-sustained oscillations. In the divergent orifice, the in-

[a]Now at the Department of Mechanical Engineering, Brigham Young University, Provo, Utah 84602. Electronic mail: thomson@byu.edu
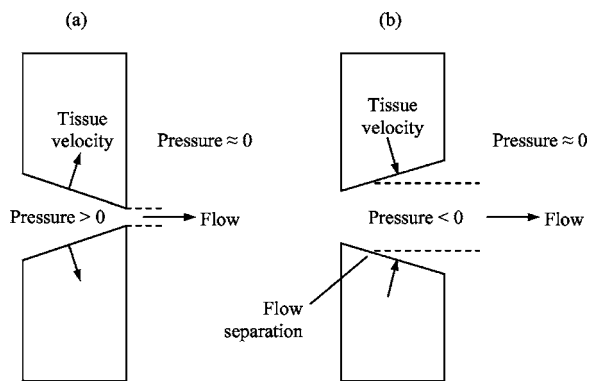
FIG. 1. Schematic of the orifice profile, intraglottal and supraglottal pressures, and normal component of tissue velocity for no supraglottal loading. (a) Opening phase; (b) closing phase.

crease in pressure along the streamwise direction, referred to as an adverse pressure gradient, may result in separation of the boundary layer from the vocal fold surface (illustrated in Fig. 1). The location of the separation point is important because of its effect on surface pressure, but a precise analytical prediction for many orifice flows is not possible. The pressure along the vocal fold surface between the flow separation point and the glottis exit is nearly zero (again, for no supraglottal loading). Therefore, the surface area over which the pressure increases to a value of zero is smaller when flow separation occurs, resulting in a change in the net intraglottal pressure.

This explanation for the energy transfer mechanism from airflow to the vocal folds has been generally accepted, and related studies have been performed. However, no studies involving both the detailed flow and structural data necessary for quantifying the energy transfer to the surface have been reported. To verify this theory, simultaneous cyclic vocal fold wall pressure profile data and vocal fold surface velocity data are required. A self-oscillating model (as opposed to a driven model) must be used. Reduced-order, finite-element, and experimental self-oscillating models have been previously developed to study various aspects of laryngeal fluid-structure interactions.

The two-mass model of Ishizaka and Flanagan (1972) has been most widely used. This model consists of two ideal coupled mass-damper-spring systems driven by the pressure loading on the masses as predicted by Bernoulli's obstruction theory. Numerous improvements of this model have been proposed, including, for example, increasing the number of masses used to represent the vocal folds (Titze, 1973, 1974), incorporating the contribution of the vocal fold body (Story and Titze, 1995), and modifying the geometry of the masses and using additional flow theory to estimate the point of flow separation within the glottis (Pelorson et al., 1994). A summary of the modified two-mass model was presented by Titze (2002). The advantage of this model is its simplicity; its disadvantage is the need for detailed experimental data or more sophisticated models to establish the needed lumped parameter values and orifice coefficients.

Numerical simulations with more detailed flow and solid modeling have also been used. A finite-element model of the vocal folds was developed by Jiang et al. (1998); previously

measured (Jiang, 1991) pressures on the human vocal folds were applied to the model to provide the driving force. This study was used to investigate the influence of vocal fold pathology on vocal fold response, and aerodynamic energy transfer mechanisms were not discussed.

A simulation of glottal flow was performed by Alipour et al. (1996), in which the flow details were solved using the Navier-Stokes equations and the vocal fold motion was prescribed. Direct numerical simulations of the airflow through an orifice with prescribed wall motion were performed by Zhao et al. (2002) and Zhang et al. (2002) to elucidate the different sound sources in the vocal tract. These three studies provided useful information about the nature of glottal and supraglottal flow fields during speech, as well as the acoustics of the sound source, but the energy transfer to the vocal folds was not investigated.

In the numerical simulation studies referenced above, either the loading directly applied to the wall surface or the wall motion itself was prescribed. Other studies have been performed that have included fully coupled fluid and solid responses. Alipour et al. (2000) used continuum vocal fold models coupled with flow predictions to investigate modal entrainment caused by aerodynamic loading; the precise nature of energy transfer from the glottal airflow to the vocal fold domain was not reported.

Other studies have similarly used fully coupled models to investigate different aspects of voice production, but they have not explicitly addressed the issue of aerodynamic energy transfer mechanisms. In one of the early implementations of detailed self-oscillating numerical models, Alipour and Titze (1996) used finite-element analysis and the Navier-Stokes equations to predict tissue deformation. Glottal airflow velocity and wall pressure profiles were obtained. Alipour and Scherer (2000a) used the finite element method to study the influence of bulging of the medial surface of the vocal folds by contraction of the thyroarytenoid muscles. The influence of subglottal pressure and the effects of the presence of the false folds were investigated using a three-dimensional finite element vocal tract model by De Oliveira Rosa et al. (2003). In this study, the fluid and solid domains were modeled, although the fluid domain mesh was coarse. The above simulations demonstrate the superiority of the finite element approach over multi-mass models in terms of the ability to predict detailed flow and structural behavior.

Experimental self-oscillating models that have been investigated include excised canine and human larynx models and synthetic models. For example, Alipour and Scherer (1995, 2000b) investigated glottal airflow and pressure distribution in excised larynges. Tissue velocity measurements were not included. Berry et al. (2001) obtained high-speed images of the vocal fold medial surface, but without detailed aerodynamic measurements. Excised larynx experiments have the advantage of physiological accuracy, but have limited self-oscillation viability; furthermore, the material properties and some geometric quantities (such as tissue layer thicknesses) are unknown and vary between larynges.

Self-oscillating membranous synthetic models have also been used. Kakita (1988) used thin transparency film to simulate vocal fold vibration. Measurements of membrane
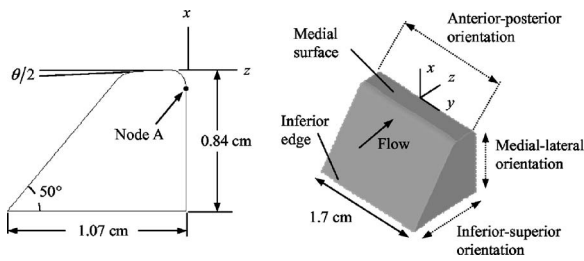
FIG. 2. Synthetic model design dimensions and coordinate system orientation.

FIG. 3. (Color online) Orifice plate geometry. (a) Rigid mold used to cast the synthetic model; (b) and (c) two views of a rubber model and its acrylic support. $w = 1.2$ cm; $h = 2.5$ cm; $l = 5.1$ cm.

shape, jet velocity, and radiated acoustic pressure were obtained. Isogai *et al.* (1988) used a silicone-backed membrane to simulate the vocal fold cover and body and obtained flow and frequency data, as well as intraglottal pressure data at certain points. Neither of the above two studies reported sufficient data for measuring aerodynamic energy transfer. The membranous model of Titze *et al.* (1995) and Chan *et al.* (1997) was used to verify the theory of phonation threshold pressure proposed by Titze (1988) as it relates to prephonatory glottal width and glottal angle. Onset and offset pressure measurements were reported, but no detailed information on the motion of the membrane or of the orifice flow was provided.

In the present study, the hypothesis that a cyclic converging/diverging orifice profile is needed for self-sustained vocal fold oscillations to be maintained was verified using complementary experimental and numerical models. This was accomplished by:

(i)   Designing and fabricating a self-oscillating physical model, with geometry and behavior similar to the human larynx;

(ii) Developing a corresponding two-dimensional finite element model capable of simulating fully coupled fluid-structure interactions, with geometry and material properties matching the physical model and boundary condition definitions based on physical model experimental data;

(iii) Exercising the numerical model to quantify the energy transfer from the flow to the solid.

The use of similar physical and numerical models facilitated development of key numerical model parameters, including material properties and upstream and downstream pressure boundary conditions. The numerical analysis supplied data that could not be measured experimentally, including the surface velocity and the total fluid stress on the vibrating surface. A determination of these two values is necessary to quantify aerodynamic energy transfer.

II. SYNTHETIC MODEL

The physical model was designed to achieve reasonable similarity with the human vocal folds with respect to (1) onset pressure; (2) oscillation frequency; (3) size and shape; and (4) material properties.
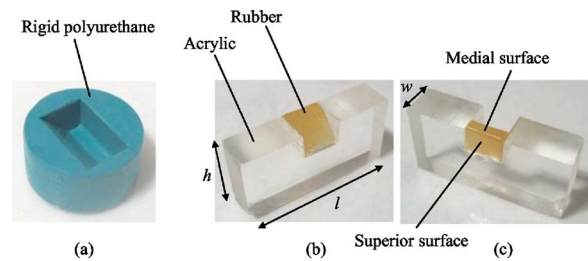
A. Geometry and dimensions

A schematic of the synthetic model is shown in Fig. 2, along with the coordinate system used. The model geometry is the same as that of the rigid model described by Scherer *et al.* (2001) (in which complete dimensions, including specifications for the glottal entrance and exit radii, can be found). This geometry was chosen to facilitate comparisons with data available from static models. The cross section was uniform in the anterior-posterior direction; the anterior-posterior length was approximately 1.7 cm, a value close to the length of the vibrating portion of the vocal folds of an average adult human male (Titze, 2000). A slightly convergent angle of $\theta = 5°$ was used.

B. Fabrication procedures

A rigid "positive" piece was constructed using a stereolithographic (SLA) numerically controlled rapid prototyping tool. The rapid prototyping tool input file was generated using the commercial computer-aided modeling package Pro/Engineer. The rigid model was of the desired size and shape. The rigid SLA prototype was immersed in a two-part polyurethane liquid plastic compound (Smooth-Cast™ 300) and removed after the plastic had cured (10 min). A photograph of the resulting rigid plastic mold is shown in Fig. 3(a).

The rigid mold was sprayed with a mold-release agent (Universal Mold Release™). A three-component liquid polymer solution was poured into the mold and allowed to cure for 24 h. The polymer solution consisted of Evergreen™ 10, a two-part polyurethane addition cure polymer, mixed with Everflex™ for increased flexibility. These products were obtained from a commercial vendor.[1]

Varying the proportion of the three compounds in the solution resulted in cured materials of different stiffness and strength. Volume ratios of Evergreen 10 part A : Evergreen 10 part B : Everflex in the range of 1:1:4 to 1:1:5 typically resulted in a material that, when cast into the shape shown in Fig. 2, was flexible enough for self-sustained flow-induced vibrations to be achieved for reasonable subglottal pressures. Increasing the Everflex content resulted in increased compliance but decreased strength and curability. A ratio of approximately 1:1:7 seemed to be near the practical upper limit for structural integrity and tear resistance.

After curing, the polymer was removed from the mold. The composition of the physical model appeared to be homogeneous and isotropic. Material shrinkage and viscoelastic creep caused small deviations of the final model dimen-

sions from the design dimensions shown in Fig. 2. A thin coating of talcum powder was applied to reduce the surface tackiness.

Tensile and compressive tests were performed using a cured cylindrical sample (2.54 cm in diameter; 1.3 cm in height) that was cast at the same time as the physical model. Stress, σ, and strain, ε, data obtained using an EnduraTEC ELF 3200 test instrument yielded a Young's modulus value of approximately 13.7 kPa in the range $-20\% \leqslant \varepsilon \leqslant 20\%$, with a nearly linear stress versus strain slope in this range. Details of the material testing are described elsewhere (Thomson, 2004).

The structure of the human vocal fold is generally characterized as having three layers of very different Young's moduli: the cover, consisting of the epithelium and the superficial layer of the lamina propria; the transition, which includes the intermediate and deep layers of the lamina propria; and the body, comprised of the vocalis muscle (Hirano and Kakita, 1985). Relatively little data is available regarding the transverse Young's modulus of human vocal fold tissue. In one report, Tran *et al.* (1993) measured the transverse Young's modulus of the human vocal folds and found the mean value to be 12.7 kPa (with no recurrent laryngeal nerve stimulation). The Young's modulus of the synthetic material was therefore representative of the average of the transverse Young's moduli of the different vocal fold tissue layers. The use of the one-layer model simplified the fabrication process and the comparison with numerical simulations.

Fabrication and operation of the synthetic model was relatively straightforward and inexpensive. There was no need to condition the flow to prevent the structure from dehydrating, and there was no need to freeze the model before use, as is necessary with excised larynges. It was possible to reuse the model, even after several months, with reasonable repeatability. The model could be operated for several hours with uninterrupted oscillations. Such long-term operation is not feasible for excised larynges.

C. Experimental apparatus

The anterior-posterior and lateral surfaces of the model (orientation shown in Fig. 2) were attached to a 1.2 cm thick rigid acrylic orifice plate using a liquid polyurethane adhesive (ProBond®, Elmer's Products, Inc.), as represented in Fig. 3(b) (note that shown is a slightly shorter physical model than that used in the experiments discussed here). Two such components were fastened together to form a single orifice plate, representing a full laryngeal configuration. For consistency, two rigid molds were created using the same SLA part, and the two physical models were cast simultaneously using the same liquid polymer solution. The separation distance between the support plates was variable; for the experiments reported here, the medial surfaces of the model were just touching in the no-flow condition (for a zero orifice width).

The orifice plate was installed at the end of a 20 cm long duct, or test section, as shown in Fig. 4. The test section had a square cross section with an area of approximately 5.1 cm². The sidewalls were constructed of 0.635 cm thick aluminum,
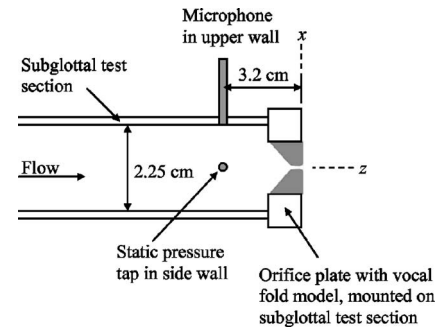
FIG. 4. Test section schematic.

and the upper and lower walls were made of 0.95 cm thick aluminum. A square cross-sectional shape (as opposed to a circular shape) was selected to facilitate instrumentation installation. There was no supraglottal duct.

A 0.635 cm diameter static pressure tap was mounted in one wall, flush with the interior surface. The time-averaged static pressure was measured using a pressure gauge and readout (MKS 100 Torr Baratron pressure gauge; MKS PDR-C-1C power supply and digital readout). A microphone (B&K type 4939) was mounted in the adjacent wall, flush with the interior surface. A B&K type 5935 microphone power supply was used. The microphone and pressure tap were placed at the same streamwise location, $z = -3.2$ cm (referenced to the orifice plate exit plane). The microphone sensitivity was provided by the manufacturer.

The flow supply was compressed shop air regulated to the desired pressure. The flow rate was measured using an MKS 0558A flow meter. A cylindrical plenum, of 27 cm diameter and 36 cm height, was connected to the flow meter via flexible tubing. A tube of approximately 1.27 cm diameter and 1.6 m length connected the plenum to the subglottal test section.

A strobe light (GenRad 1546 Strobotac®) illuminated the model superior surface at a frequency close to that of the model oscillation frequency to capture frontal-view images of the oscillating model using a digital single-frame camera.

The signals of the microphone, the flow meter, and the static pressure gauge were fed to an HP 35650 Data Acquisition System. The sampling frequency was 32 kHz, and the sampling time for each dataset was typically one second.

III. FINITE ELEMENT MODEL

A. Numerical domain

The computational domain, which consisted of distinct but fully coupled fluid and solid domains, is shown in Fig. 5. Both domains were two-dimensional.

The solid domain allowed for large deformation and large strain by using a hyperelastic material model, with input parameters based on the measured stress-strain data (discussed above). A Poisson's ratio of 0.45 was assumed. The shape and size of the structural domain was the same as the physical model, with the exception of the included angle of the medial surface, which was initially zero ($\theta = 0°$). The initial orifice width, d_g, defined as twice the distance between the medial surface of the structural domain and the line of
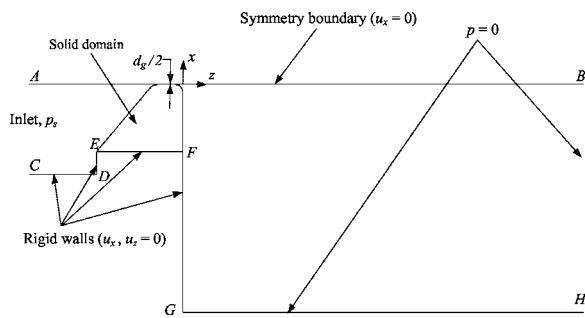
FIG. 5. Outline of the numerical domain. The inlet and outlet boundaries are located farther away from the solid domain than what is shown.

symmetry (see Fig. 5), was 0.05 mm; this was chosen to approximate initial contact of the medial surfaces (a nonzero value of d_g was required for the fluid domain mesh).

The fluid domain was laminar, incompressible, and governed by the Navier-Stokes equations; this allowed for prediction of flow separation and inclusion of unsteady effects. The fluid was air, with density $\rho = 1.2$ kg/m^3 and viscosity $\mu = 1.8 \times 10^{-5}$ kg/(m s).

B. Boundary conditions

The domain was symmetric about line AB; the normal component of velocity, u_x, was set to zero along AB. Zero-pressure boundary conditions were applied along GH and BH to simulate a free jet. Rigid walls with zero-velocity boundary conditions were defined along CD, DE, EF, and FG. The lengths of the inlet (CD) and the outlet (GH) sections were 2.1 and 20 cm, respectively. The inlet location corresponded to the location of the upstream pressure probes (microphone and static pressure tap) in the experiments. The height of the inlet channel (AC) was 1.1 cm; the height of the outlet domain (BH) was 2.8 cm.

The inlet pressure, p_s, was imposed along AC. The magnitude of p_s was prescribed to vary sinusoidally about a mean value of 2 kPa with an amplitude of 1.2 kPa; these values were based on experimental data acquired using the physical model. The physical model oscillation frequency was around 120 Hz, which was near the average of the first two modal frequencies (85 and 140 Hz, respectively, as predicted by a corresponding three-dimensional solid model modal analysis; see Thomson, 2004). However, while the physical model was rigid at the anterior-posterior surfaces, the numerical model was purely two-dimensional. Thus, the modal frequencies of the numerical model (60 and 126 Hz, respectively) were lower than for the physical model. The average of these two frequencies, 93 Hz, was chosen as the inlet pressure driving frequency. It is noted that Alipour *et al.* (2000) reported that the vibration frequency of their finite element vocal fold model was near the average of the first two modal frequencies.

C. Numerical methods

The finite element method was applied to the numerical domain to simulate the motion of the synthetic model. In general, application of the finite element method consists of defining a domain of interest representing the physical sys-
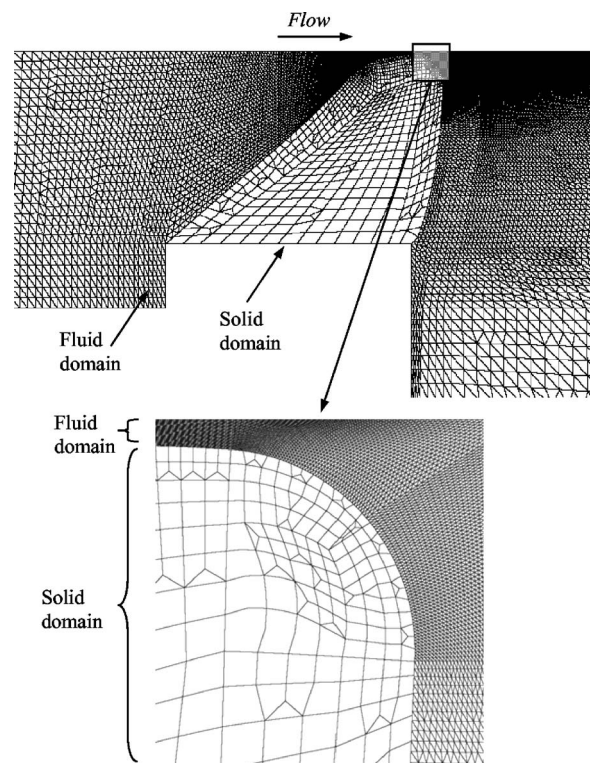
FIG. 6. Images of the numerical solid and fluid domain meshes during oscillation.

tem, defining boundary conditions, identifying the governing equations in variational form, reducing the governing equations to a set of linearized algebraic equations, and solving the resulting system of equations to predict the behavior of the physical system. The continuous variational function is discretized into small regions or "elements" over the domain. The elements are defined using interpolation polynomials that approximate the continuous function over the region occupied by the element.

For the present system, the domain and boundary conditions were defined as discussed above. The commercial code ADINA (ADINA R&D, Inc., Boston, MA), which is designed for analyzing coupled fluid-structure systems, was used to implement the finite element method; details of the numerical methods employed by ADINA can be found elsewhere (Bathe, 1996; ADINA, 2002).

Simulations were performed to ensure that the solution was reasonably independent of time step size and mesh density. The structural domain consisted of 966 9-node (quadratic) plane strain elements and 3974 nodes; the fluid domain consisted of 62 639 3-node (linear) elements and 31 897 nodes. The meshes are shown in Fig. 6 (shown is one instant during oscillation). The time marching algorithm used the backward Euler method with a time step size of 10^{-4} s (for over 100 time steps per cycle at 93 Hz).

To investigate accuracy of the flow solver, the predicted pressure along the solid domain wall was compared with that predicted by Fluent using geometrically identical fluid domains. The Fluent model used here was the same as that used by Scherer *et al.* (2001), in which a good agreement was demonstrated between their Fluent predictions and surface
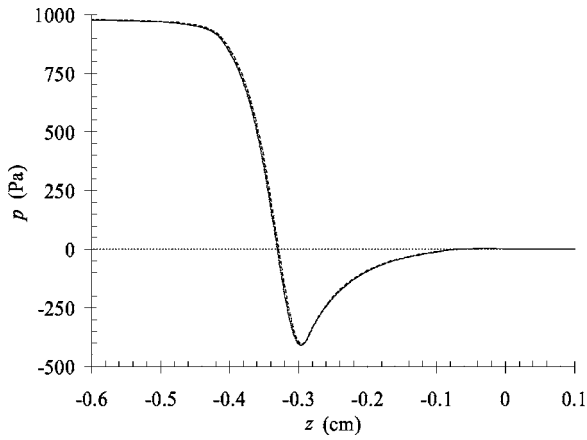
FIG. 7. A comparison of ADINA and FLUENT predicted pressure distributions along a portion of the solid domain wall (without wall motion). —: FLUENT; $\cdots$: ADINA.

pressure measurements obtained using their rigid vocal fold model. The shape of the Fluent model discussed here was the same as that shown in Fig. 2, only with a 5.1° divergent orifice profile. The minimum glottal width was 0.4 mm and the inlet pressure was 0.98 kPa (constant). The downstream section was a channel, rather than a free jet. The finite element model domain inlet and outlet locations were shortened to match the Fluent model, but the mesh densities of the two ADINA finite element models (one with the rigid solid domain and the other with the flexible solid domain) were nearly identical. The results, shown in Fig. 7, show excellent agreement between the pressure distributions of the finite element and Fluent models.

To simulate mechanical energy dissipation in the solid, a Rayleigh proportional damping model was used. This model assumes that the overall damping is determined by the damping of the individual modes. The damping matrix is calculated from the stiffness and mass matrices (Bathe, 1996):

$$\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}, \tag{1}$$

where $\mathbf{C}$, $\mathbf{M}$, and $\mathbf{K}$ are the structural damping, mass, and stiffness matrices, respectively. The Rayleigh constants α and β are related to the damping ratio ζ_i corresponding to the modal frequency ω_i through

$$\alpha + \beta\omega_i^2 = 2\omega_i\zeta_i. \tag{2}$$

The parameters α and β can be calculated from two unknown damping ratios and two modal frequencies. For the two-dimensional solid model, the first and third modal frequencies were $f_1 = 60$ Hz ($\omega_1 = 377$ rad/s) and $f_3 = 142$ Hz ($\omega_3 = 892$ rad/s). The damping ratios of $\zeta_1 = 0.1$ and $\zeta_3 = 0.2$ were assumed, based in part on measured data indicating that the damping ratios of vocal tissues are between 0.1 and 0.4, and also noting that the damping ratios depend only weakly on frequency (Chan and Titze, 1999). The calculated Rayleigh damping constants were $\alpha = 14.4$ and $\beta = 4.3 \times 10^{-4}$.

One consequence of Rayleigh damping is that higher frequencies may be significantly more damped than lower

frequencies. However, it has been shown (Berry *et al.*, 1996) that vocal fold motion is mostly characterized by contributions from the first three modes of vibration.

It is noted that no impact modeling was necessary, since, for the flow rates used here, the oscillations of the numerical model (as well as the physical model) were such that collision between opposing folds did not occur and the orifice area remained open throughout the oscillation cycle.

IV. ENERGY TRANSFER THEORY

In general, the rate of energy transfer (power) is the product of force and velocity. For the case of fluid flow and a moving surface, the rate of work done by fluid stress on the surface is given by (Kundu, 1990; Slattery, 1972)

$$\dot{E}_T = \int_S u_i\tau_{ij}(-n_j)dS, \tag{3}$$

where $\dot{E}_T$ is the rate of energy transfer resulting from the total fluid stress, u is the fluid velocity, S denotes the moving control surface (the vocal fold wall in the present case), and n is the control surface outward normal unit vector. The fluid stress tensor, τ_{ij}, includes normal and viscous forces, and is given by

$$\tau_{ij} = -p\delta_{ij} + 2\mu e_{ij}, \tag{4}$$

where p is the fluid pressure and δ_{ij} is the Kronecker delta, defined as $\delta_{ij} = 1$ for $i = j$ and $\delta_{ij} = 0$ for $i \neq j$.

Equation (3) can be further decomposed into normal and viscous stress components using Eq. (4):

$$\underbrace{\int_S u_i\tau_{ij}(-n_j)dS}_{\dot{E}_T} = \underbrace{\int_S u_i(-p\delta_{ij})(-n_j)dS}_{\dot{E}_p} + \underbrace{\int_S u_i(2\mu e_{ij})(-n_j)dS}_{\dot{E}_\mu} \tag{5}$$

Since the terms in Eqs. (3) and (5) denote rates of energy transfer, or energy "flow," they have units of power (Joules per second), and each integrand represents intensity (energy per unit time and per unit area, or energy flux). The total energy transfer during a certain time period (units of Joules) is calculated by integrating the $\dot{E}$ terms over a specified time interval. The variables $\dot{E}$ and $\dot{I}$ represent energy flow and energy flow intensity, respectively, with the subscripts denoting whether the calculated quantity includes total, normal, or viscous fluid stress terms. Note that since the domain was two-dimensional, all energy flux quantities are per unit depth.

V. RESULTS AND DISCUSSION

A. Synthetic model response

The onset pressure and flow rate values were $p_{on} = 1.15$ kPa and $Q_{on} = 169$ cm^3/s, respectively, and the offset pressure and flow rate values were $p_{off} = 1.00$ kPa and $Q_{off} = 124$ cm^3/s, respectively. These values are consistent with previous observations (e.g., Lucero, 1995; Titze *et al.*, 1995) that the onset pressure is greater than the offset pressure. The
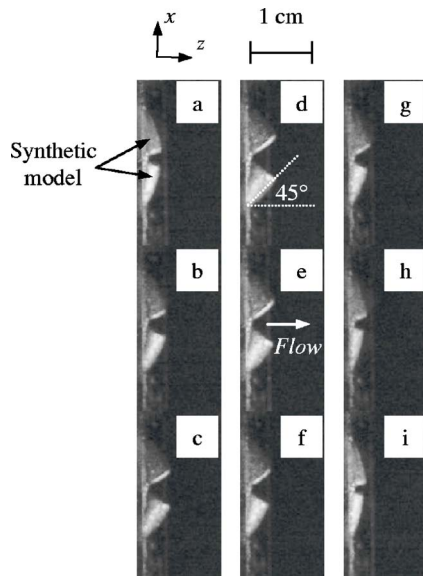
FIG. 8. Images of the synthetic model (frontal view, obtained using the strobe light) during oscillation.

ratio p_{off}/p_{on} for the physical model was 0.87. Lucero (1999) showed analytically that the ratio should be between 0.5 and 1. Using a physical model of the cover, Titze *et al.* (1995) reported a ratio of approximately 0.87 for an initially straight glottis with a prephonatory glottal half-width of 1 mm (as reported by Lucero, 1995), and Chan *et al.* (1997) reported ratios between 0.75 and 0.8.

Jiang and Titze (1993) reported threshold pressures of approximately 0.9 kPa for excised canine larynges, whose values were between 0.2 and 0.3 kPa greater than previously reported values obtained in a human excised larynx, and between 0.3 and 0.5 kPa greater than previously reported values obtained using human subjects. Flanagan (1972) reported that the onset pressure for an adult human male is around 0.4 kPa. Thus, the model onset pressure was comparable to that for excised canine larynges, greater than that for excised human larynges, and comparable to lung pressure for loud speech (Titze, 2000). The onset pressure was sensitive to the upstream loading (the length of the upstream tube, for example). It is noted that a 1.6 m tube has a half-wavelength resonance frequency of 108 Hz, which is close to the physical model oscillation frequency. The influence of the upstream acoustical loading on the vibrations of the physical model has been reported elsewhere (Zhang *et al.*, in review). In their study, the influence of subglottal duct length on physical model vibration characteristics, including onset pressure and vibration frequency, were investigated. It was found that the physical model fundamental frequencies entrained to the subglottal tube resonances. A significant variation in measured onset pressure with subglottal duct length was also reported.

Figure 8 shows a frontal view of the oscillating model at several times during the oscillation cycle, obtained using the digital camera and the strobe light. The motion of the physical model included significant inferior-superior motion as well as medial-lateral motion. The magnitude of the displacement varied with flow rate. At the time of minimum

orifice area [$t/T \approx 0$, where T is the period; see, for example, Fig. 8(i)], the maximum z displacement, z_{max}, was typically about 1.5 mm and occurred at the medial, superior tip of the model. At the time of maximum area ($t/T \approx 0.5$; see Fig. 8(e)], z_{max} was approximately 3.5 mm, for a total inferior-superior displacement of around 2 mm. The orifice width, which also varied with flow rate, ranged from 0 mm at t/T =0 (at flow rates sufficiently high to produce collision) to approximately 3.5 mm at $t/T=0.5$.

As the flow rate was increased prior to the onset of oscillations, the orifice width increased; this likely contributed to the high onset pressure. Recall that the medial surfaces were in contact in the absence of flow. As the flow rate increased, the preoscillatory orifice width also increased as the medial-superior edge moved superiorly and laterally. It has been shown analytically (Titze, 1988) and verified experimentally (Titze *et al.*, 1995) that the onset pressure generally decreases as the prephonatory glottal width decreases (except for very small glottal widths, where the aerodynamic viscous resistance becomes significant; Lucero, 1996). In the human vocal folds, a small glottal width is maintained via muscle control prior to the onset of phonation. By contrast, the preoscillatory orifice width of the physical model was relatively large (typically greater than approximately 1 mm).

Since the synthetic material composition can be varied prior to curing to control stiffness, further experiments may be performed with the model cast in layers, each with a different stiffness, to represent the multilayer structure of the vocal folds. Creating a multilayer model may result in a decreased onset pressure by allowing a smaller preoscillatory orifice width to be maintained. Additionally, material transverse isotropy may be achievable by embedding small elastic fibers, oriented in the anterior-posterior direction, during casting.

B. Numerical model response

For the numerical model, time traces of the subglottal pressure and the x and z positions of node A on the solid domain surface are shown in Fig. 9 (the positions are shown relative to the coordinate system shown in Fig. 2; the location of node A is also shown in Fig. 2). The system reached nearly steady-state oscillation after approximately 50 ms.

The solid model surface position over the cycle from t =0.055 s to t=0.0658 s is shown in Fig. 10. The inferior-superior motion is evident, and the shape of the medial surface can be discerned. The medial surface was convex near the time of minimum opening, and concave or flat over the rest of the cycle. An alternate convergent-divergent profile is seen, with the convergent profile occurring during opening and divergent during closing; this profile-phase relationship is consistent with human vocal fold oscillations (Titze, 1988).

A comparison of the solid domain superior surface motion shows reasonable agreement with the images of the synthetic model. Using the scale and images shown in Fig. 8, the magnitudes of the z and x displacements of the medial surface near the glottis exit over one cycle were both estimated to be approximately 2.4 mm. Near the time of maximum

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Thomson *et al.*: Energy transfer to the vocal folds   1695
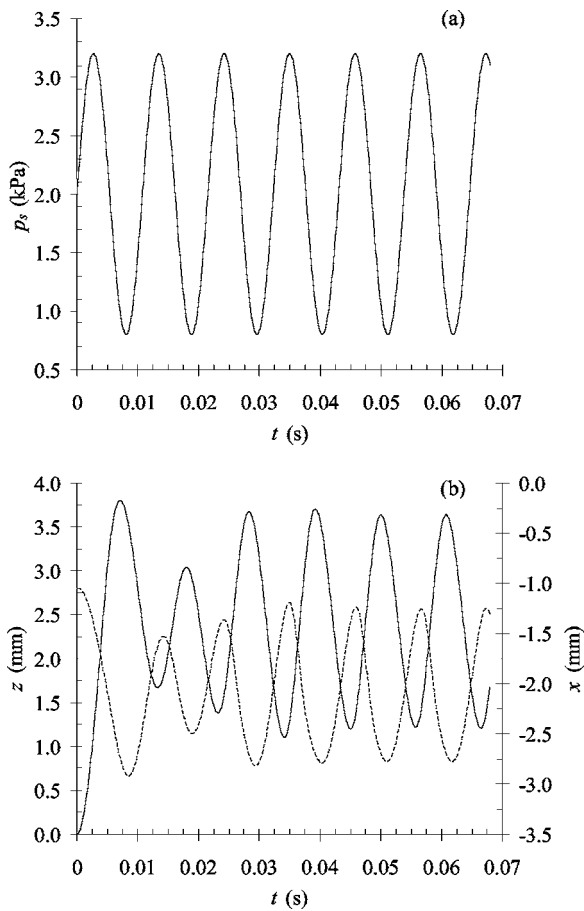
FIG. 9. Numerical model response. (a) p_s versus time; (b) position of node A on the solid surface versus time; —:z position; ---:x position.
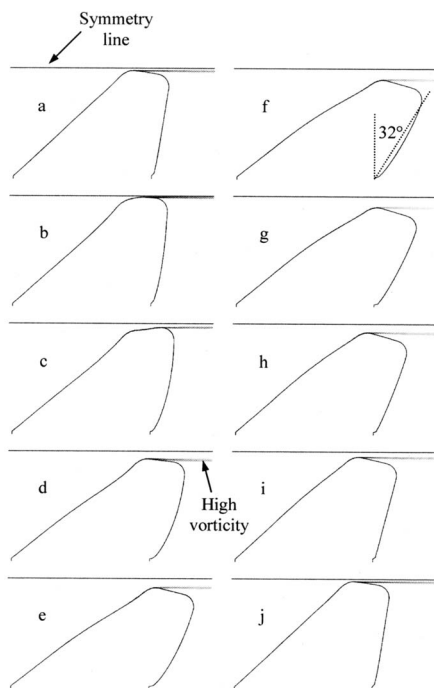
FIG. 10. Images showing the numerical model solid domain during different times of one oscillation, along with flow vorticity, from times $t_1=0.055$ s to $t_2=0.0658$ s, at $\Delta t=0.0012$ s intervals. The gray line denotes regions of high vorticity and illustrates the flow separation point.
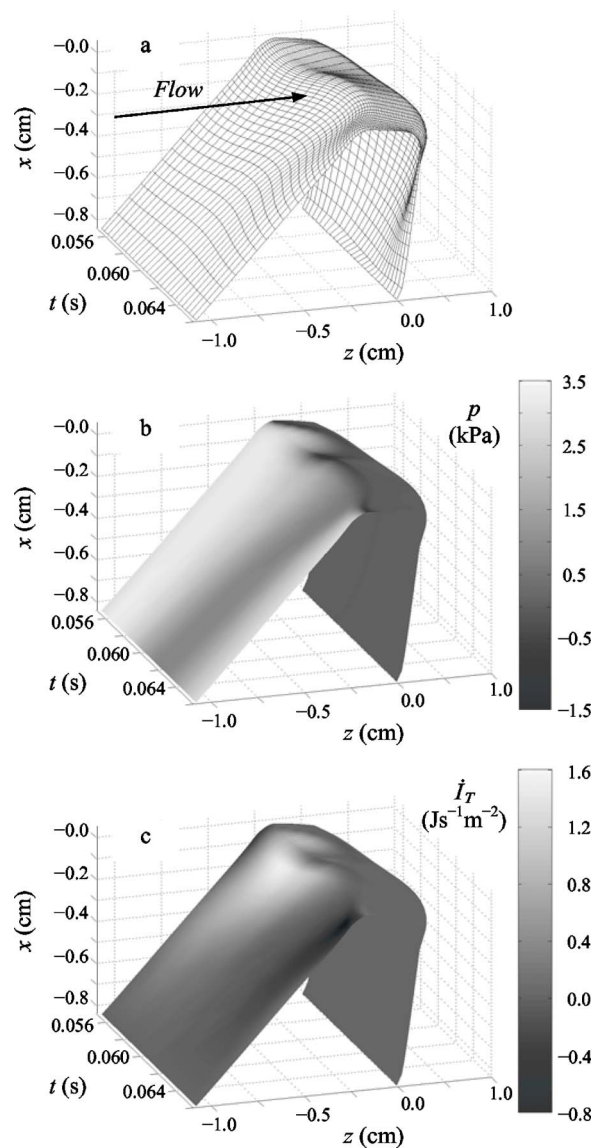
FIG. 11. For the numerical model solid domain over one cycle: (a) Wall position; (b) pressure superimposed on wall position; (c) energy flow intensity superimposed on the wall position.

opening, the synthetic model superior surface had rotated by approximately 45°. This compares to z- and x-displacement magnitudes of approximately 2.4 and 1.5 mm, respectively, for a superior surface rotation of approximately 32°, for the numerical model. A comparison of the medial surface motion was not possible due to limited optical access.

The difference in displacement magnitude was attributed to the choice of inlet pressure driving frequency and damping parameters in the numerical model. The anterior-posterior rigid constraints in the physical model were not modeled in the two-dimensional simulations, and one would expect a larger vibration amplitude with a two-dimensional model. For the numerical model, however, the frequency of the inlet pressure boundary condition was adjusted according to modal analysis predictions (see Sec. III B); it is noted that larger amplitudes did indeed result when the inlet frequency matched the model fundamental frequency. Three-dimensional simulations with detailed flow predictions were not feasible due to excessive computational costs.

1696   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Thomson *et al.*: Energy transfer to the vocal folds
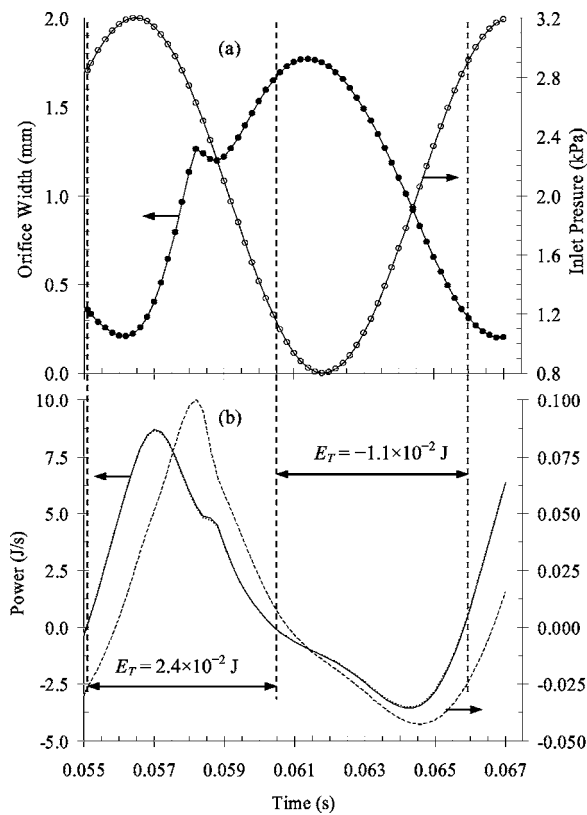
FIG. 12. For one cycle of the vocal fold continuum model: (a) -●-: orifice width; -○-: inlet pressure; and (b) net rate of energy transfer to the physical model surface due to total, pressure, and viscous stresses: Left ordinate: —:$\dot{E}_T$; ⋯:$\dot{E}_p$; right ordinate: ---:$\dot{E}_\mu = \dot{E}_T - \dot{E}_p$.

FIG. 13. Quantities along the fold surface at time $t=0.0551$ s: (a) —:x vs z; →:$u_i n_i$ vs z; (b) p vs z; (c) $\dot{I}_T$ vs z.

main). The local maximum at approximately $t=0.058$ s coincided with the time near the transition from a convergent to a divergent orifice profile [see Figs. 10(c) and 10(d)].

C. Aerodynamic energy transfer

The rate of energy transfer to the numerical model surface due to the total fluid stress, $\dot{E}_T$, over one period is shown in Fig. 12(b). Also shown are the individual contributions of the pressure and viscous stresses ($\dot{E}_p$ and $\dot{E}_\mu$, respectively). The total energy transfer was nearly entirely due to the normal pressure. The maximum energy transfer due to viscous stresses occurred when the orifice profile transitioned from a convergent to a divergent shape ($t=0.0582$ s); however, at this time the energy flow was only about 1% of the maximum total energy flow. Thus the contribution of viscous forces to the overall energy balance was negligible in terms of the total energy transfer. At the minimum orifice width, however, energy flow due to normal forces was zero, and viscous forces were responsible for all of the energy transfer.

The net area under the $\dot{E}_T$ curve over one cycle in Fig. 12(b) is positive; approximately twice as much energy was imparted to the solid domain than what was recovered. The remainder was dissipated in the solid due to damping. The rate of energy transfer was slightly out of phase with the orifice width, but most of the net energy transfer to the solid occurred as the orifice was opening.

Figures 13–15 show the wall position, the normal component of velocity, the wall pressure distribution, and the energy flow intensity, $\dot{I}_T$, at three different times. At time t

Figure 11(a) shows the position of the solid domain surface over one oscillation period. Figures 11(b) and 11(c) show the surface position with gray scale values used to show distributions of pressure and energy flow intensity, respectively. The pressure on the superior surface was nearly zero for the entire cycle, was positive on the inferior surface, and fluctuated between positive and negative values on the medial surface.

The lateral portion of the inferior surface and the entire superior surface both experienced minimal energy transfer with the surroundings. The reason for the minimal energy transfer at the superior surface was due to the small pressure on that surface (the pressure was nearly zero after the flow had separated from the surface, as shown below). The regions of greatest energy transfer were along the medial portion of the inferior surface and along the entire medial surface during the opening phase of the cycle (this is also apparent in Figs. 13–15, as discussed later). The energy transfer near the time of maximum opening was nearly zero over the whole fluid domain.

Figure 12(a) shows the orifice width over one cycle. Also shown in Fig. 12(a) is the "subglottal" pressure, p_s, that was applied at the fluid domain inlet. The orifice width varied from approximately 0.2 mm at the time of minimum opening to 2.8 mm at the maximum opening. Note that the orifice width refers to the full orifice width (twice the distance between the point on the medial surface of the solid domain closest to the plane of symmetry in the fluid do-

FIG. 14. Quantities along the fold surface at time $t=0.057$ s: (a) —:x vs z; →:$u_i n_i$ vs z; (b) p vs z; (c) $\dot{I}_T$ vs z.

FIG. 15. Quantities along the fold surface at time $t=0.0642$ s: (a) —:x vs z; →:$u_i n_i$ vs z; (b) p vs z; (c) $\dot{I}_T$ vs z.

$=0.0551$ s (Fig. 13), the orifice was nearly closed and the profile was slightly divergent and transitioning to a convergent shape. The inlet pressure was approaching its maximum specified value. The net energy flow to the solid was zero since the regions of the positive and negative energy flow cancelled. At the time of maximum energy flow to the solid ($t=0.0570$ s; Fig. 14), the pressure was in phase with the surface velocity over most of the surface; thus the energy flow was distributed along nearly the entire length of the inferior and medial surfaces. The orifice profile was convergent and the medial surface was slightly concave. At the time of maximum energy flow from the solid to the flow ($t=0.0642$ s; Fig. 15), the orifice profile was divergent and the orifice was closing. The energy flow was negative since the velocity and pressure terms were out of phase. The region of greatest energy transfer was near the medial, inferior tip [compare Figs. 15(a) and 15(c)].

Figure 16 shows the spatially averaged pressure, p_{avg}, with p_{avg} defined as

$$p_{avg} = \frac{1}{L} \int p \, dl, \tag{11}$$

and where L is the wetted length of the vocal fold surface. It can be seen from Fig. 16 that the average pressure was never negative. Even though localized regions did experience negative pressure, as seen in Figs. 13–15, these regions of negative pressure did not significantly contribute to the energy transfer. Therefore, there was never a resultant negative Bernoulli-type force "pulling" the solid model surface medially. However, the time-averaged pressure, p_{mean}, during

positive energy flow to the solid (orifice opening, p_{mean} $=1.3$ kPa) was greater than that during negative energy flow (orifice closing; $p_{mean}=0.72$ kPa). This is in agreement with the observation by Titze (1988) that for self-sustained oscillations to be achieved, the net intraglottal pressure during closing must be either negative or "less positive" than the net intraglottal pressure during opening.

Note that "orifice opening" has been defined here as the time during which the energy flow to the solid was positive ($0.0551 \leqslant t \leqslant 0.0604$), and "orifice closing" as the time during which the energy flow was negative ($0.0604 \leqslant t \leqslant 0.0658$). The energy flow was slightly out of phase with the orifice width [see Figs. 12(a) and 12(b)]. If the orifice opening and closing phases are instead defined using the orifice width, then the mean pressure terms become p_{mean} $=1.1$ kPa during opening ($0.0562 \leqslant t \leqslant 0.0614$) and p_{mean} $=0.92$ kPa during closing ($0.0614 \leqslant t \leqslant 0.0668$). The temporal pressure asymmetry exists for either definition of orifice opening and closing.

In further agreement with Titze's theory, a comparison between Figs. 14 and 15 shows that the energy flow to the solid is indeed assisted by the medial surface convergent-divergent motion. When the profile was convergent, the flow was attached and the magnitude of the pressure was significant over the entire medial surface of the solid [Fig. 14(b)]. The energy flow was thus also significant over the medial surface of the solid [Fig. 14(c)]. When the profile was divergent, the flow separated from the medial surface, causing the pressure to be nearly zero along a significant portion of the medial surface [Fig. 15(b)]. This resulted in the energy flow
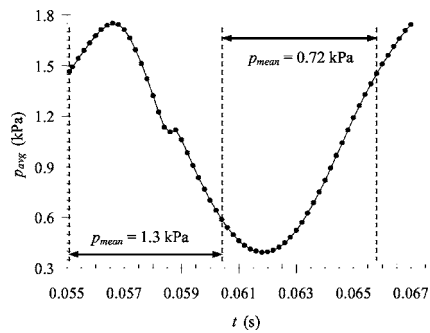
FIG. 16. Pressure, spatially averaged along the numerical model solid domain surface, over one cycle of oscillation.

being confined to a smaller region along the vocal fold surface [Fig. 15(c)]. This temporal asymmetry in pressure distribution causes a similar temporal asymmetry in the energy flow, thus aiding oscillation. Note that the spatial regions of high-energy transfer were strongly dependent on the location of flow separation. This emphasizes the importance of accurate prediction of the flow separation point in numerical models.

VI. CONCLUSIONS

A self-oscillating physical model of the vocal folds was constructed that demonstrated regular oscillations over an extended period of time. The model demonstrated similarities with human vocal folds, including oscillation amplitude, frequency, and subglottal pressure. Numerical simulations of a complementary two-dimensional finite element model were used to calculate the energy transferred from the fluid domain to the solid domain during one period of regular oscillations. The aerodynamic viscous effects, which only minimally contributed to the overall energy transfer, were quantified. It was shown that the alternating convergent-divergent orifice profile resulted in a greater net pressure within the orifice during the time of positive energy flow to the solid than during the time of negative energy flow. This temporal asymmetry in net orifice pressure caused a temporal asymmetry in the energy flow to the solid domain, which allowed for the self-sustained oscillations to be maintained. The results support the hypothesis that a cyclic variation of the orifice profile from a convergent to a divergent shape leads to a temporal asymmetry in the average wall pressure, which is the key factor for self-sustained oscillations of the vocal folds to be achieved.

ACKNOWLEDGMENTS

[1]Smooth-On, Inc., 2000 Saint John Street, Easton, PA, 18042. Phone: (800) 762-0744. Website: www.smoothon.com.

ADINA (**2002**). *ADINA Theory and Modeling Guide Volume I: ADINA*, ADINA R&D, Inc.

Alipour, F., and Scherer, R. C. (**1995**). "Pulsatile airflow during phonation: An excised larynx model," J. Acoust. Soc. Am. **97**, 1241–1248.

Alipour, F., and Scherer, R. C. (**2000a**). "Vocal fold bulging effects on phonation using a biophysical computer model," J. Voice **14**, 470–483.

Alipour, F., and Scherer, R. C. (**2000b**). "Dynamic glottal pressures in an excised hemilarynx model," J. Voice **14**, 443–454.

Alipour, F., and Titze, I. R. (**1996**). "Combined simulation of two-dimensional airflow and vocal fold vibration," in *Vocal Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular Publishing Group, Inc., San Diego, CA, 1996), pp. 17–29.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal-fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Alipour, F., Scherer, R. C., and Knowles, J. (**1996**). "Velocity distributions in glottal models," J. Voice **10**, 50–58.

Bathe, K. J. (**1996**). *Finite Element Procedures* (Prentice–Hall, Englewood Cliffs, NJ).

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Berry, D. A., Montequin, D. W., and Tayama, N. (**2001**). "High-speed digital imaging of the medial surface of the vocal folds," J. Acoust. Soc. Am. **110**, 2539–2547.

Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.

Chan, R. W., and Titze, I. R. (**1999**). "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021.

De Oliveira Rosa, M., Pereira, J. C., Grellet, M., and Alwan, A. (**2003**). "A contribution to simulating a three-dimensional larynx model using the finite element method," J. Acoust. Soc. Am. **114**, 2893–2905.

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception*, 2nd ed. (Springer-Verlag, New York).

Hirano, M., and Kakita, Y. (**1985**). "Cover-body theory of vocal fold vibration," in *Speech Science: Recent Advances*, edited by R. G. Daniloff (College-Hill Press, San Diego), pp. 1–46.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**, 1233–1268.

Ishizaka, K., and Matsuidara, M. (**1968**). "What makes the vocal cords vibrate," *Proceedings of the International Congress of Acoustics*, B1-3, Tokyo, Japan.

Isogai, Y., Horiguchi, S., Honda, K., Aoki, Y., Hirose, H., and Saito, S. (**1988**). "A dynamic simulation model of vocal fold vibration," in *Vocal Physiology: Voice Production, Mechanisms, and Functions*, edited by O. Fujimura (Raven Press, Ltd., New York, 1988), pp. 191–206.

Jiang, J. J. (**1991**). "A methodological study of hemilaryngeal phonation and the measurement of vocal fold intraglottal pressure and impact stress," Ph.D. dissertation, University of Iowa, Iowa City, Iowa.

Jiang, J. J., and Titze, I. R. (**1993**). "A methodological study of hemilaryngeal phonation," Laryngoscope **103**, 872–882.

Jiang, J. J., Diaz, C. E., and Hanson, D. G. (**1998**). "Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: Implications for the pathogenesis of vocal nodules," Ann. Otol. Rhinol. Laryngol. **107**, 603–610.

Kakita, Y. (**1988**). "Simultaneous observation of the vibratory pattern, sound pressure, and airflow signals using a physical model of the vocal folds," in *Vocal Physiology: Voice Production, Mechanisms, and Functions*, edited by O. Fujimura (Raven Press, Ltd., New York, 1988), pp. 207–218.

Kundu, P. K. (**1990**). *Fluid Mechanics* (Academic Press, San Diego, CA).

Lucero, J. C. (**1995**). "The minimum lung pressure to sustain vocal fold oscillation," J. Acoust. Soc. Am. **98**, 779–784.

Lucero, J. C. (**1996**). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," J. Acoust. Soc. Am. **100**, 2551–2554.

Lucero, J. C. (**1999**). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Thomson *et al.*: Energy transfer to the vocal folds   1699

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Slattery, J. C. (**1972**). *Momentum, Energy, and Mass Transfer in Continua* (McGraw-Hill, New York).

Story, B. H., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Thomson, S. L. (**2004**). "Fluid-structure interactions within the human larynx," Ph.D. dissertation, Purdue University, West Lafayette, Indiana.

Titze, I. R. (**1973**). "The human vocal cords: A mathematical model. Part I," Phonetica **28**, 129–170.

Titze, I. R. (**1974**). "The human vocal cords: A mathematical model. Part II," Phonetica **29**, 1–21.

Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536-1552.

Titze, I. R. (**2000**). *Principles of Voice Production*, 2nd printing (National Center for Voice and Speech, Iowa City, Iowa).

Titze, I. R. (**2002**). "Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model," J. Acoust. Soc. Am. **111**, 367–376.

Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.

Tran, Q. T., Berke, G. S., Gerratt, B. R., and Kreiman, J. (**1993**). "Measurement of Young's modulus in the in vivo human vocal folds," Ann. Otol. Rhinol. Laryngol. **102**, 584–591.

Zhang, C., Zhao, W., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aerocoustics of phonation, part II: Effects of subglottal pressure, glottal oscillation frequency, and ventricular folds," J. Acoust. Soc. Am. **112**, 2147–2154.

Zhang, Z., Neubauer, J., Berry, D. A., and Berke, G. S. (submitted). "Influence of subglottal acoustics on vocal fold vibration," J. Acoust. Soc. Am.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). "Computational aeroacoustics of phonation, part I: Numerical methods, acoustic analogy validation, and effects of glottal geometry," J. Acoust. Soc. Am. **112**, 2134–2146.

Effects of later-occurring nonlinguistic sounds on speech categorization

Travis Wade[a] and Lori L. Holt
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213

Nonspeech stimuli influence phonetic categorization, but effects observed so far have been limited to precursors' influence on perception of following speech. However, both preceding and following speech affect phonetic categorization. This asymmetry raises questions about whether general auditory processes play a role in context-dependent speech perception. This study tested whether the asymmetry stems from methodological issues or genuine mechanistic limitations. To determine whether and how backward effects of nonspeech context on speech may occur, one experiment examined perception of CVC words with [ga]-[da] series onsets followed by one of two possible embedded tones and one of two possible final consonants. When the tone was separated from the target onset by 100 ms, contrastive effects of tone frequency similar to those of previous studies were observed; however, when the tone was moved closer to the target segment assimilative effects were observed. In another experiment, contrastive effects of a following tone were observed in both CVC words and CV nonwords, although the size of the effects depended on syllable structure. Results are discussed with respect to contrastive mechanisms not speech-specific but operating at a relatively high level, taking into account spectrotemporal patterns occurring over extended periods before and after target events. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1984839]

I. INTRODUCTION

Human perception of speech sounds is heavily influenced by the acoustic properties of the contexts in which the sounds are heard. This context-dependence is of importance with respect to speech communication; due to coarticulatory influences from neighboring sounds, productions of a single phoneme may vary considerably across different phonological environments and perception must often take this variability into account in order to be effective. An oft-cited example of the reciprocal effects of context dependence in perception and production involves place of articulation in voiced stop consonants. The English phone [d] is distinguishable from the sound [g] primarily by a higher third formant (F3) just after the stop release. Due to coarticulation, however, this resonant frequency is a function not only of the intended consonant category but also of the speech sounds that neighbor it in a word or utterance. A [d] preceded by the liquid [r], for instance, tends to have a much lower, more "[g]-like" F3 than one preceded by [l]; conversely, a [g] preceded by [l] is more [d]-like (with a higher F3 frequency) than one following an [r]. Fortunately, however, these differences do not result in perceptual confusion, because listeners perceive the consonants relative to their environments. Mann (1980) observed that, given the same [d]-[g] series, listeners accept sounds with lower F3 values as instances of [d] when the sounds follow [r] as compared to [l], effectively compensating for the context-dependent differences in production of the consonants. Interestingly, listeners who are unlikely to rely on knowledge of English sound sequences also display

this perceptual pattern. Japanese speakers, who do not perceptually distinguish between English [l] and [r]—and for whom liquid-stop sequences of any kind are phonotactically impossible—also identify the consonants in a context-dependent manner parallel to that shown by English speakers (Mann, 1986). Even 4-month-old infants discriminate [d]-like consonants from [g]-like ones at different F3 values depending on preceding liquids (Fowler *et al.*, 1990). Thus, listeners with limited or no phonetic knowledge nevertheless exhibit phonetic context effects. Effects of this nature are not limited to consonant identification, but seem to be ubiquitous in human speech perception. Across various places and manners of articulation [see Repp (1982) for a review], the pattern is the same: perception of speech sounds shows context influences which mirror—and thus tend to compensate for—the coarticulatory effects of neighboring sounds, allowing for accurate identification.

However, although context-dependent perception is of obvious linguistic benefit as one listens to continuous, naturally produced speech, it does not only occur under these circumstances. Listeners' perception of the same speech sounds is also influenced by nonspeech precursors such as pure tones. Lotto and Kluender (1998) observed that much like preceding [l] and [r] segments, frequency-modulated sine-wave glides approximating the [l] and [r] F3 trajectories and even steady-state F3 tones situated at the [l] and [r] F3 offset frequencies affect categorization of following [g]–[d] consonants. Importantly, the observed sine-wave-induced influence on speech categorization is contrastive in direction: lower-frequency ([r]-like) precursors cause more consonants to be identified as the relatively higher-frequency consonant alternative, [d], compared to higher-frequency ([l]-like) pre-

cursors. The directionality of the influence of nonspeech contexts is thus consistent with the effects of the speech contexts they model; coarticulation in speech is generally assimilative in nature and perceptual compensation is therefore generally in the opposite, contrastive direction. The observation that nonspeech stimuli that model very limited spectral characteristics of speech contexts produce similar effects of context, in conjunction with the finding that nonhuman animals also exhibit context-dependent identification of speech (Lotto *et al.*, 1997) has been taken to suggest that human context-dependent perception might rely on a general contrastive auditory mechanism rather than processes specialized for speech.

Considering Lotto and Kluender's (1998) finding that speech categorization was influenced by the spectral properties of an immediately preceding nonspeech sound, such a contrastive mechanism might be attributed to some relatively peripheral (cochlear or perhaps auditory nerve) process or processes. Early accounts of contrastive context-dependence proposed low-level processes such as peripheral neural adaptation as candidate mechanisms (Holt, 1999; Holt *et al.*, 2000) and forward masking has also been suggested as a possible cause for the nonspeech effects (Fowler *et al.*, 2000). However, it is unlikely that low-level sensory processes such as these are responsible for many of the context-dependent properties of speech perception. Phonetic context effects are observed even when context speech precedes a target speech sound by hundreds of milliseconds or is presented to the ear contralateral to that of speech target presentation (Holt and Lotto, 2002; Mann, 1980). Moreover, speech context may also influence phonetic categorization when it *follows* rather than precedes a speech target (Mann and Repp, 1980; Mann and Soli, 1991; Miller and Liberman, 1979). Mann and Repp (1980), for example, observed that although the frequency region of frication noise is a primary cue in distinguishing [s] from [š], fricative spectra also vary substantially depending on a following vowel. Fricatives followed by [u] have more energy at lower frequencies, and are therefore more [š]-like, than fricatives preceding [a]. However, as in the case of liquid-stop combinations, listeners tend to compensate for this effect perceptually, by accepting frication in lower frequency regions as signifying [s] in the presence of a following [u] compared to [a]. Since long-term, intra-aural, and backward-operating effects are unlikely to derive from masking or peripheral adaptation, it seems that context dependence in speech perception must be a product of higher-level processes involving more central mechanisms operating over longer time windows.

A majority of existing accounts assume that these high-level processes involve either some type of knowledge specific to speech (Gaskell, 2003; Gow, 2003; Nearey, 1997; Smits, 2001) or the perceptual recovery of articulatory events (Fowler, 1986; Fowler *et al.*, 2000; Liberman and Mattingly, 1985). These assumptions are called into question, however, by results suggesting that the influence of nonspeech precursors on speech categorization involves higher-level (nonperipheral) processing as well. Lotto *et al.* (2003) found that sine wave tones like those used by Lotto and Kluender (1998) influenced perception of following consonants even

when nonspeech context and speech target were separated by up to 175 ms of silence or presented to opposite ears; these findings effectively rule out purely peripheral sensory mechanisms of masking and adaptation. Holt (2005) further observed that precursor "melodies" composed of multiple sine-wave tones similarly affected [g]-[d] categorization, even though the spectral characteristics of these nonspeech contexts were defined distributionally by sequential acoustic events that unfolded over seconds. When listeners heard a [g]-[d] series preceded by "high" and "low" sequences comprised of tones overlapping in absolute frequency but with means based on the [r] and [l] F3 values used by Lotto and Kluender (1998), "low" precursors robustly caused listeners to identify more members of a consonant series as [d], parallel to previous speech and nonspeech findings. Moreover, this nonspeech context effect persisted when as much as 1300 ms of silence or up to 13 intervening neutral (midfrequency) acoustic events separated the nonspeech precursors and the speech targets. Effects such as these are most consistent with higher-level (i.e., central, perhaps cortical) auditory processing and cannot be accounted for by solely peripheral mechanisms. Nor, however, can they be attributed to speech- or gesture-specific knowledge or abilities, since the context sounds involved were unambiguously synthetic. Thus, an emerging alternative explanation posits a *higher-level, nonlinguistic* mechanism whereby auditory events (either speech or nonspeech) occurring over time are incorporated into a general, contrast-providing context, relative to which targeted speech and other sounds are perceived (e.g., Diehl *et al.*, 2004; Holt, 2005; Wade and Holt, in press).

An important remaining asymmetry between speech and nonspeech context effects on speech categorization involves the directionality of the effects observed thus far. Whereas both preceding and following speech segments are known to affect phonetic categorization, documented effects of nonspeech contexts have been limited to the influence of precursors on following speech targets. The lack of data on backward-operating nonspeech contrastive effects leaves open the possibility that general auditory contrast may only operate in the forward direction. Conversely, accounts of perceptual compensation for coarticulation that invoke recovery of articulatory gestures (Fowler, 1986; Fowler *et al.*, 2000) or of temporally distributed acoustic features (Gow, 2004; Gow, 2003) readily account for patterns observed in the perception of both progressively and regressively coarticulated speech sounds. As a result, proponents of these accounts have described the general contrastive approach as unparsimonious in providing a mechanism for only a subset of the context effects observed in speech perception. Certainly, it is not necessarily the case that context dependence in the forward and backward directions arises from the same mechanism, even in speech perception. However, if the same central, high-level contrastive process reflected in the robust dichotic, temporally distributed influence of nonspeech precursors on speech observed in recent reports indeed contributes to compensation for carryover coarticulation in speech perception, it seems a reasonable hypothesis that the mechanism may also operate in the reverse direction, with later-occurring nonspeech contexts influencing perception of ear-

lier speech as well. The present study was designed to address this issue, testing whether spectral information from following nonspeech tones may effect the phonetic categorization of a preceding speech series.

In designing such a test, a major concern was that observation of backward effects of nonspeech context on speech perception may be constrained by methodological issues to a greater extent than observation of forward effects, perhaps accounting for the lack of data on the issue thus far. Specifically, we predicted that two methodological obstacles might relate to the way sequences of sounds are segregated into units and streams as a function of their temporal proximity (e.g., Bregman, 1990). First, insufficient temporal proximity between target speech and following nonspeech context might discourage listeners from considering the context in arriving at a speech categorization response. Since the contrastive influence of context-providing stimuli seems to be contingent on their perceptual continuity with target events (e.g., McCollough, 1965; Walker and Irion, 1979), it may be essential that listeners group the nonspeech events into a single stream with targeted speech sounds. This is especially critical when the context follows, rather than precedes, the target speech sound since the context can exert no effect if listeners make phonetic decisions before it is presented. It is known, for example, that following speech information has an attenuated contextual influence on the perception of earlier target speech sounds if listeners are not encouraged to take this information into account. Miller and Dexter (1988) found that although the total length of a syllable affects the classification of its onset consonant as voiced or voiceless, this influence is diminished if listeners respond very quickly after hearing the onset, suggesting that phonetic categorization decisions may be made without considering potentially informative following information. Under less speeded response conditions, Newman and Sawusch (1996) examined the extent to which temporally separated later-occurring speech context influenced the perception of word-initial consonants and found effects only when the context was within a short temporal window (a phoneme or two) after the target. Thus, we predicted that it would be necessary to place the nonspeech context events as closely as possible following target speech sounds to observe their effects.

A potentially competing issue, however, involves perceptual grouping at a lower hierarchical level. If the nonspeech context is presented too close to the target segment, it may be assimilated not only with the appropriate perceptual stream but with the target segment itself. That is, the spectral information provided by the nonspeech tone may be perceived as information *for* the speech target. Bregman (1990) has described processes of this type as fusion within a *unit* (as compared to a *stream*), and it may be similar or related to what occurs in duplex perception (e.g., Ciocca and Bregman, 1989; Liberman *et al.*, 1981) when a nonspeech chirp presented to the ear opposite an acoustically incomplete speech segment contributes to the perception of that segment. Should perceptual grouping along these lines occur, observable context effects of the nonspeech tones on speech categorization would be assimilative rather than contrastive in di-

rection and therefore could override or obscure any effects of perceptual contrast that might occur.

In light of these methodological concerns, the present study attempts to strike a balance between providing for acoustic continuity such that nonspeech context is grouped into a context-providing stream with an earlier-occurring speech target while ensuring that the two are not perceived as information for a single event. Two experiments employed stimuli in which a brief nonspeech tone was inserted just after a target speech segment within a larger word context. The general reasoning behind this design was that a tone that was both preceded and followed by speech was more likely to be incorporated into a context-providing stream than a tone or sequence simply appended following a speech target. In addition, it was hypothesized that a word identification task requiring categorization of both initial and final consonants might further encourage listeners to take the nonspeech context into account; if acoustic information both preceding and following the nonspeech context were required for word recognition, listeners would be unable to respond (and therefore perhaps less likely to make a phonetic judgment on the target) before the nonspeech context sound occurred.

Experiment 1 tested for contrastive effects of following tones on target onset consonants in CVC words and also examined whether the opposite, assimilative, effects might occur if acoustic continuity (signaled by temporal proximity) between speech target and nonspeech context was sufficient. Experiment 2 tested the limits of this experimental design, examining whether any observed patterns would hold up in the absence of the lexical task or the CVC stimulus structure.

II. EXPERIMENT 1

Experiment 1 was designed as a first step in observing whether and how later-occurring nonspeech acoustic events may influence speech perception. Nonspeech context tones were embedded following the initial consonant in a CVC word stimulus. In keeping with previous studies (Holt, 2005; Lotto *et al.*, 2003; Lotto and Kluender, 1998; Mann, 1980), the initial (target) consonant was a stimulus drawn from a series of speech stimuli varying perceptually from [ga]-[da] and nonspeech contexts were pure tones of one of two frequencies shown in previous research to produce contrastive context effects on categorization of these [ga]-[da] series stimuli (Lotto and Kluender, 1998; Holt, 2005). In an effort to understand the situations in which perceptual grouping might lead to competing contrastive and assimilative effects on preceding speech, two conditions were tested, differing only in the temporal proximity of the embedded tone to the target consonant. It was hypothesized that nonspeech information occurring very close to the target consonant would be more likely to contribute to the perception of the consonant's spectral properties (e.g., Bregman, 1990; Wertheimer, 1923), whereas events occurring somewhat later would more likely contribute to a contrast-providing context. In Experiment 1a, nonspeech tones immediately followed the initial formant transitions, whereas in Experiment 1b they did not occur until well into the following vowel. To the extent that contrast effects occurred in either case, it was predicted that

listeners would label consonants [d] (the alternative with the higher F3 frequency) more often before a lower frequency tone than before a higher frequency tone. Assimilation, on the other hand, would result in more [d] responses in the high tone than in the low tone condition.

A. Method

1. Participants

Twenty-three college-age native English speakers from the Carnegie Mellon University community with no known speaking or hearing disorders participated in the study. Eleven participants were arbitrarily assigned to Experiment 1a and the remaining twelve participated in Experiment 1b. Participants were paid at least $7 per hour for their time.

2. Stimuli

Stimuli were English CVC words with pure tones replacing part of the vowel nucleus. Initial CV portions were drawn from a series varying perceptually from [ga] to [da] and the final consonant was an unambiguous [t] or [k], comprising the four-word set [dot, got, dock, gawk]. (For the experimenter, a speaker of North Midland American English, and most subjects interviewed, the vowel in *gawk* was identical to the [a] in the other three words. This vowel did not seem to present any difficulty identifying any of the words. Moreover, the acoustic characteristics of the vowel were identical across tokens, encouraging responses based only on the initial and final consonants.)

The consonant-vowel [ga]-[da] series was identical to that used by Holt (2005). The stimuli were derived from natural [da] and [ga] recordings from a monolingual male native English speaker (Computer Speech Laboratory; Kay Elemetrics, Lincoln Park, NJ; 20 kHz sample rate, 16 bit resolution), in the following manner. From a number of natural productions, one [ga] and one [da] token were selected that were nearly identical in spectral and temporal properties except for the onset frequencies of F2 and F3. LPC analysis was performed on each of the tokens, and a nine-step continuum of filters was created (Analysis-Synthesis Laboratory, Kay Elemetrics) such that the onset frequencies of F2 and F3 varied approximately linearly between [d] and [g] end points. These filters were excited by the LPC residual of the original [ga] production to create an acoustic series spanning the natural [da] and [ga] end points in approximately equal steps. The series was judged by the experimenters to comprise a gradual shift between natural-sounding [da] and [ga] tokens, and this impression was confirmed by regular shifts in phonetic categorization across the continuum by participants in the Holt (2005) study.

For the present experiment, these CV segments were trimmed to 275 ms in length from the initial burst and their amplitudes were attenuated linearly to zero over the final 50 ms. They were followed by a 25 ms silent interval, representing a stop closure, and 170 ms of burst from a [t] or [k] taken from recordings of natural native-English male productions of the consonants in monosyllabic word-final position following a low back vowel. The resulting stimuli were judged
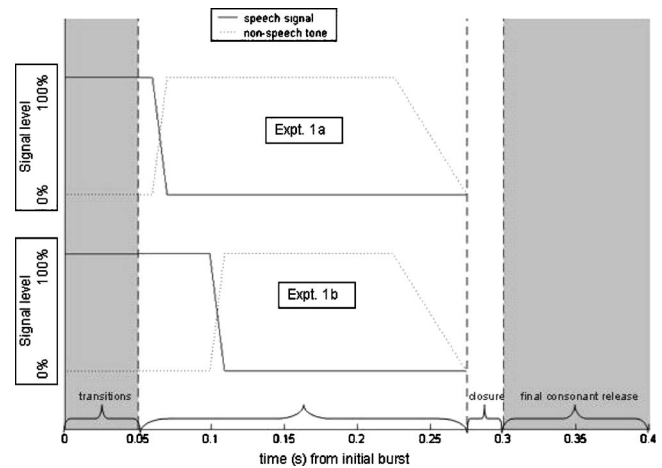
FIG. 1. Schematic illustration of speech and nonspeech stimuli in Experiment 1 conditions.

to resemble careful natural productions of the words *dot, dock, got*, and *gawk*.

Figure 1 shows the composition of Experiment 1 stimuli. In two conditions, nonspeech sine-wave tones were inserted to replace the final portion of the vowel. In Experiment 1a, tones were introduced immediately after the initial formant transitions (transitions lasted approximately 50 ms following the initial burst), replacing the final 215 ms of the vowel. In Experiment 1b tones did not occur until further into the vowel, replacing only the final 175 ms. Tone insertion was achieved as follows: beginning either 60 ms (Experiment 1a) or 100 ms (Experiment 1b) after an initial burst, a tone was gradually introduced over the duration of the vowel, so that the vowel:tone amplitude ratio decreased linearly to zero over 10 ms. The vowel was completely replaced by the tone over the remaining 205 ms (Experiment 1a) or 165 ms (Experiment 1b), with a linear amplitude off-ramp over the final 50 ms. The resulting stimuli gave the impression of spoken word recordings to which synthetic beeps had been added. Speech onset, nucleus, and offset portions were identifiable in all stimuli, although the vowel was necessarily more ambiguous in Experiment 1a stimuli.

Tone frequencies were selected to provide spectral contrast for the initial consonant F3 frequencies. Following Lotto and Kluender (1998), the sine-wave tones had a periodicity of either 1800 or 2800 Hz.

3. Procedure

Participants heard each [ga]-[da] stimulus series member with each of the embedded tone conditions [2800 Hz tone, 1800 Hz tone] and each final consonant 18 times; 18 additional repetitions of each CVC combination without embedded tones were included as filler items, for a total of 972 stimuli. Presentation was controlled by Tucker Davis Technologies (TDT) System II hardware; stimuli were converted from digital to analog, low-pass filtered at 4.8 Hz, amplified, and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL to participants in sound-attenuated booths. Stimuli were presented in random order in
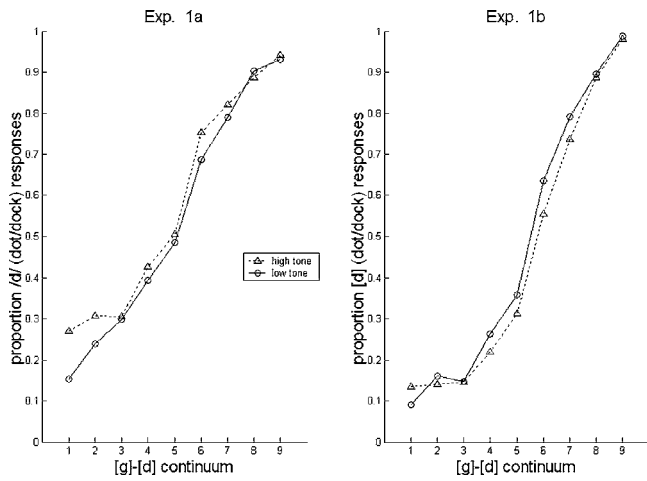
FIG. 2. Initial consonant classification in word response patterns for Experiment 1 conditions, averaged over offset consonants and presented as the proportion of [d] responses across CV stimuli.

two identical sessions separated by a short break; participants were instructed to listen to each word and press one of four buttons labeled *got, dot, gawk, dock* in response.

B. Results

Figure 2 shows listener responses to the [ga]-[da] continua for the two tone conditions, averaged across final consonants. No-tone filler items were not included in the analysis, since interpretable hypotheses focused on context effect differences depending on tone frequencies. Overall, Experiment 1a participants were somewhat more variable in responding, with [d] and especially [g] end points identified at less than 100%. This was not unexpected, since Experiment 1a stimuli lacked speech information immediately following the onset formant transitions that was present in Experiment 1b.

1. Context influence on target consonants

In Experiment 1a, a reliable effect of tone frequency was observed; participants responded [d] (*dot* or *dock*) more often when the higher frequency tone immediately followed the initial consonant than when the lower frequency tone followed it, $t(10)=3.14$, $p=0.011$. This indicates that the immediately adjacent tone contexts of Experiment 1a influenced perception of the preceding consonants assimilatively. Spectral energy in a *higher* frequency range resulted in more [d] (i.e., *high*-F3) responses. Experiment 1b participants also showed a reliable overall effect of the following context tone frequency on categorization, $t(11)=3.81$, $p=0.003$. However, the effect in this case, as in previous studies involving preceding speech (Mann, 1980), preceding nonspeech (e.g., Lotto and Kluender, 1998) and following speech (Mann and Repp, 1980) contexts, was *contrastive* in direction. The *lower*-frequency tone caused listeners to categorize the consonants more often as [da], the alternative with the *higher* F3 frequency.

Thus, later-occurring nonspeech may influence speech categorization, both assimilatively and contrastively. Moreover, it seems that a shift between these two influences (or a

shift in the effectiveness with which one or both of them operates) may be induced by a change of as little as 40 ms in the asynchrony of the nonspeech context with respect to the target speech segments. To confirm this shift in the present between-subjects design, the effects of the nonspeech contexts were compared across participant groups. A 2(tone frequency)×2(tone-onset, Experiments 1a, 1b) mixed-model ANOVA with tone-onset condition as a between-subjects factor revealed a significant interaction, $F(1,21)=21.96$, $p<0.001$, indicating that tone frequency indeed had a differential effect across the two groups as a function of the temporal proximity of the tone to the speech target.

In speculating on the processes driving these context effects, their magnitude (the mean difference in [d] responses across tone conditions) was compared with participants' median reaction times. For Experiment 1a, for which assimilative context effects of temporally adjacent tones were observed, bivariate (Pearson) correlation revealed no relationship between context effect size and reaction time, $p>0.5$. For Experiment 1b where tone onset was somewhat later in the vowel, however, a reliable positive correlation was observed, $r=0.606$, $p=0.037$. Participants who took longer to respond generally showed larger contrastive effects of the following tone on speech categorization. This pattern may simply have been due to a subset of participants whose fast, careless responses led to decreased effect size; however, investigation revealed no similar correlation between response time and any measure of accuracy or consistency. Possible implications of this observation with respect to the workings of a high-level contrastive perceptual mechanism are addressed in Sec. IV.

2. Final consonant effects

Since this experiment involved word decisions that depended on final consonants as well as initial, target consonants, and since the word-final release burst stimuli used had not been previously tested experimentally, we also examined listener responses with respect to these consonants. Listeners responded consistently with the intended consonant (e.g., *dot* or *got* for a final [t] burst) on 94% of trials, indicating that the consonants were indeed unambiguous. This observation was further confirmed by the fact that no effect of tone frequency on final consonant identification was observed in either test condition, $(t)<1$.

A 2(final consonant)×2(tone frequency)×2(condition) mixed model ANOVA with proportion [d] responses as the dependent variable revealed a main effect of final consonant [$F(1,21)=25.5$; $p<0.001$] but no interactions involving the consonant, $F<1$, indicating that listeners robustly categorized the initial consonant as [d] more often when the word ended in [k] than when it ended in [t]. The present results cannot conclusively determine the source of this effect, but it is not surprising. One possibility is bias deriving from word frequency (Connine *et al.*, 1993). While frequency measures are not available for the word *gawk*, it is almost certainly many times less common than any of the other words [Kucera-Francis (1967) written frequencies for *dot, dock,* and *got* are 13, 8, and 482, respectively] which may have

biased listeners disproportionately against this word. Alternatively, the final consonant may have been exerting additional spectrally contrastive perceptual effects on the onset, since the final [k] spectra were probably more similar to those of initial [g] than [d] consonants. In any case, this effect did not seem to interfere in any way with the main focus of the experiment, the effects of embedded tones on onset consonants.

III. EXPERIMENT 2

Experiment 1 demonstrated the existence of backward-operating contrastive influence of nonspeech on speech and addressed one of the two methodological issues that were raised concerning the observation of this influence, namely whether contrastive effects may be obscured by an assimilative effect as a function of temporal proximity to target speech. The results demonstrated that context effects produced by nonspeech context tones that follow a speech target are modulated by the temporal proximity of the context and target sounds. Listeners may assimilate the nonspeech sound as information for the speech target when it follows the speech target immediately, but perceive it in contrast to the target when it occurs somewhat later.

The other potential methodological issue involved ensuring that the acoustic continuity between target and following context was sufficient to drive the contrastive effect seen in Experiment 1b. Experiment 1 showed that the backward nonspeech effects may be observed in certain favorable circumstances where (1) additional speech information follows the nonspeech event and (2) this information is required for a proper response. To determine whether these stimulus and task construction characteristics are essential in observing context effects of later-occurring nonspeech sounds on speech categorization and in investigating the limits of the context effects observed in Experiment 1, an additional study was designed in which the final consonant present in Experiment 1 was either eliminated or was not critical to the required categorization response.

A. Methods

1. Participants

Participants were 16 college-age native English speakers with no known speaking or hearing disorders. Participants were paid at least $7 per hour for their time.

2. Stimuli

There were two conditions tested in Experiment 2. Stimuli for Condition 1 were English CVC words and embedded tones identical to those for which contrastive effects were observed in Experiment 1b. Condition 2 stimuli were CV syllables created by eliminating the final-consonant burst of the Experiment 1b stimuli, leaving only the consonant-vowel and embedded tones.

3. Procedure

Participants heard 486 Condition 1 stimuli [(2 tone conditions+no-tone fillers)×2 final consonants×9 continuum
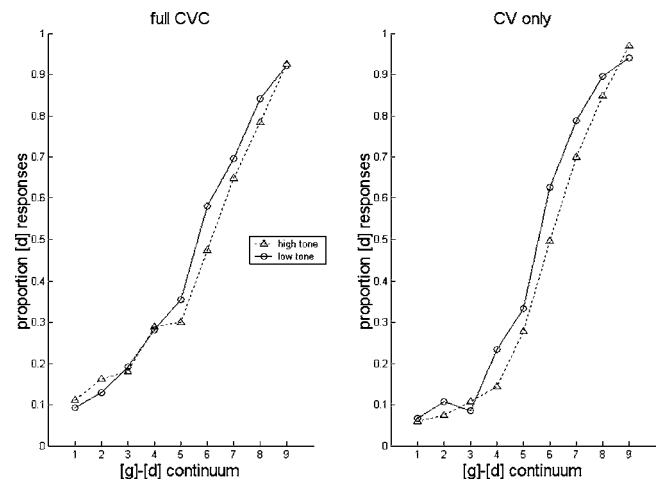
FIG. 3. Initial consonant classification in Experiment 2 conditions.

members×9 repetitions] and 486 Condition 2 stimuli in consecutive sessions separated by a short break. Half of the participants heard Condition 1 stimuli in the first session and Condition 2 stimuli in the second, and the other half heard Condition 2 stimuli first. Stimulus presentation was identical to that of Experiment 1. Rather than making a word decision, however, in both conditions participants were instructed simply to quickly press one of two buttons depending on the first consonant (*d* or *g*) of a stimulus.

B. Results

One participant displayed an irregular, apparently random response pattern; this participant's data were discarded. Figure 2 shows response curves to the initial [g]-[d] continuum as a function of the embedded tone for the remaining 15 participants. As in Experiment 1b, tone frequency reliably influenced consonant identification in a contrastive manner; participants more often reported hearing an initial [d] in the presence of the lower-frequency (1800 Hz) tone in both CVC, $t(15)=2.62$, $p=.02$, and CV, $t(14)=3.95$, $p=0.001$, syllables. For neither condition did the size of the effect $[p([d],\text{low tone})-p([d],\text{high tone})]$ differ reliably from that of Experiment 1b, indicating that neither the CVC word context nor the lexical task requiring the final syllable in responses was essential in driving the effect.

Figure 3 shows a comparison of the effect sizes across Experiment 2 conditions and session orders. A 2(condition, syllable type)×2(session order) mixed model ANOVA with session order as a between-subjects factor revealed that participants demonstrated slightly larger contrastive effects overall when no final consonant was present, $F(1,13)=8.83$, $p=0.011$. Additionally, across conditions a robustly larger contrastive effect was seen in the participants' second sessions, without respect to whether the stimuli were CVCs or CVs, resulting in a Condition×Session Order interaction $(F(1,13)=21.8, p<0.001)$. This does not seem to have been an artifact of the task or an effect of practice. In Experiment 1b, which was identical to Experiment 2 except for the response choices and the shift in syllable structure between sessions, there was a nonsignificant shift in the opposite direction, with participants showing

T. Wade and L. L. Holt: Backward effects of nonspeech on speech

greater effects in the first session. A 2(Experiment 1b versus 2) × 2(Session) mixed model ANOVA with effect size as the dependent variable revealed a significant interaction, $F(1,25)=9.35$, $p=0.005$, confirming that a session-to-session pattern different from Experiment 1b emerged in Experiment 2. Implications of this pattern are discussed in the following section.

IV. GENERAL DISCUSSION

The present study demonstrates that later-occurring nonspeech acoustic events can influence categorization of preceding speech. In two experiments, the periodicity of brief pure tones inserted after members of a target [g]-[d] continuum reliably affected listeners' speech identification. Like previously documented effects of preceding speech (Mann, 1980) and nonspeech (Holt, 2005; Lotto and Kluender, 1998) on speech perception, of preceding speech on nonspeech perception (Stephens and Holt, 2003), and following speech on speech perception (Mann and Repp, 1980), this influence was primarily contrastive in nature. Higher-frequency following tones caused ambiguous speech targets to be identified as the lower-F3 consonant alternative ([g]), and *vice versa*.

This finding adds to the array of findings in which nonspeech acoustic context has been shown to parallel speech context in its perceptual influence on speech. As such, it is consistent with the view that a general contrastive perceptual mechanism may play a role in driving the context-dependent nature of speech perception and calls into question the extent to which specialized knowledge or processes specific to articulation are needed to explain context-dependent speech perception. In particular, demonstration that nonspeech context effects are not constrained in the temporal direction of their operation increases the parsimony with which a contrast-based explanation can account for the set of observed speech context effects. Like feature parsing (Gow, 2004; Gow, 2003) and gesture recovery (Fowler, 1986; Fowler *et al.*, 2000) accounts, contrast effects could potentially explain observed perceptual compensation for both progressively and regressively coarticulated speech. This is not, however, to say that the contrastive effect demonstrated here is the only factor contributing to the context-dependent nature of speech perception. As is often noted in conjunction with proposals of contrastive effects, it seems likely that speech- or language-specific learning also plays a role (Diehl *et al.*, 2004; Holt and Kluender, 2000; Lotto, 2000). Since contrast effects do not arise as compensation for speech production, it is unreasonable to predict that they would result in ideal perceptual compensation for all possible patterns of coarticulation, which probably vary somewhat from language to language. Although general contrastive effects probably simplify the perceptual problem by perceptually eliminating some context-conditioned spectral variability, residual variability might be learned as part of language acquisition. Indeed, there is evidence that speakers of different languages process context-dependent variability differently. Beddor, Harnsberger, and Lindemann (2002), for example, observe cross-language differences in perceptual compensa-

tion for vowel coarticulation, citing differences between English and Shona speakers' category boundary shifts depending on the position of context-providing vowels.

The results of Experiment 1 also demonstrate the sensitivity of nonspeech-induced context effects to the spectrotemporal characteristics of the stimuli. A robust shift from assimilative to contrastive effects resulted from a 40 ms change in the temporal proximity of the nonspeech context to the target. This observation underscores the variety of ways nonspeech (or speech) context sounds might influence speech perception and also suggests some caution in the interpretation of null effects of nonspeech context (e.g. Fowler *et al.*, 2000), since such observations might stem from the operation of multiple, competing perceptual effects like those observed here.

A. Temporal range of context effects

A principle concern in designing Experiment 1b and Experiment 2 stimuli was to provide sufficient temporal proximity between target and context stimuli for contrastive processes to operate. We did observe contrastive influences when the context was placed some 40 ms after the target in the following vowel segment, but the present experiments do not address whether these influences might persist if the context were further separated in time. Newman and Sawusch (1996) observed that perception of a segmental contrast involving duration may only be affected by speaking rate context information that follows the target segment by no more than a phoneme or two. A parallel study involving the limits of following nonspeech context will be informative in further comparing the perceptual influences of speech and nonspeech contexts.

B. Contrast and higher levels of linguistic processing

These results and those of previous studies involving nonspeech context effects on speech categorization (e.g., Holt, 2005; Lotto *et al.*, 2003; Wade and Holt, in press) suggest that a *higher-level* (central), *nonlinguistic* contrastive mechanism may play a role in the interpretation of speech sounds. However, we wish to emphasize that such a mechanism need not be independent of, or incompatible with, the various language-specific processes involved in human speech perception. In particular, the present results resonate with spoken word recognition models that posit interaction between multiple representational levels over the course of perception. In the TRACE model (McClelland and Elman, 1986), for example, it is assumed that speech processing takes the form of activation at three hierarchically arranged levels: acoustic features, abstract phonemes, and words. Although the initial input to the model involves the feature units, there are bidirectional excitatory connections between units at neighboring levels; critically, the units at all levels corresponding to all of the segments in a word remain active—and continue to interact—for the entire time a word is processed. While we will not argue that speech processing involves precisely these three levels of representation, it follows straightforwardly from the interactive dynamics of the model that if perception of the acoustic events comprising

speech sequences is assumed to be modulated by the acoustic context—speech or nonspeech—preceding and following the events, the unit activations at lower representational levels might adjust over time as context becomes available. As input representations evolve over time, the activation of the higher-level units they excite should also change, although the effects of the change would be delayed by factors including the strength of the top-down connections of higher-level units already activated by the precontext feature representations. Thus, assuming a general auditory representation as a lower-level input representation to an interactive model like TRACE provides a means by which a general contrastive mechanism might interact with higher-level, language-specific processes over the course of perception.[1]

Experiment 1b revealed a pattern in listeners' reliance on nonspeech context that might shed some light on the dynamics of such interaction. As response latencies increased, so did the magnitude of the contrastive effect of following tone frequency on their speech categorization. While care should be taken not to over-interpret this correlational observation, it is consistent with the notion that top-down influences delay the effects of acoustic context at the input level. Since this process might continue well beyond the physical duration of the word, delaying word recognition (presumably reflected in the longer response times) should allow the context to exert a greater contrastive effect.

The results of Experiment 2 demonstrate the robustness of the backward contrastive effect and provide some additional evidence regarding the types of processing it might involve. In two conditions where participants were asked to label only the initial consonants of stimuli as [d] or [g], later-occurring tones identical to those used in Experiment 1b evoked similar contrastive influences on categorization. This indicates that the CVC word identification task, which required participants to make a (word) response based on parts of a speech signal both preceding and following the nonspeech tone, was in fact not critical in producing the backward effect. Additionally, participants showed a slightly greater contrastive effect when the stimuli were CV nonwords than when they were CVC words. The cause of this difference is not clear; it may have been that the additional later-occurring context provided by the final consonant actually obscured the contrastive effect of the tones. Alternatively, the lexical status of the stimuli may have played a role. It has been observed that listeners are faster at phoneme monitoring in word than in nonword contexts (Cutler *et al.*, 1987; Eimas *et al.*, 1990), though not always with stimuli as short as those used here (Foss and Gernsbacher, 1983). If, as suggested by these authors, word stimuli resulted in a different mode of processing in for the words in Condition 1 than the nonwords in Condition 2, the resulting faster processing of onsets in word stimuli might have resulted in less contrastive influence of later-occurring context. Reaction time comparisons in this experiment were confounded by the additional differences in syllable length and structure (CV vs CVC) across conditions and were thus uninformative in this respect.

Comparison of effects across Experiment 2 conditions also revealed an interaction involving session order. Regard-
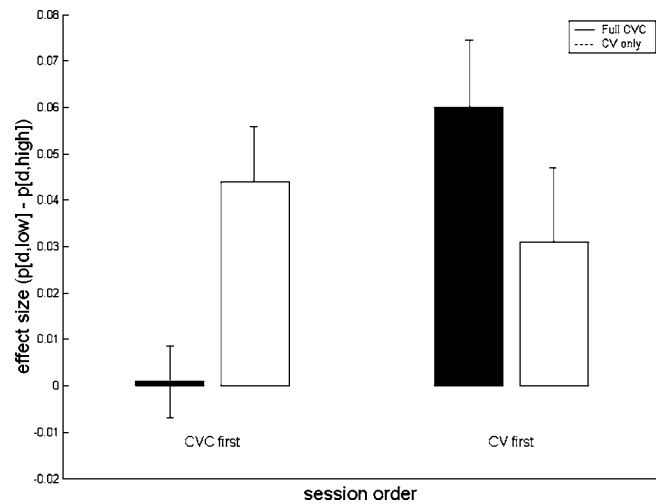
FIG. 4. Experiment 2 contrastive effect sizes across syllable types and session orders (error bars indicate standard errors).

less of syllable structure, participants exhibited reliably greater contrastive effects during their second test sessions, an effect not observed in Experiment 1. [See Fig. 4.] While further study will be needed to determine the precise causes of this difference, it seems likely that attention played a role. Second test session stimuli always involved novel word-final information compared to those of the first session (either the addition or the sudden absence of a coda consonant); it seems likely that this novelty resulted in increased attention to portions of the stimulus following the initial onset, perhaps including the nonspeech tones. This increased attention, then, might have enhanced the tones' contrastive influence on the onset consonants, resulting in the session-to-session effect-size difference seen in Experiment 2.

C. Masking as a possible account

Effects of preceding nonspeech acoustic context on speech categorization have previously been attributed to auditory masking (Fowler *et al.*, 2000), although this interpretation is challenged by the extended time courses and dichotic presentation paradigms (Holt, 2005; Lotto *et al.*, 2003) for which the nonspeech context effects have been observed. Holt (2005), for example, presents evidence that nonspeech precursor contexts influence speech perception even when 13 constant acoustic stimuli spanning 1.3 s intervene between nonspeech context and speech targets. Masking is very unlikely to account for such context effects of nonspeech on speech. Nonetheless, the backward-operating effects documented here might similarly be attributed to a process such as backward masking (e.g., Tyler and Small, 1977) or informational masking (Neff *et al.*, 1993; Pollack, 1975). By selectively masking spectral energy in the neighborhood of 2800 Hz, for example, a higher-frequency following tone could cause an ambiguous consonant to be perceived as more [g]-like, resulting in contrastive effects as reported here. The present experiments do not rule out these possibilities; indeed, given the relatively poor current understanding of the precise mechanisms responsible for backward and informational masking, it is difficult to determine

whether context-dependence in natural speech might derive from similar or related processes. Backward masking, for example, has been observed to operate even when the masking stimulus is presented to the ear contralateral to the target (Weber and Green, 1979), indicating a more central mechanism. In fact, backward masking has been mentioned as a potential contributor to the context-dependent perception of various speech contrasts (Dent *et al.*, 1997; Jamieson, 1987; Sinnott *et al.*, 1998). Further research and a better understanding of these processes will be required to determine whether this is indeed the case.

D. Conclusions

This study demonstrates that later-occurring nonspeech events contrastively influence the categorization of preceding speech sounds in certain circumstances. Contrastive backward effects of pure tones on onset consonant identification were observed across syllable structures (CV and CVC) and tasks (word versus phoneme identification), although they depended on the temporal proximity of nonspeech tones to the preceding speech target and perhaps on attention to later portions of the stimuli. These findings are taken as support of a central, nonlinguistic contrastive perceptual mechanism, whereby auditory events occurring over time are incorporated into an acoustic context, relative to which preceding or following speech sounds are perceived.

[1]The original TRACE model, it should be noted, has a different, phonologically oriented account of context-dependent perception which does not account for nonspeech context effects (McClelland and Elman, 1986).

Beddor, P. S., Harnsberger, J. D., and Lindemann, S. (**2002**). "Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates," J. Phonetics **30**, 591–627.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Ciocca, V. and Bregman, A. S. (**1989**). "The effects of auditory streaming on duplex perception," Percept. Psychophys. **46**, 39–48.

Connine, C. M., Titone, D., and Wang, J. (**1993**). "Auditory word recognition: Extrinsic and intrinsic effects of word frequency," Cognition **19**, 81–94.

Cutler, A., Mehler, J., Norris, D., and Segui, J. (**1987**). "Phoneme identification and the lexicon," Cogn. Psychol. **19**, 141–177.

Dent, M. L., Brittan-Powell, E. F., Dooling, R. J., and Pierce, A. (**1997**). "Perception of synthetic /ba/-/wa/ speech continuum by budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **102**, 1891–1897.

Diehl, R., Lotto, A., and Holt, L. L. (**2004**). "Speech perception," Annu. Rev. Psychol. **55**, 149–179.

Eimas, P., Hornstein, S. B., and Payton, P. (**1990**). "Attention and the role of dual codes in phoneme monitoring," J. Mem. Lang. **31**, 375–395.

Foss, D. J. and Gernsbacher, M. A. (**1983**). "Cracking the dual code: Toward a unitary model of phoneme identification," J. Verbal Learn. Verbal Behav. **22**, 609–632.

Fowler, C. A. (**1986**). "An event approach to the study of speech perception from a direct-realist perspective," J. Phonetics **14**, 3–28.

Fowler, C. A., Best, C. T., and McRoberts, G. W. (**1990**). "Young infants' perception of liquid coarticulatory influences on following stop consonants.," Percept. Psychophys. **48**, 559–570.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). "Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans," J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Gaskell, M. (**2003**). "Modelling regressive and progressive effects of assimilation in speech perception," J. Phonetics **31**, 447–463.

Gow, D. (**2004**). "A cross-linguistic examination of assimilation context effects," J. Mem. Lang. **51**, 279–296.

Gow, D. W. (**2003**). "Feature parsing: Feature cue mapping in spoken word recognition," Percept. Psychophys. **65**, 575–590.

Holt, L. L. (**1999**). "Auditory constraints on speech perception: An examination of spectral contrast," Diss. Abstr. Int., C **61**, 556.

Holt, L. L. (**2005**). "Temporally non-adjacent non-linguistic sounds affect speech categorization," Psychol. Sci. **16**, 305–312.

Holt, L. L. and Kluender, K. R. (**2000**). "General auditory processes contribute to perceptual accommodation of coarticulation," Phonetica **57**, 170–180.

Holt, L. L. and Lotto, A. J. (**2002**). "Behavioral examinations of the level of auditory processing of speech context effects," Hear. Res. **167**, 156–169.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.

Jamieson, D. (**1987**). "Studies of possible psychoacoustic factors underlying speech perception," in *The Psychophysics of Speech Perception*, edited by M. E. H. Schouten (Nijhoff, Dordrecht).

Kucera, H. and Francis, W. N. (**1967**). *Computational Analysis of Present-day American English* (Brown University Press, Providence).

Liberman, A., Isenberg, D., and Rakerd, B. (**1981**). "Duplex perception of cues for stop consonants: Evidence for a phonetic mode," Percept. Psychophys. **30**, 133–143.

Liberman, A. M. and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Lotto, A., Sullivan, S., and Holt, L. L. (**2003**). "Central locus for nonspeech context effects on phonetic identification," J. Acoust. Soc. Am. **113**, 53–56.

Lotto, A. J. (**2000**). "Language acquisition as complex category formation," Phonetica **57**, 189–196.

Lotto, A. J. and Kluender, K. R. (**1998**). "General contrast effects of speech perception: Effect of preceding liquid on stop consonant indentification.," Percept. Psychophys. **60**, 602–619.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1997**). "Perceptual compensation for coarticulation by Japanese quail (*Coturnix coturnix japonica*)," J. Acoust. Soc. Am. **102**, 1134–1140.

Mann, V. A. (**1980**). "Influence of preceding liquid on stop-consonant perception," Percept. Psychophys. **28**, 407–412.

Mann, V. A. (**1986**). "Distinguishing universal and language-dependent levels of speech perception: Evidence from Japanese listeners' perception of English 'l' and 'r,' " Cognition **24**, 169–196.

Mann, V. A. and Repp, B. H. (**1980**). "Influence of vocalic context on perception of the [sh]-[s] distinction," Percept. Psychophys. **28**, 213–228.

Mann, V. A. and Soli, S. (**1991**). "Perceptual order and the effect of vocalic context on fricative perception," Percept. Psychophys. **49**, 399–411.

McClelland, J. L. and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.

McCollough, C. (**1965**). "Color adaptation of edge-detectors in the human visual system," Science **149**, 1115–1116.

Miller, J. L. and Dexter, E. R. (**1988**). "Effects of speaking rate and lexical status on phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **14**, 369–378.

Miller, J. L. and Liberman, A. M. (**1979**). "Some effects of later-occurring information on the perception of stop consonant and semivowel," Percept. Psychophys. **25**, 457–465.

Nearey, T. M. (**1997**). "Speech perception as pattern recognition," J. Acoust. Soc. Am. **101**, 3241–3254.

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.

Newman, R. S. and Sawusch, J. R. (**1996**). "Perceptual normalization for speaking rate: Effects of temporal distance," Percept. Psychophys. **58**, 540–560.

Pollack, I. (**1975**). "Auditory informational masking," J. Acoust. Soc. Am. **57**, S5.

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**(1), 81-110.

Sinnott, J. M., Brown, C. H., and Borneman, M. A. (**1998**). "Effects of syllable duration on stop-glide identification in syllable-initial and syllable-final position by humans and monkeys," Percept. Psychophys. **60**, 1032–1043.

Smits, R. (**2001**). "Evidence for hierarchical categorization of coarticulated phonemes," J. Exp. Psychol. Hum. Percept. Perform. **27**, 1145–1162.

Stephens, J. D. W. and Holt, L. L. (**2003**). "Preceding phonetic context affects perception of non-speech sounds," J. Acoust. Soc. Am. **114**, 3036–3039.

Tyler, R. S. and Small, A. M. (**1977**). "Two-tone suppression in backward masking," J. Acoust. Soc. Am. **62**, 215–218.

Wade, T. and Holt, L. L. "Perceptual effects of preceding non-speech rate on temporal properties of speech categories," Percept. Psychophys. (in press).

Walker, J. T. and Irion, A. L. (**1979**). "Two new contingent aftereffects: Perceived auditory duration contingent on pitch and on temporal order," Percept. Psychophys. **26**, 241–244.

Weber, D. L. and Green, D. M. (**1979**). "Suppression effects in backward and forward masking," J. Acoust. Soc. Am. **65**, 1258–1267.

Wertheimer, M. (**1923**). "Untersuchungen zur Lehre der Gestalt" ("Investigations of Gestalt Theory"), Psychol. Forsch. **4**, 301–350.

Voice gender identification by cochlear implant users: The role of spectral and temporal resolution

Qian-Jie Fu,[a] Sherol Chinchilla, Geraldine Nogaki, and John J. Galvin III
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

The present study explored the relative contributions of spectral and temporal information to voice gender identification by cochlear implant users and normal-hearing subjects. Cochlear implant listeners were tested using their everyday speech processors, while normal-hearing subjects were tested under speech processing conditions that simulated various degrees of spectral resolution, temporal resolution, and spectral mismatch. Voice gender identification was tested for two talker sets. In Talker Set 1, the mean fundamental frequency values of the male and female talkers differed by 100 Hz while in Talker Set 2, the mean values differed by 10 Hz. Cochlear implant listeners achieved higher levels of performance with Talker Set 1, while performance was significantly reduced for Talker Set 2. For normal-hearing listeners, performance was significantly affected by the spectral resolution, for both Talker Sets. With matched speech, temporal cues contributed to voice gender identification only for Talker Set 1 while spectral mismatch significantly reduced performance for both Talker Sets. The performance of cochlear implant listeners was similar to that of normal-hearing subjects listening to 4–8 spectral channels. The results suggest that, because of the reduced spectral resolution, cochlear implant patients may attend strongly to periodicity cues to distinguish voice gender. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1985024]

I. INTRODUCTION

Speech recognition primarily involves the extraction of linguistic meaning. However, perception of some secondary features such as voice gender and/or talker identity is not only important for understanding the nature of the information conveyed by spoken language (e.g., Scherer, 1986), but also for recovering linguistic information despite acoustic variability among talkers (e.g., Pisoni, 1997). Voice gender and talker identity are two important aspects of indexical cueing in speech. Previous studies have shown that individual speech utterances include acoustic correlates of voice gender and individual talker identity; the fundamental frequency (F0) and the formant frequencies are two important measures for voice gender classification (e.g., Bachorowski and Owren, 1999).

F0, which corresponds to perceived voice pitch, is determined by the pulsing of the vocal folds. On average, the vibrating segment of each vocal fold is 60% longer in males; typically, the F0 of male talkers is 45% lower that of female talkers (Titze, 1989). The supralaryngeal vocal tract can be described in terms of resonance characteristics (i.e, formant frequencies). On average, the supralaryngeal vocal-tract length (VTL) is about 15% greater in males (Fant, 1966); formant frequencies tend to be 15% lower in male talkers than in female talkers for the same vowel (Bachorowski and Owren, 1999; Lieberman and Blumstein, 1993). Although combining F0 and formant frequencies provide the most ac-

curate voice gender identification, F0 or formant frequencies alone can provide sufficient cues to identify voice gender (Bachorowski and Owren, 1999).

Cochlear implant (CI) patients do not typically have access to the rich spectral and temporal information available to normal-hearing (NH) listeners. Despite these reduced spectro-temporal cues, many CI users are capable of some degree of voice gender identification. Recent studies with NH listeners have demonstrated that voice gender identification is possible in conditions of reduced spectral resolution, such as 3-sinewave replicas of natural speech (Remez *et al.*, 1997; Sheffert *et al.*, 2002). A few studies have also explored voice gender and speaker identification ability in CI listeners (Cleary and Pisoni, 2002; Spahr and Dorman, 2004; Fu *et al.*, 2004). Fu *et al.* (2004) explored the relative importance of temporal and spectral cues in voice gender identification and vowel recognition by CI users and NH subjects listening to an acoustic simulation of CI speech processing. In the acoustic simulation, the number of spectral channels ranged from 4 to 32, thereby varying the spectral resolution; the cut-off frequencies of the channels' envelope filters ranged from 20 to 320 Hz, thereby varying the available temporal cues. For NH subjects, results showed that both voice gender identification and vowel recognition scores improved as the number of spectral channels was increased. When only 4 spectral channels were available, voice gender identification significantly improved as the envelope filter cut-off frequency was increased from 20 to 320 Hz. For all spectral conditions, increasing the amount of temporal information had no significant effect on vowel recognition. Both voice gender identification and vowel recognition scores

[a] Electronic mail: qfu@hei.org

TABLE I. Relevant information for cochlear implant subjects who participated in the experiment.

Subject	Age	Etiology	Prosthesis	Age of onset hearing loss	Years with prosthesis	Speech strategy
S1	64	Trauma/unknown	N-22	45	14	SPEAK
S2	48	Trauma	N-22	35	12	SPEAK
S3	62	Hereditary	N-22	20	13	SPEAK
S4	74	Noise induced	N-22	40	8	SPEAK
S5	72	Unknown	N-24	37	5	ACE
S6	62	Hereditary	N-24	40	2	ACE
S7	35	Pre-natal rubella	MedEL	0	3	CIS+
S8	41	Congenital	MedEL	7	5	CIS+
S9	71	Unknown	CII	65	4	HiRes
S10	55	Unknown	CII	49	6	HiRes

were highly variable among CI users. CI listeners' performance was similar to that of NH subjects listening to 4–8 spectral channels.

Taken together, the results suggest that both spectral and temporal cues contribute to voice gender identification, and that temporal cues may be especially important for CI users because of the reduced spectral resolution. F0 information, while not explicitly coded in present-day CI speech processors, is well-represented by the temporal envelope fluctuations in each channel, making voice gender identification possible in the absence of explicit place coding of F0 information. CI patients' access to these periodicity cues may be limited by their sensitivity to modulation frequency. Although temporal processing in the F0 range (80–500 Hz) may be relatively poor, CI patients may be able to distinguish voice gender if the F0's of the male and female talkers are widely separated, as the voice qualities associated with these periodicity cues may be sufficiently different to identify voice gender. However, if the F0 of a male voice is relatively high or the F0 of a female voice is relatively low, it is unclear whether CI users will be able to identify the talker gender.

When there is sufficient spectral resolution to preserve formant structure (16 or more spectral channels), high levels of voice gender identification are possible, even when temporal cues are severely reduced (Fu *et al.*, 2004). CI patients' perception of spectral profiles is determined by the number of spectral channels (electrodes) and by the degree of mismatch between acoustic signal and the electrode location. Because of the limited insertion depth of the electrode array, all CI patients are subject to some degree of spectral mismatch as the acoustic frequency range is allocated to the implanted electrodes. For patients with shallow insertion depths, the spectral shift may be more severe and may significantly distort the spectral profile. While previous results revealed a significant interaction between spectral resolution and temporal cues (Fu *et al.*, 2004), it is not clear how spectral mismatch might affect the relative contribution of temporal cues to voice gender identification.

In the present study, voice gender identification was measured in CI listeners and in NH subjects listening to acoustic simulations of CI processing. CI subjects were tested using their everyday speech processors while NH subjects were tested using sinewave processors that simulated various degrees of spectral resolution, temporal resolution,

and spectral mismatch. To more deeply explore the relative contributions of spectral and temporal cues to voice gender identification, two talker sets were tested, in which there was either a large or small difference between the F0's of male and female talkers. We predicted that both spectral and temporal cues would significantly contribute to voice gender identification when the F0 differences between male and female talkers were relatively large. We also expected that spectral cues would contribute more strongly and temporal cues less strongly when the difference between the F0's of male and female talkers was relatively small.

II. METHODS

A. Subjects

Ten NH subjects (3 males and 7 females, 22–38 years old) and 10 CI users (7 males and 3 females, 35–74 years old) participated in the study. All NH subjects had pure tone thresholds better than 15 dB HL at octave frequencies from 250 to 8000 Hz; CI subject details are listed in Table I. All subjects were native English speakers. All subjects were paid for their participation.

B. Test materials

The tokens used for closed-set voice gender identification tests were digitized natural productions from 10 males and 10 females, drawn from speech samples collected by Hillenbrand *et al.* (1995). For each talker, there were 12 phonemes, including 10 monophthongs (/i ɪ ɛ æ u ʊ ɑ ʌ ɔ ɝ/) and 2 diphthongs (/o e/), presented in a /h/-vowel-/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, hayed). For each talker, F0 values for each token were extracted by an autocorrelation method (Markel, 1972); the mean of these F0 values across the 12 tokens was calculated to be the mean F0 of the talker. After calculating the mean F0 for each talker, talkers were divided into two groups, each of which contained 5 male and 5 female talkers. Group assignment was solely determined by the mean F0 of each talker. Five female talkers with the highest F0 values and five male talkers with the lowest F0 values were grouped together as Talker Set 1 while the remaining talkers were grouped together as Talker Set 2.

TABLE II. Acoustic measures for each talker in Talker Sets 1 and 2. F0 is the fundamental frequency, F1, F2, and F3 refer to the lowest three formant frequencies, respectively, and VTL refers to vocal-tract length. Note that the formant frequency measurements are for the vowel /ɛ/ as in "head" only.

| | Talker Set 1 (Nonoverlapping Talker Set) | | | | | | | | | |
| | Male talkers | | | | | Female talkers | | | | |
Measure	M1	M2	M3	M4	M5	F1	F2	F3	F4	F5
F0 (Hz)	95	128	117	138	148	255	212	238	223	202
F1 (Hz)	562	627	649	584	627	735	757	800	757	643
F2 (Hz)	1686	1729	1751	1751	1881	2097	1902	1967	2097	1961
F3 (Hz)	2421	2529	2551	2594	2616	2789	2918	2918	3135	2745
VTL (cm)	16.12	15.21	14.95	15.35	14.63	13.14	13.22	12.87	12.46	14.06
	Talker Set 2 (Overlapping Talker Set)									
F0 (Hz)	170	155	166	160	196	182	188	175	197	154
F1 (Hz)	627	596	580	612	612	737	659	596	659	659
F2 (Hz)	1882	1678	1757	1788	2024	2071	2196	2055	1992	2353
F3 (Hz)	2651	2729	2651	2714	2635	2745	2839	2965	2824	3090
VTL (cm)	14.54	15.18	15.24	14.78	14.37	13.26	13.32	13.83	13.74	12.64

Acoustic analysis of F0 and formant frequencies F1, F2, and F3 for the vowel "head" was conducted for all talkers using custom software; analysis was conducted on the midpoint of the steady portion of the vowel. Linear predictive coding (LPC) was used to derive the frequencies of F1, F2, and F3 using 14 coefficients. All formant frequencies were directly measured from the LPC-based spectral envelope using cursors, while a simultaneously displayed wideband spectrogram served as a reference. Table II lists the acoustic measures for each talker in Talker Sets 1 and 2. Note that for each talker, F0 values were averaged across all vowel stimuli while F1, F2, and F3 values were for the vowel "head" only. Each talker's VTL was estimated from the formant frequency measurements using the following equation (adapted from Lieberman and Blumstein, 1993):

$$\text{VTL} = \frac{(2k + 1)c}{4F_{k+1}},$$

where $k=(0,1,2)$, F_{k+1} is the formant frequency, c is the speed of sound (34 400 cm/s). Individual calculations were made for each of the three formant frequencies and the mean of these calculations was used to estimate talkers' VTL. In Talker Set 1, the F0 values between male and female talkers differed greatly; there was no overlap in F0 between any pair of male and female talkers. Mean F0 values were 125 and 226 Hz for male and female talkers, respectively. A one-way ANOVA test showed a significant difference in mean F0 values between male and female talkers in Talker Set 1 $[F(1,8)=59.33, p<0.001]$. Mean VTL values were 15.25 and 13.15 cm for male and female talkers, respectively. A one-way ANOVA test showed a significant difference in VTL values between male and female talkers in Talker Set 1 $[F(1,8)=33.37, p<0.001]$. In Talker Set 2, there was a much smaller difference between the F0 values of male and female talkers; male talkers were higher-pitched and female talkers were lower-pitched, relative to Talker Set 1. There was significant overlap between the F0's of male and female talkers. Mean F0 values were 169 and 179 Hz for male and female talkers, respectively. A one-way ANOVA test showed no

significant difference in mean F0 values between male and female talkers $[F(1,8)=0.90, p=0.37]$. Mean VTL values were 14.82 and 13.36 cm for male and female talkers, respectively. A one-way ANOVA test showed a significant difference in VTL values between male and female talkers $[F(1,8)=29.06, p<0.001]$.

Acoustic analysis of formant frequencies F1 and F2 for an additional three vowels (/i æ u/) was conducted. Two measures, which would reflect the vowel space size of the talker, were derived from the formant frequency measurements (Hazan and Markham, 2004). One measure is the difference between the first format frequencies for /i/-/æ/ while the other measure is the difference between the second formant frequencies for /i/-/u/. In Talker Set 1, the mean difference in F1 values for /i/-/æ/ was 233 and 244 Hz for male and female talkers, respectively. The mean difference in F2 values for /i/-/u/ was 1284 and 1506 Hz for male and female talkers, respectively. In Talker Set 2, the mean difference in F1 values for /i/-/æ/ was 242 and 257 Hz for male and female talkers, respectively, while the mean difference in F2 values for /i/-/u/ was 1308 and 1427 Hz for male and female talkers, respectively. A two-way ANOVA test showed a significant difference in vowel space size between male and female talkers $[F(1,36)=4.758, p=0.036]$ but no significant difference in vowel space size between Talker Set 1 and Talker Set 2 $[F(1,36)=0.307, p=0.583]$.

C. Signal processing

For NH listeners, sinewave processors were used to simulate a CI speech processor fitted with the continuously interleaved sampling (CIS) strategy (Wilson et al., 1991). The processor was implemented as follows. The signal was first processed through a pre-emphasis filter (high-pass with a cut off frequency of 1200 Hz and a slope of 6 dB/octave). An input frequency range (100–5000 Hz) was band-passed into a number of frequency analysis bands (4, 8, 16, or 32 bands) using Fourth-order Butterworth filters; the distribution of the analysis filters was according to Greenwood's (1990) formula. The corner frequencies (3 dB down) for the
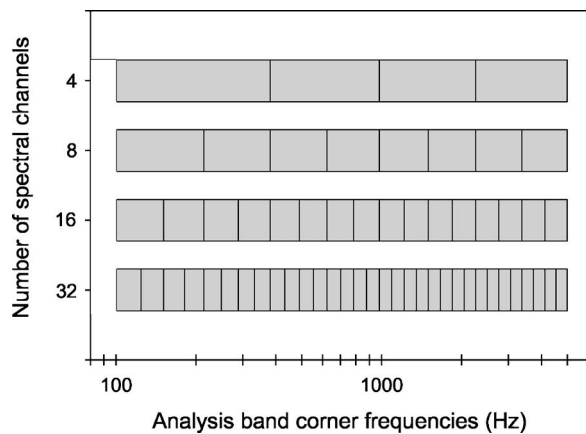
FIG. 1. Corner frequencies for analysis band filters used in cochlear implant simulation.

analysis band filters are shown in Fig. 1. The temporal envelope was extracted from each frequency band by half-wave rectification and low-pass filtering. The cut-off frequency of the envelope filter was varied according to the temporal envelope experimental conditions (40, 80, or 160 Hz). After extracting the temporal envelope for each channel, a sinusoidal carrier was generated for each channel; the carrier frequency depended on the experimental condition. For the matched condition, the carrier frequency was equal to the center frequency of the analysis filter. For the shifted condition, the output carrier bands were shifted up one octave relative to the analysis bands to simulate the spectral shift associated with CIs; the frequency of the sinewave carrier was equal to the center frequency of the upshifted carrier bands. A one-octave shift was utilized to reduce the contribution of the spectral profile (or formant frequencies) to voice gender identification, thereby forcing listeners to make use of temporal cues. Also, the spectral shift approximated the mismatch between acoustic input frequency and electrode location experienced by CI patients, due to the limited insertion depth of the electrode array. The extracted temporal envelope from each band was used to modulate the corresponding sinusoidal carrier. The amplitude of the modulated sinewave was adjusted to match the rms energy of the temporal envelopes. Finally, the modulated carriers of each band were summed and the overall level was adjusted to be equal to that of the original, unprocessed speech.

All CI subjects were tested using their clinically assigned speech processors (details of CI subjects' speech processors are shown in Table I). Subjects were instructed to use their normal volume and sensitivity settings, and not to use any noise suppression settings. Depending on the implanted device, four different speech-processing strategies were used by the CI subjects in their clinically assigned speech processors. Four Nucleus-22 patients were fitted with the SPEAK strategy (Seligman and McDermott, 1995). Two Nucleus-24 patients were fitted with the ACE strategy (Arndt *et al.*, 1999). Two Med-El patient devices were fitted with the CIS strategy (Wilson *et al.*, 1991). Two Clarion II patients were fitted with the high-resolution speech coding strategy (HiRes).

D. Procedure

Voice gender identification was measured using a closed-set, 2AFC paradigm. For each trial, a stimulus token was chosen randomly, without replacement, from the stimulus set and presented to the subject. The subject responded by clicking on one of two response choices displayed on a computer screen; the response buttons were labeled "male" or "female." If subjects were unsure of the correct answer, they were instructed to guess as best that they could. To minimize any learning effects, no preview of the stimuli was allowed and no feedback was provided. Voice gender identification was tested for the two Talker Sets with CI listeners using their clinically assigned speech processors. A minimum of three test runs was completed for each Talker Set, and the test order was randomized between the two Talker Set conditions. For NH listeners, the test order was randomized among Talker Set conditions, spectral resolution conditions (4, 8, 16, or 32 channels), temporal envelope conditions (envelope filter cut-off frequencies of 40, 80, or 160 Hz) and spectral shift conditions. A minimum of three test runs was completed for each test condition. Both NH and CI listeners were tested in free field in a sound-treated booth (IAC). Stimuli were presented over a single loudspeaker (Tannoy Reveal) at 65 dBA (long-term average level). This presentation level corresponded to conversational speech levels and was well-matched to the acoustic input dynamic range of CI users' speech processors.

III. RESULTS

Figure 2 shows mean voice gender identification results for CI and NH subjects. Panels (a) and (c) show data for Talker Set 1 and panels (b) and (d) show data for Talker Set 2. Panels (a) and (b) show data for the spectrally matched condition and panels (c) and (d) show data for spectrally shifted condition. In all panels, closed symbols represent data from NH listeners while open symbols represent data from CI users. As shown in the figure, CI listeners generally performed very well with Talker Set 1; mean voice gender recognition across CI users was 93.83% correct. Mean CI performance dropped by nearly 26 percentage points for Talker Set 2 (67.91% correct). Because of the limited number of CI subjects, the uneven distribution of devices across CI patients and the variability in performance among patients, it was not possible to correlate patient performance with device type or speech processing strategy.

Similar to CI users, the best NH performance was with Talker Set 1. For Talker Set 1 with spectrally matched speech [panel (a)], performance with relatively low temporal envelope cut-off frequencies (40 Hz) sharply increased as more spectral channels were added. Similarly, voice gender identification with relatively low spectral resolution (4 spectral channels) sharply improved as the amount of temporal information was increased. A two-way ANOVA revealed main effects for spectral resolution [$F(3,108)=344.07$, $p<0.001$] and envelope cut-off frequency [$F(2,108)=118.56$, $p<0.001$]; there was a significant interaction between the spectral resolution and the amount of temporal information [$F(6,108)=42.01$, $p<0.001$]. *Post-hoc* Bonferroni *t*-tests re-
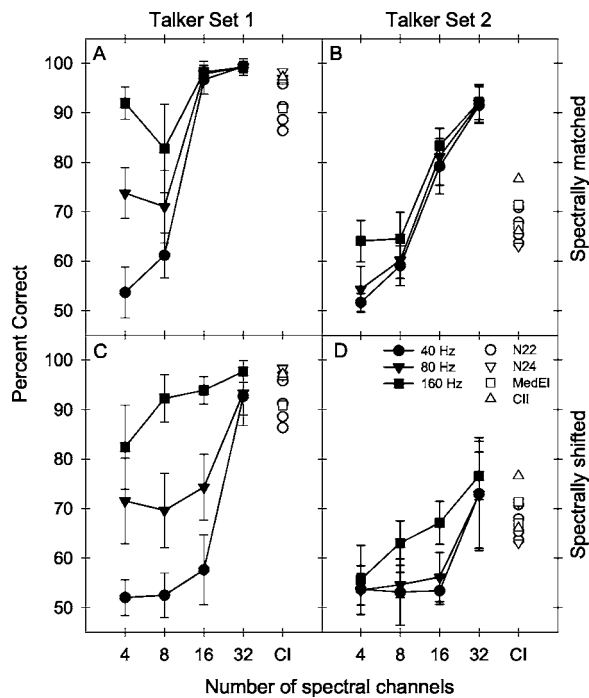
FIG. 2. Voice gender identification scores for CI and NH listeners. The left panels [(a) and (c)] show results for Talker Set 1 and the right panels [(b) and (d)] show results for Talker Set 2. In all panels, individual CI results (open symbols) are shown for performance with clinically assigned speech processors. Mean NH listener data are shown for spectrally matched speech in the top panels [(a) and (b)] and spectrally shifted speech in the bottom panels [(c) and (d)], as a function of the number of channels; the closed symbols show performance for the different envelope cut-off frequency conditions. Error bars indicate ±1 standard deviation. N22, N24, MedEL, and CII refer to data from CI patients using the Nucleus-22, Nucleus-24M, MedEL Combo+, and Clarion CII implant devices, respectively.

vealed that the envelope filter cutoff frequency significantly affected performance when there were 8 or fewer channels ($p < 0.001$); with 16 or more channels, the envelope cutoff frequency had no effect on performance, as mean performance was better than 98% correct for all envelope filter conditions. With 8 or fewer channels, there were inconsistent interactions between the spectral resolution and the envelope filter. For example, with the 40 Hz envelope filter, performance significantly improved as the number of channels was increased from 4 to 8 ($p = 0.002$), while with the 160 Hz filter, performance significantly declined as the number of channels was increased from 4 to 8 ($p < 0.001$); for the 80 Hz envelope filter, increasing the number of channels from 4 to 8 had no significant effect on performance ($p = 1.00$). Because the envelope filters used in CI users' speech processors were generally comparable to those used in the acoustic processors with the 160 Hz envelope filter, the CI data were compared to the NH data with the 160 Hz envelope filter. A one-way ANOVA test revealed that there was no significant difference in performance between CI users and NH subjects listening to the 4-channel processor ($p = 0.280$). However, CI users performed significantly poorer than NH subjects listening to either the 16-channel processor ($p = 0.011$) or 32-channel processor ($p = 0.003$). Interestingly, CI users performed significantly better than NH subjects listening to the 8-channel processor.

Similar to CI users, NH listeners' performance was generally poorer with Talker Set 2 (right column) than with Talker Set 1 (left column). With spectrally matched speech (panel B), for all envelope filter conditions, performance generally improved as the number of spectral channels was increased, reaching a peak level of ~92% correct with 32 spectral channels. One exception to this trend was with the 160 Hz envelope filter condition, in which performance was unchanged (~64% correct) as the number of spectral channels was increased from 4 to 8. A two-way ANOVA revealed main effects for spectral resolution [$F(3, 108) = 454.75$, $p < 0.001$] and envelope cut-off frequency [$F(2, 108) = 18.95$, $p < 0.001$]. As with Talker Set 1, there was a significant interaction between the spectral resolution and the amount of temporal information [$F(6, 108) = 3.88$, $p = 0.002$]. Post-hoc Bonferroni t-tests revealed that the 160 Hz envelope filter cutoff frequency significantly affected performance when there were 8 or fewer channels. With 4 channels, there was no significant difference between the 40 and 80 Hz filter ($p = 0.484$); increasing the cutoff frequency from 80 to 160 Hz significantly improved performance ($p < 0.001$). With 8 channels, there was no significant difference between the 40 and 80 Hz filter ($p = 1.00$), or between the 80 and 160 Hz filter ($p = 0.074$); however, increasing the cutoff frequency from 40 to 160 Hz significantly improved performance ($p = 0.015$). With 16 or more channels, the envelope cutoff frequency had no effect on performance. For all envelope filter conditions, each doubling of the spectral resolution significantly improved performance ($p < 0.001$), except for the 160 Hz envelope filter, in which increasing the number of channels from 4 to 8 did not significantly affect performance ($p = 1.000$). CI users' performance was also compared to that of NH subjects with the 160 Hz envelope filter. A one-way ANOVA test revealed that there was no significant difference between CI users' performance and that of NH subjects listening to the 4-channel processor ($p = 0.053$) or the 8-channel processor ($p = 0.132$). However, CI users' performance was significantly worse than that of NH subjects listening to the 16-channel or the 32-channel processor ($p < 0.001$).

For Talker Set 1 with spectrally shifted speech [panel (c)], performance with the 40 and 80 Hz envelope filters increased slowly as the number of spectral channels was increased from 4 to 16. Performance improved sharply with the 160 Hz envelope filter, for all spectral resolution conditions except 32 channels. A two-way ANOVA revealed main effects for spectral resolution [$F(3, 108) = 117.07$, $p < 0.001$] and envelope cut-off frequency [$F(2, 108) = 221.59$, $p < 0.001$]. There was a significant interaction between the spectral resolution and the amount of temporal information [$F(6, 108) = 18.58$, $p < 0.001$]. Post-hoc Bonferroni t-tests revealed that increasing the envelope filter cutoff frequency significantly improved performance when there were 16 or fewer channels ($p < 0.001$); with 32 channels, the envelope cutoff frequency had no effect on performance. There was a different pattern of interaction between the envelope filter and spectral resolution for spectrally shifted speech, as compared to those with spectrally matched speech [panel (a)]. For the 40 and 80 Hz envelope filter conditions, performance did not significantly improve until 32 spectral channels were

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Fu *et al.*: Voice gender identification by cochlear implant users   1715

available ($p<0.001$); with the 160 Hz envelope filter, performance significantly improved as the spectral resolution was increased from 4 to 8 channels ($p=0.002$); increasing the spectral resolution beyond 8 channels did not significantly affect performance.

For Talker Set 2 with spectrally shifted speech [panel (d)], performance with the 40 and 80 Hz envelope filters did not improve until 32 spectral channels were provided. Performance with the 160 Hz envelope filter gradually improved as more spectral channels were added. However, peak performance was less than 80% correct, even with 32 spectral channels, significantly worse than the other Talker Set/spectral shift conditions. A two-way ANOVA revealed main effects for spectral resolution [$F(3,108)=60.16$, $p<0.001$] and envelope cut-off frequency [$F(2,108)=15.89$, $p<0.001$]. Different from the other test conditions, there was no significant interaction between spectral resolution and the amount of temporal information [$F(6,108)=2.06$, $p=0.064$]. *Post-hoc* Bonferroni *t*-tests showed that the envelope filter cutoff frequency did not significantly affect performance with 4 or 32 spectral channels. With 8 or 16 spectral channels, the 160 Hz envelope filter provided significantly better performance than that with the 40 Hz or 80 Hz envelope filter ($p<0.01$). For the 40 and 80 Hz envelope filter conditions, performance did not significantly improve until 32 spectral channels were provided ($p<0.001$). For the 160 Hz envelope filter, increasing the spectral resolution 4 to 16, or from 16 to 32 channels significantly improved performance ($p=0.006$).

IV. DISCUSSION

The results from the present study show that spectral cues, periodicity cues, and spectral mismatch significantly contributed to NH subjects' voice gender identification performance when listening to an acoustic CI simulation. These effects were significantly different for the two Talker Sets. With Talker Set 1 (large difference between the F0's of male and female talkers), NH listeners attended to both periodicity and spectral cues. With Talker Set 2 (small difference between the F0's of male and female talkers), NH listeners primarily attended to spectral cues. The effect of Talker Set was similar for NH and CI listeners, with voice gender identification performance generally much better for Talker Set 1. While overall recognition scores were fairly consistent across subjects, it is interesting to note that subjects' response patterns (or confusion matrices) significantly differed in various speech processing conditions. These results suggest that the response patterns may not be simply predicted by the individual talkers' acoustic features. The effects of the speech processor parameters used in the acoustic CI simulation and their relation to CI patients' performance will be described in more detail below.

For both Talker Sets, when the periodicity cues were removed (40 Hz envelope filter), improving the spectral resolution provided better voice gender identification, similar to results reported for vowel identification (Friesen *et al.*, 2001). For vowel recognition, improving the spectral resolution may have allowed listeners to perceive important for-

mant frequencies and formant transitions more accurately. Similarly, for voice gender identification, improving the spectral resolution may have allowed listeners to make use of spectral envelope differences between male and female talkers, as the formant frequencies of female talkers are roughly 16% and 11% higher than those of male talkers for Talker Sets 1 and 2, respectively. The performance improvement with increasing spectral resolution suggests that NH listeners may have performed some sort of spectral profile analysis when the F0 information was ambiguous between talkers (as in Talker Set 2), especially at higher levels of spectral resolution. The difference in performance between the two Talker Sets further underscores the importance of the spectral profile to voice gender identification. When the F0's of talkers did not overlap (Talker Set 1), 16 or more spectrally matched channels provided near-perfect performance, regardless of temporal cues; with spectrally shifted speech, 32 or more spectral channels were needed if periodicity cues were not provided. When the F0's of talkers significantly overlapped (Talker Set 2), increasing the spectral resolution improved performance across all envelope filter conditions when there was no spectral mismatch. When there was a spectral mismatch, NH subjects were able to take advantage of the improved spectral resolution only with the 160 Hz envelope filter; with the 40 and 80 Hz envelope filters, performance remained near chance-level until 32 spectral channels were provided.

Given the importance of the spectral profile to voice gender identification, distortions to the spectral profile because of spectral mismatch may affect performance. In the present study, the effect of spectral mismatch interacted with spectral resolution and Talker Set. With Talker Set 1, when 8–16 spectral channels were provided, the one-octave spectral shift caused performance to significantly decline; with 4 or 32 spectral channels, performance was less affected by the spectral shift. With Talker Set 2, the spectral shift caused performance to decline more severely. When 16 spectral channels were available, the spectral mismatch caused mean NH performance to drop nearly 20 percentage points. With 32 spectral channels, mean NH performance dropped nearly 15 percentage points. Thus, when the difference among talkers' F0 values was relatively small (Talker Set 2), and listeners must rely more strongly on the spectral profile to discriminate voice gender, distortions to the spectral profile because of spectral shifting may further reduce performance relative to that with spectrally-matched speech. When speech is spectrally shifted, a minimum of 32 spectral channels is necessary for voice gender identification among talkers whose F0 values overlap. While NH listeners were tested with the experimental speech processors without any prior exposure to the processing, it is possible that subjects might eventually accommodate the spectral shift, as suggested by previous studies (e.g., Fu *et al.*, 2002; Rosen *et al.*, 1999). Also, because vowel tokens were used to test voice gender identification, rather than sentence materials, subjects were provided with minimal speech samples with which to analyze the spectral profile; sentence materials might have al-

lowed listeners to average the formant information across a larger sample. In such a case, spectral shifting might have affected performance less severely.

Along with the spectral resolution, the distribution of frequency analysis and carrier bands will significantly affect the spectral profile. With 4–8 spectral bands, the frequency allocation may be very important to preserve the spectral cues necessary for voice gender identification. With 32 spectral channels, the relatively high degree of spectral resolution may lessen the effect of frequency allocation. With Talker Set 1, performance for NH subjects listening to spectrally matched speech with the 160 Hz envelope filter dropped significantly as the spectral resolution was increased from 4 to 8 channels; a similar trend was noted for the 80 Hz envelope filter. The data are not consistent with our previous study (Fu *et al.*, 2004), in which performance improved as the spectral resolution was increased from 4 to 8 channels. This difference in performance may have been due to the different frequency allocations used in the previous and present study. In the previous study, the overall input frequency range was between 200 and 7000 Hz, distributed according to Greenwood's (1990) function. In the present study, the overall input frequency range was between 100 and 5000 Hz, also distributed according to Greenwood's function; this range was used to accommodate the one-octave spectral shift test condition. Thus, although the frequency bands were distributed according to Greenwood's function in both studies, the corner frequencies of the analysis filter bands significantly differed. It is possible that the frequency allocation used in the present study introduced an ambiguity for certain talkers for the 8-channel processor.

While previous work has shown that periodicity cues can significantly affect voice gender identification performance (Fu *et al.*, 2004), the degree of difference between talkers' F0 values may determine how strongly temporal cues may contribute to performance. In the present study, temporal cues contributed strongly to NH listeners' voice gender identification under conditions of reduced spectral resolution when there was no overlap between male and female talkers' F0 values (Talker Set 1). When there was significant overlap between male and female talkers' F0 values (Talker Set 2), temporal cues contributed much less to voice gender identification. Also, periodicity cues contributed less to voice gender identification when 16 or more spectral channels were provided. With spectrally shifted speech, periodicity cues contributed more strongly to performance than with spectrally matched speech. Taken together, these results suggest that when the spectral profile is sufficiently weakened or distorted, listeners will rely more strongly on temporal cues to identify voice gender.

For NH subjects listening to an acoustic CI simulation, voice gender recognition was significantly affected by the spectral resolution, especially for talkers whose F0 values significantly overlap. For CI subjects, the amount of spectral resolution differed among implant devices (22 intracochlear electrodes for the Nucleus-22 and -24 devices, 12 for the MedEl device and 16 for the Clarion II device). In general, there was no correlation between performance and the number of channels/electrodes used by the various implant systems. Compared to NH listeners, CI users' perceived spectral pattern representation may be further weakened and/or distorted by the different speech processing strategies used in the CI systems. Patient-related factors (e.g., electrode locations, proximity of electrodes to healthy neural regions) may further distort the spectral patterns perceived by CI users, relative to NH listeners. For example, a shallow insertion depth would compress the acoustic input frequency range onto relatively fewer electrodes and poor electrode selectivity might "smear" the spectrum of the input speech signal. For both Talker Sets, CI performance was comparable to that of NH subjects listening with 4–8 spectral channels with the 160 Hz envelope filter. Because of the relatively weaker spectral profile, CI patients may rely more strongly on temporal cues to identify voice gender.

V. CONCLUSION

NH listeners' performance with multichannel CI simulations showed that both temporal cues and spectral cues significantly affected voice gender identification, and that there were significant interactions among speech processor parameters such as spectral resolution, the envelope filter cutoff frequency and the degree of spectral shift. When male and female talkers' F0 values were widely separated (Talker Set 1), NH listeners were able to attend to periodicity cues to discriminate voice gender. When talkers' F0 values overlapped (Talker Set 2), NH listeners relied primarily on spectral cues. For both Talker Sets, NH listeners required 16 spectral channels to effectively use the spectral profile to discriminate voice gender; when the speech spectrum was shifted, 32 spectral channels were required. CI users' performance was comparable to that of NH subjects listening to between 4 and 8 channels with temporal cues up to 160 Hz. Because of the reduced spectral resolution, temporal cues were sufficient to discriminate voice gender when male and female talkers' F0's were widely separated. When talkers' F0's overlapped, CI listeners could not effectively attend to the spectral profile because of limited spectral resolution. Because increased temporal cues did not provide better performance with NH listeners for Talker Set 2, the results suggest that more spectral channels, rather than enhanced periodicity cues, are needed to improve CI patients' voice gender identification performance.

Arndt, P., Staller, S., Arcaroli, J., Hines, A., and Ebinger, K. (**1999**). "Within-subject comparison of advanced coding strategies in the Nucleus 24 Cochlear Implant." Cochlear Corporation, Englewood, CO.
Bachorowski, J. A., and Owren, M. (**1999**). "Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech," J. Acoust. Soc. Am. **106**, 1054–1063.
Cleary, M., and Pisoni, D. B. (**2002**). "Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results," Ann. Otol. Rhinol. Laryngol. Suppl. **189**, 113–118.

Fant, G. (**1966**). "A note on vocal tract size factors and non-uniform F-pattern scaling," Speech Trans. Lab. Quart. Prog. And Status Report **4**, 22–30.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: A comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Fu, Q.-J., Chinchilla, S., and Galvin, J. J. (**2004**). "Voice gender discrimination and vowel recognition in normal-hearing and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.

Fu, Q.-J., Shannon, R. V., and Galvin, J. J. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

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

Hazan, V., and Markham, D. (**2004**). "Acoustic-phonetic correlates of talker intelligibility for adults and children," J. Acoust. Soc. Am. **116**, 3108–3118.

Hillenbrand J., Getty, L. A., Clark, M. J., and Wheeler K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Lieberman, P., and Blumstein, S. E. (**1993**). *Speech Physiology, Speech Perception, and Acoustic Phonetics Cambridge* (University Press, Cambridge, MA).

Markel, J. D. (**1972**). "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio Electroacoust., **AU-20**, 367–377.

Pisoni, D. B. (**1997**). "Some thoughts on 'normalization' in speech perception," in *Talker Variability in Speech Processing*, edited by K. Johnson and J. Mullenix (Academic, New York), pp. 9–32.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.

Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.

Scherer, K. R. (**1986**). "Vocal affect expression: A review and model for future research," Psychol. Bull. **99**, 143–165.

Seligman, P. M., and McDermott, H. J. (**1995**). "Architecture of the Spectra-22 speech processor," Ann. Otol. Rhinol. Laryngol. **104**, 139–141.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (**2002**). "Learning to recognize talkers from natural, sinewave, and reversed speech samples," J. Exp. Psychol. Hum. Percept. Perform. **28**, 1447–1469.

Spahr, A. J., and Dorman, M. F. (**2004**). "Performance of subjects fit with the advanced Bionics CII and Nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624–628.

Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–707.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "New levels of speech recognition with cochlear implants," Nature (London) **352**, 236–238.

The effects of high presentation levels on consonant feature transmission[a)]

Benjamin W. Y. Hornsby[b)]
Dan Maddox Hearing Aid Research Laboratory, Vanderbilt Bill Wilkerson Center, 1215 21st Avenue South, Room 8310, Medical Center East, South Tower, Nashville, Tennessee 37232

Timothy D. Trine
Starkey Laboratories, Inc., 6600 Washington Avenue South, Eden Prairie, Minnesota 55344

Ralph N. Ohde
Department of Hearing and Speech Sciences, Vanderbilt Bill Wilkerson Center, 1215 21st, Avenue South, Room 8310, Medical Center East, South Tower, Nashville, Tennessee 37232

The effect of high speech presentation levels on consonant recognition and feature transmission was assessed in eight participants with normal hearing. Consonant recognition in noise (0 dB signal-to-noise ratio) was measured at five overall speech levels ranging from 65 to 100 dB SPL. Consistent with the work of others, overall percent correct performance decreased as the presentation level of speech increased [e.g., G. A. Studebaker, R. L. Sherbecoe, D. M. McDaniel, and C. A. Gwaltney, J. Acoust. Soc. Am. **105**(4), 2431–2444 (1999)]. Confusion matrices were analyzed in terms of relative percent information transmitted at each speech presentation level, as a function of feature. Six feature sets (voicing, place, nasality, duration, frication, and sonorance) were analyzed. Results showed the feature duration (long consonant duration fricatives) to be most affected by increases in level, while the voicing feature was relatively unaffected by increases in level. In addition, alveolar consonants were substantially affected by level, while palatal consonants were not. While the underlying mechanisms responsible for decreases in performance with level increases are unclear, an analysis of common error patterns at high levels suggests that saturation of the neural response and/or a loss of neural synchrony may play a role. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1993128]

I. INTRODUCTION

Substantial research has shown that presenting speech at higher-than-normal levels [i.e., above 80 to 90 dB sound pressure level (SPL)] results in decreased speech understanding (French and Steinberg 1947; Fletcher and Galt 1950; Pollack and Pickett 1958; Studebaker *et al.* 1999; Molis and Summers 2003). This decrease is also observed among persons with hearing loss (Jerger and Jerger 1971; Studebaker *et al.* 1999). The absolute presentation level at which speech recognition begins to break down is affected by several factors. These factors include the presence or absence of a background noise, the spectral shape of the background noise, the signal-to-noise ratio (SNR), and frequency region being presented (French and Steinberg 1947; Pollack and Pickett 1958; Loven and Collins 1988; Studebaker *et al.* 1999; Molis and Summers 2003).

French and Steinberg (1947) reported decreases in intelligibility, in quiet, when speech levels increased above about 80 dB SPL. The results of their work are incorporated into the ANSI S3.5-1997 (ANSI 1997) Calculation of Speech Intelligibility Index (SII) and are used to model the effects of high presentation levels on speech recognition. Pollack and Pickett (1958) also reported slight decreases in performance when speech was presented in quiet at levels between 80 to 130 dB SPL. They noted substantially larger decrements over the same range of levels when the speech was mixed with a broadband white noise.

The spectral shape of the masking noise also affects speech recognition at high levels. Decrements in speech recognition are observed at much lower levels when the speech signal is mixed with a speech-shaped background noise (Goshorn and Studebaker *et al.* 1994; Studebaker, *et al.* 1999) or a single-talker competing message (Speaks, *et al.* 1967) as opposed to a broadband white noise (Pollack and Pickett 1958). Studebaker *et al.* (1999) presented bandpass filtered NU-6 monosyllables at levels ranging from 64 to 99 dB SPL at multiple fixed SNRs. The background noise was filtered to match the average spectral shape of the speech stimuli. Studebaker *et al.* (1999) reported decreases in speech intelligibility at presentation levels above 69 dB SPL. This level is substantially lower than that reported by earlier researchers (i.e., Fletcher and Galt 1950; Pollack and Pickett 1958) who examined the effect of high presentation levels in quiet and in a broadband white noise.

Finally, recent work by Molis and Summers (2003) suggests that high-frequency speech information may be more susceptible to the negative effects of high presentation levels than lower-frequency information. Using filtered sentence materials presented in quiet, they observed significantly greater "rollover" with increases in level for high-pass filtered (>1938 Hz) than for low-pass filtered (<1366 Hz) sentence materials. Given, however, that these materials were presented in isolated bands, it is not clear that these results will generalize to broadband speech conditions. Goshorn and Studebaker (1994) found that when low-frequency (200–400 Hz) and high-frequency bands (1000–5000 Hz) of speech were presented together, increasing the level of a low-frequency band had a larger negative impact on speech understanding than an equivalent increase in the high-frequency band, a finding they attribute to the upward spread of masking.

In addition to a general decrease in speech recognition, researchers have also examined the effects of high speech levels on the perception of isolated speech features using synthetic speech stimuli. Several studies have demonstrated that high presentation levels influence the perception of place of articulation (Dorman and Dougherty 1981; Van Tasell and Crump 1981; Hedrick and Jesteadt 1997). Dorman and Dougherty (1981) examined shifts in phonetic identification due to changes in the presentation level of synthetic stimuli. These authors generated identification functions using a 12-step continuum of two-formant synthetic consonant-vowel (CV) stimuli ranging from /ba/ to /da/ to /ga/. Participants identified stimuli as /ba/, /da/ or /ga/ at three presentation levels of 55, 70, and 90 dB SPL. The identification functions they obtained showed a substantial change in shape as the overall level of the speech stimuli increased. A large increase in velar /ga/ responses and a concomitant decrease in the labial /ba/ and alveolar /da/ responses occurred as the level increased from 55 to 90 dB SPL. Two formant stimuli, however, do not represent /g/ well (Ohde 1988), thus it is not entirely clear why an increase in velar /ga/ responses was observed. Hedrick and Jesteadt (1997) reported similar results when evaluating the effect of overall and relative levels on the labeling functions of the voiceless /p/-/t/ contrast. Their experiments also used synthetic CV stimuli, with overall levels ranging from 55 to 100 dB SPL. They reported an increase in alveolar /t/ and decrease in labial /p/ responses as level increased.

In contrast to work using synthetic stimuli, there has been very little research examining the effects of high speech levels on the consonant-feature transmission of natural speech stimuli. In other words, very little work has examined how our perception of natural speech features changes at high presentation levels. Much of the existing work has examined changes in feature transmission over a range of relatively low presentation levels. These studies have typically begun testing near threshold and have increased the presentation level up to normal conversational levels or until maximum performance was reached (Kent 1978; Gutnick 1982). One exception to this is a study by Wang and Bilger (1973) who used Sequential Information Analysis (SINFA) in an attempt to identify salient features that "consistently account for transmitted information" across multiple presentation levels, SNRs, and stimulus sets. Their participants identified CV and vowel-consonant syllables in the presence of a broadband white noise at levels ranging from 50 to 95 dB SPL. They reported that the percent of correct identification decreased only slightly at the highest noise levels. This is consistent with the findings of other studies that have used a white noise masker (Pollack and Pickett 1958). The overall level of the noise did not significantly change the pattern of salient features identified by the SINFA.

To our knowledge, no study has examined the effects of high-intensity speech on consonant feature transmission using a speech-shaped noise as a masker. It is not clear if the findings of Wang and Bilger (1973) would hold true when listening in a speech-shaped background noise. Studebaker and colleagues (1999) compared results across multiple studies and found that the negative effects of high-presentation levels were larger when the background noise was speech-shaped rather than a white noise of comparable level. The reason for this difference in effect size between white noise and speech-shaped maskers is not clear. It is possible that speech-shaped maskers, which may contain relatively sharp spectral peaks compared to a white noise, have a greater spread of masking as speech levels increase compared to a white noise of equal level.

The effects of high presentation levels have direct implications for persons with hearing loss. Persons with hearing loss frequently listen to amplified speech at high levels to ensure audibility. The research reviewed above suggests that under certain conditions, increasing the level of speech may actually reduce speech recognition performance. Basic knowledge about speech perception at high levels is necessary to explore methods to improve speech understanding in these adverse listening situations. It is possible that predictable patterns of errors may occur with changes in presentation level. Some specific features of speech may be more susceptible to degradation than others. For example, the onset frequency of the second formant (F2 locus) and spectral information in the burst and transition trajectory of a stop consonant are important acoustic cues to perception of place of articulation for stop consonants (Blumstein and Stevens 1980; Ohde et al. 1995). Poor identification of the place of articulation among stop consonants at high levels would suggest that high presentation levels somehow distort these relatively short duration cues. Information about specific error patterns and feature degradation at high levels may provide insight into the underlying mechanisms responsible for errors that occur at high levels and, subsequently, may suggest strategies for reducing them.

The purpose of this study was to determine which features of speech were most susceptible to degradation at high intensities. A second purpose was to identify common error patterns that occur at high speech presentation levels in order to reveal potential mechanisms responsible for this degradation.

Feature	/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Voicing	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
Friction	0	0	0	1	1	1	1	0	0	0	1	1	1	1	0	0	1	1	0	0	0	0
Duration	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
Nasality	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Sonorance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1
Place	1	3	5	1	2	3	4	1	3	5	1	2	3	4	1	3	4	4	5	4	4	3

Note: The 1's and 0's in the first five rows refer to the presence (1) or absence (0) of a feature for a particular consonant. The coding for the place feature is as follows: 1=Labial, 2=Dental, 3=Alveolar, 4=Palatal, and 5=Back.

II. METHODS

A. Participants

Eight participants (two male and six female) between the ages of 25 and 42 participated in this experiment. All participants had pure-tone air conduction thresholds of 15 dB hearing level (HL) or better at octave frequencies from 250–8000 Hz [ANSI S3.6 1996 (ANSI 1996)] and word recognition scores of 92% or better (NU-6, 25 item list) with speech presented at a conversational level (50 dB HL). Participants had normal middle ear function defined as normal pressure and compliance functions (Type A tympanograms) and the presence of an ipsilateral acoustic reflex at 500 and 2000 Hz in response to a 90 dB SPL pure tone stimulus.

B. Stimuli

Nonsense syllables were used to limit linguistic contribution to speech recognition and focus on the transmission of acoustic information at high intensities. The speech stimuli were 22 CV syllables from a portion of the UCLA recording of the Nonsense Syllable Test (Dubno and Schaefer 1992, 1995). CV syllables consisted of 22 consonants (/p/, /t/, /k/, /f/, /s/, /b/, /d/, /g/, /v/, /z/, /m/, /n/, /w/, /r/, /j/, /l/, /θ/, /ð/, /dʒ/, /tʃ/, /ʒ/, /ʃ/) paired with the vowel /i/. Due to the large number of stimulus presentations required for consonant confusion analysis, the single vowel (/i/) was used to limit the testing time required for each participant. In addition, overall performance on consonant recognition testing is lower for the vowel /i/ compared to the vowel /a/ (Gutnick 1982). The use of /i/, then, also helped to reduce ceiling effects. Stimuli were natural tokens spoken by one male and one female speaker resulting in a total of 44 individual tokens used during the experiment. The masking noise was a 10 s sample of digitally filtered noise spectrally shaped to match the average long-term spectra of speech reported by Byrne et al. (1994). This shape is an average based on measures made across multiple languages using both male and female speakers. Each speech stimulus was temporally centered within a 1200 ms burst of noise with a 25 ms rise and fall time. The starting point for each 1200 ms sample of noise was randomly chosen from within the 10 s sample. The average stimulus duration was 636 ms. The longest and shortest stimuli had durations of 800 and 479 ms, respectively. Noise was present for at least 200 ms prior to the onset and following the offset of each stimulus.

1. Stimulus feature sets

The features used for analysis in this study followed those suggested by Kent (1978). These authors used five binary features [voicing, frication, duration, and nasality from Miller and Nicely (1955), and sonorance], and a five-value feature for place of articulation (labial, dental, alveolar, palatal, and back). Table I shows a breakdown of consonants by acoustic-phonetic features.

2. Calibration

Speech levels were determined by concatenating all 44 CV stimuli into a single wave form with the silent portions between stimuli removed. This concatenated stimulus was then output through the experimental circuit (without the noise) to an ER3-A earphone connected to a 2-cc coupler and sound level meter (Larson Davis 800B). The output of the sound level meter was then routed to a spectrum analyzer (SRS 785) and averaged to determine an overall root-mean-square (rms) level in dB SPL using a 0 to 8000 Hz bandwidth. A gain control on a mixer (TDT SM3) was used to adjust this average level of the concatenated speech to 100 dB SPL. A pure-tone signal was digitally generated to match the rms level of the concatenated speech and was used for calibration before each session. The same procedure was used to determine calibration values for the 10 s sample of digitally filtered speech-shaped noise. A second gain control on the TDT SM3 was used to adjust the overall level of the masking noise to equal the overall rms level of the concatenated speech stimulus. Figure 1 shows the one-third octave band spectra (100–10 000 Hz) for the concatenated speech stimulus and for the 10 s sample of speech-shaped noise measured in a 2 cc coupler.

As can be seen in Fig. 1, the spectral shape of the speech and the background noise differ substantially. The use of the single vowel /i/ in a CV context, with its relatively high second formant frequency (F2), resulted in quite a peaked shape. There is a substantial midfrequency dip between F1 (about 250 Hz) and F2 (about 3 kHz) for the concatenated C/i/ stimuli. In contrast, the background noise, with its low-frequency peak and gradually falling high-frequency spectrum, mirrors average conversational speech (Byrne et al. 1994). Studebaker et al. (1999) advocate using a background noise that is spectrally matched to the average speech stimuli when testing speech understanding in noise. Given the unusual shape of the concatenated CV stimuli used in this experiment, however, we would expect substantially different
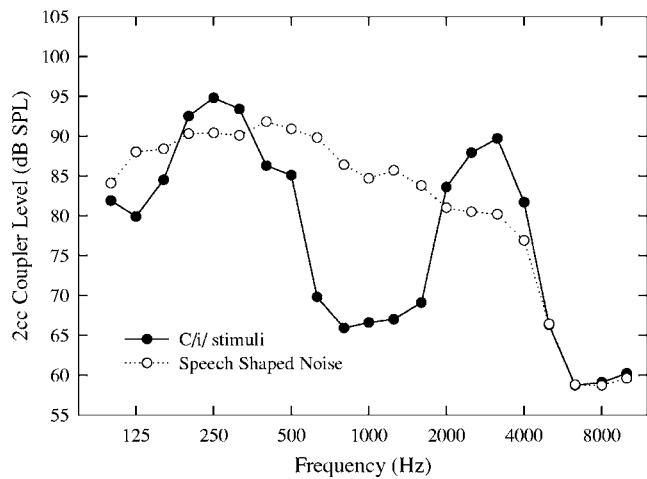
FIG. 1. One-third octave rms levels (100–10 000 Hz) of the nonsense syllable speech stimuli (male and female talker presenting consonants paired with the vowel /i/) and the speech-shaped background noise used in this study. The overall levels of the speech and noise, output through ER3 earphones and measured in a 2 cc coupler, are 100 dB SPL.

FIG. 2. Consonant recognition scores (in RAUs), averaged across participants, as a function of speech presentation level (in dB SPL in a 2 cc coupler). Speech was presented in noise at a 0 dB SNR. Filled symbols show performance in noise as a function of level. The open symbol shows performance in quiet when the speech was presented at a level of 65 dB SPL. Error bars show performance 1 standard deviation from the mean.

amounts of masking across speech tokens (i.e., a token with energy focused within the "valley" compared to one with energy focused in the "peak" regions). As a compromise, we chose a masker that maintained a "speech shape" without providing substantial spectral dips.

III. PROCEDURE

Participants typically listened to six runs per session (range of four to nine runs). A run consisted of 44 (22 consonants × 2 speakers) CV stimuli randomly presented three times at a single level (132 stimuli per run at each level). All participants completed four runs at each presentation level. This resulted in 528 syllables (132 stimuli × 4 runs) for each presentation level and 4224 stimuli per condition (8 participants × 528 syllables). Presentation levels were randomized to reduce order effects. All tokens were stored digitally and output at 33 kHz through a 16-bit digital/analog converter. The signal was then lowpass filtered (8 kHz), attenuated, and mixed with a masking noise. The combined speech and noise signal was presented monaurally (right ear) via ER-3A insert earphones at a SNR of 0 dB. Participants were tested at five overall levels: 65, 75, 85, 95, and 100 dB SPL in noise and also at 65 dB SPL in quiet.

Testing in quiet provided a baseline check of performance for all the syllables presented. Testing sessions were limited to approximately one hour to reduce fatigue. Within-session breaks were provided as needed. The task required participants to listen to stimuli while viewing their response alternatives on a computer screen. They indicated their response by clicking on the appropriate syllable with a computer mouse. All of the test consonants were presented on the screen during each trial and no correct answer feedback was given. Data from each trial were stored in a 22 × 22 confusion matrix.

IV. RESULTS

A. Effect of speech presentation level on overall performance

A repeated-measures analysis of variance (ANOVA) was used to analyze the main effect of presentation level on speech understanding. An alpha level of 0.05 was used for this, and all other, statistical analyses reported in this paper. Prior to statistical analyses, percent correct scores were transformed to rationalized arcsine units (RAUs) to stabilize the error variance (Studebaker 1985). When using the RAU transform, the percent correct scores are converted into arcsine units that are scaled to resemble the original percent of correct values over the range of 20 to 80 RAU. In addition, given the number of stimuli presented to obtain a score in each condition, scores of approximately −16 and 116 RAU correspond to percent correct scores of 0 and 100%, respectively. Figure 2 displays mean consonant recognition scores (in RAUs) as a function of the presentation level.

ANOVA testing revealed a significant main effect of presentation level ($F_{2,48} = 124.1$ $p < 0.05$). As illustrated in Fig. 2, consonant recognition decreased as the overall presentation level increased. This is consistent with the findings of previous researchers Studebaker et al. 1999. (Pollack and Pickett 1958); Posthoc testing, using a Tukey HSD test with a significance level of $p < 0.05$, showed a significant decrease in performance as soon as the presentation level increased above the lowest level tested (65 dB SPL). Average performance decreased systematically as the presentation level increased from 65 to 100 dB SPL. Significant decreases occurred with each increase in level up to 95 dB. The difference between performance at the 95 and 100 dB levels was not significant. An average decrease of approximately 22 RAU was observed as speech levels increased from 65 to 100 dB SPL.

Hornsby *et al.*: Consonant feature transmission at high levels

TABLE II. Confusion matrix for C/i/ stimuli (65 dB presentation level, 0 dB SNR).

											Stimulus												
		/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Response	/p/	122	0	29	3	6	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0
	/t/	6	192	9	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/k/	45	0	154	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	/f/	7	0	0	101	35	0	0	16	0	0	7	1	0	0	0	0	0	0	0	0	0	0
	/θ/	4	0	0	29	116	0	0	17	0	0	21	5	0	0	0	1	0	0	0	0	0	4
	/s/	0	0	0	1	20	183	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	/ʃ/	0	0	0	0	0	0	175	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/b/	0	0	0	28	3	0	0	66	0	0	18	6	1	0	0	0	0	0	0	1	0	0
	/d/	0	0	0	0	1	0	0	4	178	2	7	7	0	0	0	1	0	0	0	1	0	2
	/g/	5	0	0	1	0	0	0	0	3	140	0	0	0	0	0	1	0	3	0	0	16	0
	/v/	2	0	0	22	2	0	0	56	6	0	61	29	0	0	2	1	0	0	4	0	0	0
	/ð/	0	0	0	3	6	2	0	27	1	0	47	88	5	0	2	3	0	0	0	0	0	22
	/z/	0	0	0	0	1	7	0	5	0	0	14	28	177	0	0	0	0	0	0	0	0	3
	/ʒ/	0	0	0	0	0	0	7	0	2	0	0	0	0	177	0	0	0	14	0	0	1	0
	/m/	0	0	0	2	0	0	0	0	0	0	0	0	0	0	162	30	0	0	0	0	0	1
	/n/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	143	0	0	0	0	1	2
	/tʃ/	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	191	0	0	0	0	0
	/dʒ/	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	1	175	0	0	0	0
	/w/	0	0	0	1	0	0	0	1	0	0	1	0	1	0	2	3	0	0	173	2	0	2
	/r/	0	0	0	1	0	0	0	0	0	0	3	5	2	0	0	0	0	0	13	188	0	0
	/j/	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	1	0	0	0	0	174	1
	/l/	1	0	0	0	0	0	0	0	2	0	11	23	6	0	7	8	0	0	2	0	0	155

B. Speech feature transmission as a function of presentation level

Individual participant confusion matrices were combined to create a single matrix for each presentation level (Tables II–VI).

These matrices were analyzed using the Feature Information Transfer (FIX)[1] program to determine the percent information transmitted as a function of feature for each speech presentation level. Information analysis (Miller and Nicely 1955; Wang and Bilger 1973) was used to quantify

TABLE III. Confusion matrix for C/i/ stimuli (75 dB presentation level, 0 dB SNR).

											Stimulus												
		/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Response	/p/	113	0	35	3	6	0	0	2	0	3	0	1	0	0	0	0	0	0	0	0	0	0
	/t/	5	192	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/k/	46	0	151	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	/f/	9	0	0	78	32	0	0	12	1	0	10	2	0	0	0	0	0	0	0	0	0	0
	/θ/	6	0	0	52	123	2	0	17	3	0	20	13	1	0	0	0	0	0	0	0	0	3
	/s/	0	0	0	1	10	181	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/ʃ/	0	0	0	0	0	0	171	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/b/	1	0	0	31	8	0	0	49	1	2	22	10	0	0	0	0	0	0	2	0	0	0
	/d/	1	0	0	3	7	0	0	14	175	1	13	10	0	0	0	3	0	0	0	0	1	1
	/g/	9	0	3	0	0	0	0	0	5	132	0	0	0	1	1	2	0	4	1	0	27	0
	/v/	1	0	0	6	1	0	0	52	1	0	59	39	0	0	4	4	0	0	9	1	0	11
	/ð/	0	0	0	7	4	0	0	30	3	0	49	76	6	0	4	1	0	0	1	0	0	25
	/z/	0	0	0	5	0	9	0	4	0	0	10	14	178	0	1	0	0	0	0	0	0	2
	/ʒ/	0	0	0	0	0	0	6	0	0	1	0	0	0	165	1	1	0	10	0	0	2	0
	/m/	0	0	0	0	0	0	0	3	0	0	0	1	0	0	133	26	0	0	1	0	0	1
	/n/	0	0	0	0	0	0	0	0	1	0	0	0	0	0	25	138	0	0	0	0	3	2
	/tʃ/	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	188	0	0	0	0	0
	/dʒ/	0	0	0	0	0	0	3	0	0	0	0	0	0	25	0	0	4	178	0	0	0	0
	/w/	0	0	0	3	0	0	0	1	0	0	0	1	0	0	4	1	0	0	166	4	2	0
	/r/	0	0	0	0	0	0	0	1	0	0	0	3	0	0	2	0	0	0	11	187	0	1
	/j/	1	0	0	0	0	0	0	1	0	53	0	1	0	1	1	2	0	0	0	0	156	0
	/l/	0	0	1	3	0	0	0	6	1	0	9	21	7	0	16	14	0	0	1	0	1	146

TABLE IV. Confusion matrix for C/i/ stimuli (85 dB presentation level, 0 dB SNR).

		Stimulus																					
		/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Response	/p/	80	0	21	5	16	0	0	2	0	2	1	0	0	0	0	1	0	0	0	0	0	0
	/t/	8	192	14	2	5	0	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0
	/k/	57	0	148	4	2	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	/f/	17	0	2	97	45	0	0	8	3	0	20	5	0	0	0	0	0	0	0	0	0	0
	/θ/	6	0	1	33	87	1	0	15	9	0	24	4	0	0	0	0	0	0	0	0	0	1
	/s/	0	0	0	2	8	174	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	/ʃ/	0	0	0	0	0	0	147	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
	/b/	1	0	0	20	6	0	0	44	4	1	22	10	0	0	6	2	0	0	1	0	1	2
	/d/	7	0	1	2	0	0	0	25	120	4	17	15	0	0	1	3	0	0	0	0	1	2
	/g/	10	0	4	0	0	0	0	14	27	119	3	0	0	0	0	9	0	1	0	1	16	0
	/v/	3	0	0	11	6	0	0	42	7	2	51	61	0	1	12	8	0	0	14	1	0	21
	/ð/	1	0	0	8	9	0	0	15	5	1	31	50	10	0	3	6	0	0	4	0	1	26
	/z/	0	0	0	1	1	17	0	4	3	0	7	22	174	0	0	4	0	0	3	0	0	7
	/ʒ/	0	0	0	0	0	0	6	1	2	6	1	1	0	163	0	3	0	12	0	0	4	0
	/m/	0	0	0	2	0	0	0	4	2	0	1	2	0	0	123	53	0	0	3	1	3	10
	/n/	0	0	0	0	0	0	0	1	1	0	1	0	0	0	32	77	0	0	2	0	2	2
	/tʃ/	0	0	1	0	0	0	32	0	0	0	0	0	0	0	0	0	188	0	0	0	0	0
	/dʒ/	0	0	0	0	0	0	7	1	4	7	1	1	0	22	0	0	3	179	0	0	4	0
	/w/	0	0	0	1	1	0	0	2	0	1	3	1	1	1	0	3	0	0	125	3	0	5
	/r/	0	0	0	0	1	0	0	1	0	0	1	9	1	0	0	2	0	0	31	184	2	11
	/j/	2	0	0	1	0	0	0	9	4	49	2	0	0	3	2	7	0	0	2	0	157	1
	/l/	0	0	0	3	5	0	0	2	0	0	2	11	6	2	11	14	0	0	7	2	1	104

the error patterns in the confusion matrices. This analysis method transformed the error patterns in each confusion matrix into percent information transmitted for each feature. This transformation provided an estimate of how susceptible a particular feature was to the effects of high intensities. For example, if the feature of voicing was resistant to the effects of high levels, this feature would be retained in error patterns in the confusion matrix and the percent voicing information transmitted would be fairly stable as the speech presentation level increased. If the voicing feature was negatively affected by increases in level, then error patterns would be more random (i.e., errors would consist of voiced/voiceless confu-

TABLE V. Confusion matrix for C/i/ stimuli (95 dB presentation level, 0 dB SNR).

		Stimulus																					
		/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Response	/p/	68	0	16	4	9	0	0	9	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	/t/	13	189	24	0	1	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
	/k/	55	0	135	1	1	0	0	4	0	1	1	0	0	0	0	0	0	0	0	0	0	0
	/f/	13	0	0	81	52	3	0	17	4	1	23	3	3	0	0	0	0	0	1	0	0	1
	/θ/	9	0	0	23	78	12	0	9	3	0	14	4	4	1	0	0	0	0	1	0	0	2
	/s/	0	0	0	2	9	144	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	/ʃ/	0	0	0	0	0	1	120	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
	/b/	4	0	0	49	10	0	0	53	56	0	18	7	0	0	6	0	0	0	1	1	0	5
	/d/	1	0	0	2	2	0	0	17	71	16	17	10	1	0	0	4	0	1	0	0	0	2
	/g/	11	0	7	0	0	0	0	6	19	108	3	2	0	2	0	6	0	0	0	0	11	0
	/v/	4	0	0	13	8	3	0	29	7	1	49	50	15	0	15	8	0	0	5	0	1	23
	/ð/	0	0	0	8	15	5	0	20	4	0	27	47	10	0	2	5	0	1	0	2	0	29
	/z/	0	0	0	4	1	21	0	4	3	0	3	15	123	0	0	1	0	0	1	0	2	3
	/ʒ/	1	0	0	0	1	1	2	0	6	8	1	1	2	147	2	4	1	19	1	0	8	0
	/m/	1	0	0	1	1	0	0	2	0	2	4	1	0	0	151	85	0	0	10	0	2	5
	/n/	1	0	0	1	1	0	0	2	0	0	0	0	0	0	9	42	0	0	4	0	2	1
	/tʃ/	5	2	8	2	0	0	66	0	0	0	0	0	0	0	0	0	189	1	0	0	0	0
	/dʒ/	1	1	2	0	2	0	4	4	5	12	6	6	1	39	2	1	0	168	1	0	4	0
	/w/	2	0	0	1	0	0	0	2	3	1	4	7	4	0	1	5	0	0	114	8	2	5
	/r/	2	0	0	0	0	0	0	2	1	1	9	30	26	0	0	2	0	1	50	179	1	17
	/j/	1	0	0	0	0	0	0	8	10	41	3	1	0	2	0	9	0	0	3	1	155	3
	/l/	0	0	0	0	1	1	0	4	0	0	5	8	3	1	4	20	0	0	0	1	4	96

		Stimulus																					
		/p/	/t/	/k/	/f/	/θ/	/s/	/ʃ/	/b/	/d/	/g/	/v/	/ð/	/z/	/ʒ/	/m/	/n/	/tʃ/	/dʒ/	/w/	/r/	/j/	/l/
Response	/p/	60	1	16	2	5	0	0	6	0	0	1	0	0	0	1	1	0	0	0	0	0	0
	/t/	20	178	29	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	/k/	43	1	120	1	1	0	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	/f/	25	0	2	83	48	3	0	14	4	0	12	0	3	0	0	0	0	0	0	0	0	1
	/θ/	13	0	2	38	94	33	0	8	6	0	24	6	4	0	1	0	0	0	0	0	0	3
	/s/	0	0	1	0	4	126	3	3	0	0	2	0	2	0	0	0	0	0	0	0	0	0
	/ʃ/	0	0	1	0	0	0	111	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
	/b/	4	0	0	32	9	0	0	48	46	0	18	12	1	0	4	2	0	0	1	2	0	7
	/d/	1	0	0	1	5	0	0	14	69	21	13	10	0	0	0	1	0	1	2	0	1	1
	/g/	15	0	4	0	0	0	0	10	9	112	2	0	0	0	1	3	0	2	1	0	22	0
	/v/	4	0	1	15	10	0	0	34	16	1	47	41	13	0	18	11	0	0	8	1	1	15
	/ð/	1	0	0	12	8	3	0	19	22	2	36	44	24	0	4	12	0	0	1	0	1	42
	/z/	0	0	0	1	3	22	0	9	5	0	15	27	92	0	3	6	0	0	1	1	3	6
	/ʒ/	1	0	0	1	0	0	5	4	3	8	1	0	0	133	3	1	0	9	3	0	9	0
	/m/	0	0	0	4	1	0	0	3	2	0	3	1	0	0	134	73	0	0	9	0	3	3
	/n/	0	0	0	0	1	0	0	0	0	0	0	0	0	1	11	42	0	0	2	0	2	1
	/tʃ/	3	12	8	0	0	0	68	0	0	0	0	0	0	0	0	0	191	0	0	0	0	0
	/dʒ/	2	0	7	0	0	0	5	2	4	12	1	0	1	48	3	4	0	177	1	0	6	0
	/w/	0	0	0	2	1	0	0	3	1	0	6	11	2	0	1	3	0	0	110	2	4	8
	/r/	0	0	0	0	0	5	0	5	0	0	6	27	37	0	1	5	0	0	48	181	0	10
	/j/	0	0	1	0	0	0	0	5	3	34	1	1	0	9	1	1	0	3	2	2	138	2
	/l/	0	0	0	0	0	0	0	1	2	2	4	12	12	0	6	27	0	0	3	3	2	93

sions not just within voicing errors) and the percent information transmitted would decrease as the speech level increased. Figure 3 shows the average percentage of information transmitted for each feature as a function of speech presentation level.

Linear regression analysis of percent of transmitted information for individual participants as a function of presentation level was performed to obtain slope, intercept, and r^2 values for each feature. Slope, as a function of feature, was used to quantify the impact of presentation level across speech features. A single-factor repeated-measures ANOVA was performed to determine if significant differences existed between feature slopes. Results showed a significant main effect of speech feature ($F_{5,35}=12.65$ $p<0.05$). Table VII shows the slope, intercept, and r^2 values as a function of speech feature for data averaged across participants.

An examination of Table VII shows that the duration feature was the most affected by increases in speech levels. Posthoc analysis, using a Tukey HSD test with a significance level of $p<0.05$, showed the slopes for the nasal, place, sonorance, and frication features were not significantly different from each other. The slope for the duration feature was significantly steeper than all other features except for nasal. In addition, the slope for the voicing feature was significantly shallower than that of duration, nasal, and place.

C. Effect of speech presentation level on perception of place of articulation

Previous research has suggested high speech levels may affect perception of place of articulation (Dorman and Dougherty 1981; Van Tasell and Crump 1981; Hedrick and Jesteadt 1997). The place feature used in this experiment was a five-valued feature divided into labial, dental, alveolar,

FIG. 3. Percent information transmitted, in noise, for each of the speech features assessed in this study as a function of speech presentation level (in dB SPL in a 2 cc coupler). Filled symbols again show performance in noise while the open symbols show performance in quiet at a 65 dB SPL presentation level.

TABLE VII. Average slope, intercept, and r^2 values obtained from data in Fig. 3.

Feature	Intercept	Slope	r^2
Duration	159.1	−1.17	0.88
Nasal	139.9	−0.90	0.91
Place	120.0	−0.80	0.97
Sonorance	119.2	−0.71	0.98
Frication	106.5	−0.68	0.98
Voicing	101.7	−0.41	0.95
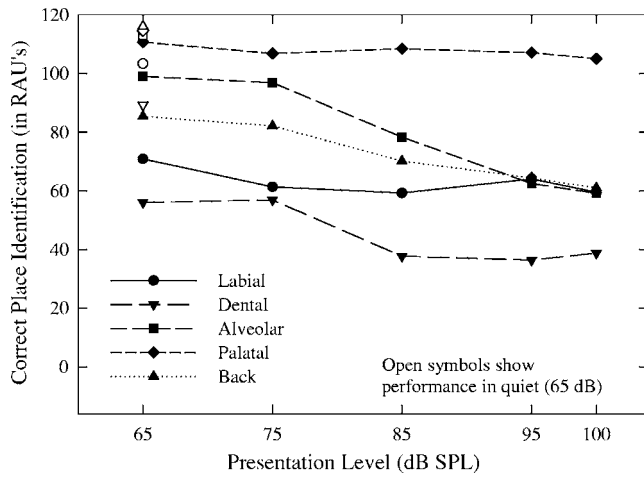
FIG. 4. Consonant recognition scores for each place of articulation examined in this study as a function of speech presentation level (in dB SPL in a 2 cc coupler). Filled symbols again show performance in noise while the open symbols show performance in quiet at a 65 dB SPL presentation level.

palatal, and back places of articulation. Figure 4 shows identification performance, as a function of presentation level, for each place of articulation.

Because the number of consonants with a given place of articulation varied substantially in this study (i.e., only two dental consonants but six alveolar consonants), only descriptive analyses of the results are provided. At low levels (i.e., 65 dB SPL), differences in performance as a function of place of articulation are due to the relative impact of the masking noise and are not of interest in the current study. For example at the 65 dB presentation level in quiet, scores for palatal consonants averaged 116 RAU. When the back-

ground noise was introduced scores remained nearly stable, averaging about 111 RAU. In contrast, dental consonant scores fell from 89 to 56 RAU with the introduction of the masker at the 65 dB presentation level. Our focus is on the effect of increasing presentation level and Fig. 4 shows that identification performance generally decreased with increasing level; however, the amount of decrement varied, based on place of articulation. A decrease in performance at higher presentation levels occurred across all places of articulation, but was greatest for the alveolar consonants. A smaller decrease was also seen for the dental and back consonants. Palatal consonants and labial consonants were only minimally affected by level.

D. Effect of speech presentation level on identification of individual consonants

Table VIII lists the performance (in RAUs) for each consonant as a function of presentation level. The last column shows the change in scores from the lowest presentation level (65 dB SPL) to the highest presentation level (100 dB SPL). As shown in Table VIII, and consistent with analysis of place of articulation slopes, performance did not vary substantially with level for the palatals /dʒ tʃ r/ or the labials /b f v/. Performance on all of these consonants decreased by less than 10 RAU over the range of levels tested. Also consistent with the place of articulation analysis, five of the seven consonants most affected by level were the alveolars /d n z s l/. Scores decreased approximately 30 to 59 RAU for these consonants as levels increased from 65 to 100 dB SPL. Identification of the palatal /ʃ/ and the back /w/ also showed substantial reductions as presentation level increased.

TABLE VIII. Score (in RAUs) for each consonant as a function of presentation level (dB SPL in a 2 cc coupler). Consonant scores are ordered based on amount of decrease in recognition as level increased from 65 dB to 100 dB SPL. The magnitude of this difference (in RAUs) is shown in the final columns.

Level dB SPL	65 dB	75 dB	85 dB	95 dB	100 dB	Difference 65–100
/d/	96.0	93.5	61.5	38.0	37.0	58.9
/n/	73.2	70.5	40.9	22.9	22.9	50.3
/z/	95.1	96.0	92.7	63.0	48.1	47.0
/ʃ/	93.5	90.4	75.4	61.5	57.1	36.4
/s/	100.7	98.7	92.7	73.7	64.4	36.3
/w/	91.9	86.9	63.9	58.6	56.7	35.3
/l/	80.0	74.8	53.8	50.0	48.6	31.4
/p/	62.5	58.1	42.4	36.6	32.5	29.9
/ʒ/	95.1	86.2	84.9	75.4	68.0	27.2
/j/	92.7	80.6	81.2	80.0	70.5	22.2
/ð/	46.2	40.5	27.3	25.7	24.1	22.2
/t/	116.0	116.0	116.0	108.2	96.0	20.1
/k/	79.4	77.6	75.9	69.0	61.5	17.9
/m/	84.3	68.0	63.0	77.6	68.5	15.8
/g/	71.6	67.5	61.0	55.7	57.6	14.0
/θ/	59.5	63.0	45.7	41.4	49.1	10.5
/b/	35.6	26.8	24.1	28.9	26.3	9.3
/f/	52.4	41.4	50.5	42.9	43.8	8.5
/r/	106.7	105.3	101.8	96.9	98.7	8.0
/v/	33.1	32.0	27.9	26.8	25.7	7.3
/tʃ/	112.0	106.7	106.7	108.2	112.0	0.0
/dʒ/	93.5	96.0	96.9	88.3	95.1	−1.6

V. DISCUSSION

The results of this study confirm previous findings of significant reductions in speech recognition with increasing speech levels (Pollack and Pickett 1958; Studebaker *et al.* 1999). When tested at a poor (0 dB) SNR, a significant reduction in performance was observed for speech levels just above 65 dB SPL and, on average, performance continued to decrease significantly with each 10 dB increase in level up to a 95 dB presentation level. Average performance dropped from approximately 77 to 55 RAU as presentation level increased from 65 to 100 dB SPL. The negative effect of increasing presentation level on consonant identification in the /i/ vowel context was comparable to that reported by Studebaker *et al.* (1999) when monosyllabic words were used as test material. Studebaker *et al.* (1999) measured speech understanding using NU-6 monosyllabic words at multiple SNRs as speech levels increased from 64 to 99 dB SPL. They reported baseline performance of approximately 73 RAU for speech presented at 64 dB at +12 dB SNR (see their Fig. 2, p. 2435). This 73 RAU starting point is comparable to the average performance of 77 RAU seen in the current study for speech presented at 65 dB SPL (0 dB SNR). The slope and intercept values derived from the current data illustrated in Fig. 1 (slope and intercept values of −0.636 and 119.4, respectively) correspond closely to data reported by Studebaker *et al.* (1999) for the +12 dB SNR (slope and intercept values of −0.6662 and 117.17, respectively), suggesting level effects were comparable, at least when performance was equated at low levels.

Also consistent with data reported by Studebaker and colleagues (1999), the effects of level observed in the current study were slightly larger than those predicted by the ANSI S3.5 (1997) (ANSI 1997) version of the SII. To compare predicted and measured reductions in performance with an increase in level, we first converted our 2 cc coupler speech and noise levels to free field levels, using the correction factor provided by Bentler and Pavlovic (1989). To predict performance, the transfer function for nonsense syllables provided in the ANSI S3.5 (1969) (ANSI 1969, Fig. 15) was used. Prior to comparison, predicted scores were converted to RAU values. SII values were reduced from 0.37 to 0.27 as levels increased from 65 to 100 dB SPL. This decrease in SII due to level increases results in a 19 RAU decrease in predicted performance compared to the 22 RAU decrease observed.

There is very little prior research on consonant feature transmission at high presentation levels. This study identified the acoustic-phonetic features of duration, nasality, and place as particularly susceptible to degradation at high presentation levels while the voicing feature was particularly resistant to level effects. (See Fig. 3 and Table II). These findings offer some support for the hypothesis that level effects are greater in the high frequency regions than the low-frequency regions (e.g., Molis and Summers 2003). Specifically, the voicing feature, which has substantial information located in the lower frequencies, was only minimally affected by level. In contrast, both the place and duration features, which were more affected by level increases than the voicing feature,

have substantial information located in the higher-frequency regions (Miller and Nicely 1955). The impact of frequency region on level effects, however, appears to vary across speech feature. Unlike voicing, the nasality feature, which also contains substantial information in the low-frequency regions (Miller and Nicely 1955), was substantially affected by increases in level.

Within the place feature, the alveolar place of articulation showed the largest decrease in identification at high presentation levels. Consonants with a palatal place of articulation were relatively resistant to the negative effects of high presentation levels. The confusions observed in this study suggest that some or all of the acoustic cues used for identification of a feature, or a specific consonant, were distorted at high presentation levels. The design of this study did not permit the specification of which cues were being distorted. An examination of some common confusions observed in this study, however, may help explain some of the perceptual consequences of presenting speech at high levels.

The most affected feature, duration, is used to differentiate four consonants /s ∫ z ʒ/ from the other consonants. Common confusions were palatal affricates /t∫/ and /dʒ/ replacing palatal fricatives /∫/ and /ʒ/ as presentation level increased. This confusion occurred in only one direction as the palatal affricates were rarely perceived as palatal fricatives. The number of confusions increased with presentation level, particularly at levels above 75 dB SPL. Fricatives have several perceptual attributes that differentiate them from affricates; including (1) the absence of a release burst, (2) a relatively long risetime between frication onset and maximum frication amplitude, and (3) long frication duration (Dorman *et al.* 1980; Howell and Rosen 1983). The difference in rise time between the voiceless fricative-affricate pair (/∫/ and /t∫/), in the CV position, is a particularly important acoustic/perceptual cue (Howell and Rosen 1983) and is clearly vulnerable to the negative effects of high presentation levels. Specifically, physiologic saturation within the cochlea or in neural firing patterns could result in changes in perceived rise times and lead to perceptual confusions of these contrasts.

Saturation of physiological response at the neural and/or cochlear levels due to high sound levels has been well documented. Wever and Bray (1938) demonstrated saturation-induced distortion in the inner ear at high intensities by analyzing changes in the cochlear microphonic (CM) in response to a 1000 Hz tone presented at high SPLs. The amplitude of the CM increased linearly for presentation levels up to 74 dB SPL. At presentation levels above 74 dB, increases in CM amplitude were nonlinear. CM amplitude began to decrease at levels above 94 dB, suggesting saturation of the physiological mechanism responsible for the CM.

Research on the neural representation of speech has also shown degradation at high presentation levels (Wong *et al.* 1998). Wong *et al.* (1998) recorded firing patterns from cat auditory nerves in response to the synthetic vowel /ɛ/ at presentation levels of 60–110 dB SPL. At low sound levels, fibers with characteristic frequencies near the F2 vowel formant "phase lock" on the stimulus. This phase locking is referred to as "synchrony capture" and occurred at levels

from threshold up to approximately 80–100 dB SPL (Wong *et al.* 1998). As presentation levels increased, the synchrony to F2 was lost and at higher levels (between 80–100 dB), the nerve lost its frequency resolution completely and demonstrated a saturated broadband response with a peak at F1.

The physiologic work of Wever and Bray (1938) and Wong *et al.* (1998) involve stimuli very different from the spectrally shaped noise found in production of fricatives. Despite this, the sharp reduction in correct identification of /ʃ/ as the presentation level increased from 75 to 85 dB is consistent with the idea of cochlear and/or neural saturation. For the consonant /ʃ/, at high presentation levels the peak levels of frication noise may begin to cause cochlear/neural saturation effectively "shortening" the rise time to peak amplitude. This would result in an apparent faster rise time from frication onset to peak amplitude and a shorter duration of audible frication noise. Both of these cues would tend to shift perception toward an àffricate response.

The cochlear and neural distortion described above may also play a role in a common place-of-articulation confusion observed in this study. The primary acoustic cues for place of articulation are frequency, and spectral discontinuities, specifically, (1) formant onset frequencies, particularly the F2 locus frequency, (2) F2 transitions, (3) spectral information in burst, frication noise, or nasal murmur, (4) the relative amplitude of the burst and frication noise, and (5) spectral discontinuities at consonant and vowel boundaries (Blumstein and Stevens 1980; Manrique and Massone 1981; Kewley-Port 1983; Ohde and Stevens 1983; Kurowski and Blumstein 1984; Ohde 1988; Ohde 1994). High presentation levels could distort any one, or a combination, of these acoustic cues.

In the current study, the alveolar consonants were most affected by increases in presentation level. An examination of specific consonants revealed a substantial decrease in correct identification for five (/n s d l z/) of the six alveolar consonants as the speech level increased from 65 to 100 dB SPL. Only identification of the voiceless /t/ remained relatively stable as speech level increased. This degradation appeared across all manners of production (i.e. stop, nasal, fricative, and liquid). The consonants that showed the largest decreases in identification were the stop consonant /d/, the fricative /z/, and the nasal /n/, with decreases of 57%, 44%, and 36%, respectively. Primary confusions for these consonants consisted of within-manner errors in which alveolar consonants were misidentified as labials. For example, the alveolar-nasal /n/ was commonly misidentified as the labial-nasal /m/ and the alveolar-stop /d/ was misidentified frequently as the labial-stop /b/.

The underlying cause of this confusion is not clear, however, the loss of neural synchrony at high levels reported by Wong *et al.* (1998) may play a role. The primary acoustic cues for identification of /d/ and /n/ are F2 onset frequency, spectral discontinuities at the CV boundary, and spectral information in the release burst and transition trajectory. A linear predictive coding analysis of a brief time window containing the release burst and formant onset reveals a rising spectral slope for alveolar place of articulation, while a flat or falling slope cues labial place of articulation (Blumstein

and Stevens 1980). The loss of neural synchrony of the F2 formant at high levels could reduce the salience of this important perceptual cue. At higher levels, the broadband or low-frequency emphasis observed in the neural response would be more consistent with labial rather than alveolar productions and may contribute to the performance degradation for these alveolar consonants. Alternatively, high presentation levels may perturb some other important static (e.g., F2 onset) and/or dynamic (e.g., time varying changes in formant transition) perceptual cues at the CV boundary that differentiate the alveolar and labial place of articulation features.

VI. CONCLUSIONS

The results of this research indicate that high presentation levels modify the salient acoustic cues of several features of speech. The current findings support the following conclusions:

(1) Overall consonant recognition in noise decreases in a relatively linear fashion as presentation levels increase above conversational speech levels.
(2) The acoustic-phonetic features of duration and nasality are particularly susceptible to the negative effects of high presentation levels.
(3) Consonants sharing alveolar place of articulation, particularly /d/, /z/, and /n/, show the largest decrement in performance at high levels.

Further research is required to isolate the specific factors responsible for the patterns of perceptual confusions seen at high levels. In addition, expansion of this work to include persons with hearing loss is needed to explore potential interactions between hearing loss and feature perception at high presentation levels.

ANSI. (**1969**). "Methods for the calculation of the articulation index," ANSI S3.5-1969, American National Standards Institute, New York.

ANSI. (**1996**). "Specification for audiometers," ANSI S3.6-1996, American National Standards Institute, New York.

ANSI. (**1997**). "The calculation of the speech intelligibility index," ANSI S3.5-1997, American National Standards Institute, New York.

Bentler, R. A. and Pavlovic, C. V. (**1989**).

Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants," J. Acoust. Soc. Am., **67**, 648–662.

Byrne, D., Dillon, H., Tran, K. Arlinger, S. Wilbraham, K., Cox, R., *et al.* (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am., **96**, 2108–2120.

Dorman, M. F., Raphael, L. J., and Isenberg, D. (**1980**). "Acoustic cues for a fricative-affricate contrast in word-final position," J. Phonetics, **8**, 397–405.

Dorman, M. F., and Dougherty, K. (**1981**). "Shifts in phonetic identification with changes in signal presentation level," J. Acoust. Soc. Am., **69**(5), 1439–1440.

1728   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Hornsby *et al.*: Consonant feature transmission at high levels

Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am., **91**, 2110–2121.

Dubno, J., and Schaefer, A. (**1995**). "Frequency selectivity and consonant recognition for hearing impaired and normal hearing listeners with equivalent masked thresholds," J. Acoust. Soc. Am., **97**(2), 1165–1174.

Fletcher, H., and Galt, R. H. (**1950**). "The perception of speech and its relation to telephony," J. Acoust. Soc. Am., **22**(2), 89–151.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am., **10**(1), 90–119.

Goshorn, E. L., and Studebaker, G. A. (**1994**). "Effects of intensity of speech recognition in high- and low-frequency bands," Ear Hear., **15**, 454–460.

Gutnick, H. (**1982**). "Consonant-feature transmission as a function of presentation level in hearing impaired persons," J. Acoust. Soc. Am., **72**(4), 1124–1130.

Hedrick, M. S., and Jesteadt, W. (**1997**). "Influence of relative amplitude and presentation level on perception of the /p/-/t/ stop contrast by normal and impaired listeners," *Modeling Sensorineural Hearing Loss*, W. Jesteadt, ed. (Lawrence Erlbaum Associates, Mahwah, N.J.), 475–486.

Howell, P., and Rosen, S. (**1983**). "Production and perception of rise time in the voiceless affricate/fricative distinction," J. Acoust. Soc. Am., **7**(3), 976–984.

Jerger, J., and Jerger, S. (**1971**). "Diagnostic significance of PB word functions," Arch. Otolaryngol., **93**, 573–580.

Kent, R. D., Wiley, T. L., and Strennen, M. L. (**1979**). "Consonant discrimination as a function of presentation level," Audiology, **18**, 212–224.

Kewley-Port, D. (**1983**). "Time varying features as correlates of place of articulation in stop consonants," J. Acoust. Soc. Am., **73**, 322–335.

Kurowski, K., and Blumstein, S. E. (**1984**). "Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants," J. Acoust. Soc. Am., **76**, 383–390.

Loven, F. C., and Collins, M. J. (**1988**). "Reverberation, masking, filtering, and level effects on speech recognition performance," J. Speech Hear. Res., **31**, 681–695.

Manrique, A. M. B., and Massone, M. I. (**1981**). "Acoustic analysis and perception of Spanish fricative consonants," J. Acoust. Soc. Am., **69**, 1145–1153.

Miller, G. A., and Nicely, P. A. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am., **27**(2), 338–352.

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low- and high-frequency speech," ARLO, **4**(4), 124–128.

Ohde, R. N. (**1988**). "Revisiting stop-consonant perception for two-formant stimuli," J. Acoust. Soc. Am., **84**, 1551–1555.

Ohde, R. N (**1994**). "The development of cues to the perception of the /m/-/n/ distinction in CV syllables," J. Acoust. Soc. Am., **96**, 675–686.

Ohde, R. N., and Stevens, K. N. (**1983**). "Effect of burst amplitude on the perception of stop consonant place of articulation," J. Acoust. Soc. Am., **74**, 706–714.

Ohde, R. N., Haley, K. L., Vorperian, H. K., and McMahon, C. W. (**1995**). "A developmental study of the perception of onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am., **97**(6), 3800–3808.

Pollack, I., and Pickett, J. M. (**1958**). "Masking of speech by noise at high sound levels," J. Acoust. Soc. Am., **30**(2), 127–130.

Speaks, C., Karmen, J. L., and Benitez, L. (**1967**). "Effect of competing message on synthetic sentence identification," J. Speech Hear. Res., **10**, 390–396.

Studebaker, G. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res., **28**, 455–462.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels." J. Acoust. Soc. Am., **105**(4), 2431–2444.

Van Tasell, D. J., and Crump, E. S. (**1981**). "Effects of stimulus level on perception of two acoustic cues in speech." J. Acoust. Soc. Am., **70**(5), 1527–1529.

Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual feature," J. Acoust. Soc. Am., **54**(5), 1248–1266.

Wever, E. G., and Bray, C. W. (**1938**). "Distortion in the ear as shown by the electrical responses of the cochlea," J. Acoust. Soc. Am., **9**, 227–233.

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel /ɛ/ in cat auditory nerve," Hear. Res., **123**, 61–77.

The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception

Catherine Mayo[a] and Alice Turk
Theoretical & Applied Linguistics, University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh, EH8 9LL, United Kingdom

Children and adults appear to weight some acoustic cues differently in perceiving certain speech contrasts. One possible explanation for this difference is that children and adults make use of different strategies in the way that they process speech. An alternative explanation is that adult-child cue weighting differences are due to more general sensory (auditory) processing differences between the two groups. It has been proposed that children may be less able to deal with incomplete or insufficient acoustic information than are adults, and thus may require cues that are longer, louder, or more spectrally distinct to identify or discriminate between auditory stimuli. The current study tested this hypothesis by examining adults' and 3- to 7-year-old children's cue weighting for contrasts in which vowel-onset formant transitions varied from spectrally distinct (/no/-/mo/, /do/-/bo/, and /ta/-/da/) to spectrally similar (/ni/-/mi/, /de/-/be/, and /ti/-/di/). Spectrally distinct cues were more likely to yield different consonantal responses than were spectrally similar cues, for all listeners. Furthermore, as predicted by a sensory hypothesis, children were less likely to give different consonantal responses to stimuli distinguished by spectrally similar transitional cues than were adults. However, this pattern of behavior did not hold for all contrasts. Implications for theories of adult-child cue weighting differences are discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1979451]

I. INTRODUCTION

A number of speech perception studies have shown that adults and children weight acoustic cues differently in identifying certain contrasts (e.g., Krause, 1982; Lacerda, 1992; Mayo *et al.*, 2003; Morrongiello *et al.*, 1984; Nittrouer and Studdert-Kennedy, 1987; Nittrouer and Miller, 1997; Ohde and Haley, 1997; Parnell and Amerman, 1978; Sussman, 2001; Watson, 1997; Wardrip-Fruin and Peach, 1984). For example, when identifying /s-vowel/-/ʃ-vowel/ contrasts on the basis of frequency of frication noise and vowel-onset formant transitions, children have been found to weight the transitions more heavily than do adults (e.g., Nittrouer and Studdert-Kennedy, 1987). Note that it is not simply that children and adults have different primary cues for these contrasts, as would be the case if children primarily used transitional cues and adults primarily used frication noise cues. Instead, Nittrouer and Studdert-Kennedy's (1987) study found that *both* adults and children gave more weight to frication noise than to transitional information. The difference between adults and children was in the *degree* of attention, or weight, given to different types of acoustic information available in the speech stream.

One explanation for these differences in cue weighting behavior is that adults and children make use of different strategies in the way that they process speech. One such hypothesis is Nittrouer and colleagues' Developmental Weighting Shift (DWS) theory, which proposes that children process speech in terms of large units (the size of a syllable

or monosyllabic word) while adults process in terms of smaller units. This processing difference then impacts on speech perception in terms of the attention that listeners give to acoustic cues. Children should therefore attend more than adults to cues that "delimit signal portions corresponding to syllables" (Nittrouer *et al.*, 2000, p. 268).

It has been suggested that cues which could be considered to be acoustic correlates of more global, syllable-based speech perception could be something like spectrally dynamic vowel-onset/-offset formant transitions (e.g., Nittrouer *et al.*, 2000). Thus, as found by Nittrouer and Studdert-Kennedy (1987) for fricative contrasts, children should give more weight than should adults to formant transitions, while adults' attention should be focused on more spectrally static cues (see also, e.g., Nittrouer, 1992, 1996; Nittrouer and Miller, 1997). However, some studies have found children to weight transitions *less* than do adults for certain contrasts (Howell *et al.*, 1992; Malech and Ohde, 2003; Mayo and Turk, 2004; Simon and Fourcin, 1978; Sussman, 2001). For example, Sussman (2001) examined 4-year-old children's and adults' (mean age 21 years, 6 months) weighting of two cues to vowel identification in CVC syllables: (*i*) vowel-onset and -offset formant transitions and (*ii*) vowel target formant frequencies (so-called vowel "steady-state" cues). The results of the study showed that although both adults and children were able to identify the vowels when only the transitional information was given to them, the children were not as successful as adults in doing so. Additionally, in identifying stimuli for which transitional information and static target frequency information were both available, but were in

[a] Electronic mail: catherin@ling.ed.ac.uk

conflict, the children made more use of the target information than did the adults.

Sussman (2001) proposes that these, and all, adult-child cue weighting differences are due to general sensory processing differences between adults and children rather than to strategy differences between the two groups (see also, e.g., Eisenberg *et al.*, 2000; Elliott, 1979; Elliott *et al.*, 1979, 1981). Under this explanation, the maturation of some aspect(s) of the central auditory system cause(s) adult-child speech perception differences as well as causing normal-hearing children to show poorer *nonspeech* perception abilities than normal-hearing adults (e.g., Allen and Wightman, 1994; Berg and Boswell, 2000; Elliott and Katz, 1980; Jensen and Neff, 1993; Maxon and Hochberg, 1982; Schneider and Trehub, 1992; Trehub *et al.*, 1988, 1995; Werner *et al.*, 1992).[1] Proponents of this theory suggest that "as auditory perceptual abilities mature over the first 10–12 years, central pattern recognition for speech may be much less robust to sensory distortion [than] that observed for adults" (Eisenberg *et al.*, 2000, p. 2705). In other words, children may be unable to cope, perceptually, with incomplete or insufficient acoustic information, and may thus need greater amounts of, or more distinct, acoustic cues than adults to identify or discriminate between auditory stimuli.

Sussman (2001) summarizes a number of these general acoustic theories in a proposal that states that children have a perceptual need for "louder or longer duration speech cues…or greater amounts of spectral information" (p. 1173). The acoustic speech cues used in Sussman's (2001) study described above were characterized in terms of this hypothetical need: The vowel-onset/-offset cues were described as shorter in duration, with "less overall energy" (p. 1174), while the steady-state cues were characterized as longer in duration and "more powerful" (p. 1174). Therefore, the fact that the children in the study were found to give more perceptual weight than were the adults to the vowel target cues is taken as evidence that "children rely more on the longer, louder or more acoustically salient cue" (p. 1179). This, in turn, is taken as support for a sensory-based, rather than a strategy-based, explanation for cue weighting behavior.

However, there are some aspects of Sussman's (2001) study that make it difficult to evaluate this claim. First, Sussman's study made use of a different paradigm from that used by Nittrouer and colleagues in the studies that formed the basis of the DWS (e.g., Nittrouer and Studdert-Kennedy, 1987; Nittrouer, 1996; Nittrouer and Miller, 1997). This makes it impossible to make a direct comparison between Sussman's study and those of Nittrouer and colleagues. Second, the primary evidence for the DWS theory comes from studies of perception of fricative contrasts (particularly /s-vowel/-/ʃ-vowel/ contrasts) while Sussman's study examined adults' and children's identification of vowel contrasts. This too makes comparison between the studies difficult: While it is possible that salience may play a role in adult-child cue weighting differences, it may (as in fact noted by Sussman 2001) simply be the case that children use different strategies for perceiving different segmental contrasts, relying more

heavily on vowel formant transitions for the identification of fricative contrasts, and more heavily on vowel target cues for the identification of vowels.

Finally, there is a significant problem with the assumption that the children in Sussman's (2001) study gave more weight to the vowel target cues than to the transitional cues because of the greater length and loudness of the target cues, rather than because of their cue type (i.e., vowel *targets* versus vowel *transitions*). First, the overall dB values for the two cues used in the study were not reported, making it impossible to evaluate the actual influence of the loudness of the cues. Additionally, the possible effect of formant frequency change in the transitions (as compared to the static vowel targets) is not taken into account. Such a frequency change could have had a subtle influence on the transitional cues' loudness (Fletcher and Munson, 1933) and could also impact for reasons other than loudness on the transitions' salience. Most critically, however, Sussman (2001) did not systematically manipulate cue salience in a way that can be readily interpreted. The study did not, for example, compare children's and adults' cue weighting patterns for long, loud vowel targets with their weighting patterns for short, soft vowel targets (paired with identical transitions in both cases). It is therefore unclear whether it was in fact solely the "loudness" of the target cue that influenced the weight given to it by children.

The aim of the current study, therefore, was to undertake a more systematic test of a general sensory explanation of adult-child cue weighting differences by addressing these issues. First, the study made use of the same methodology as that used by Nittrouer and colleagues, allowing direct comparisons between the current study and those that underpin the DWS theory. In addition, a perceptual paradigm was adopted which allowed both segmental contrast and cue type to be kept constant, and only cue salience, as characterized by loudness, length, or degree of spectral information, to be varied systematically.

The paradigm that allowed for the systematic manipulation of salience was modeled on Nittrouer and Studdert-Kennedy's original cue weighting research (Nittrouer and Studdert-Kennedy, 1987). That study examined adults' and children's weighting of vowel-onset formant transitions (as compared to frequency of frication noise) for two fricative contrasts: (i) /su/-/ʃu/ ("sue"-"shoe") and (ii) /si/-/ʃi/ ("sea"-"she"). The key difference between these two contrasts, with regard to the current study, is in the *spectral informativeness* of the vowel-onset formant transitions—one of the acoustic features that Sussman (2001) suggested should affect the perceived salience of a cue for children. Specifically, the onset frequency, extent, direction, and duration of the vowel-onset formant transitions in /su/ differ substantially from the onset frequency, extent, direction, and duration of the vowel-onset formant transitions in /ʃu/. This makes formant transitions a particularly informative cue to fricative identify for the /su/-/ʃu/ contrast. For the /si/-/ʃi/ contrast, on the other hand, there is much less difference in onset frequency, extent, direction, and duration of the vowel formant transitions across the two syllables, making transitions much less spectrally informative in this contrast. The results of Nittrouer and
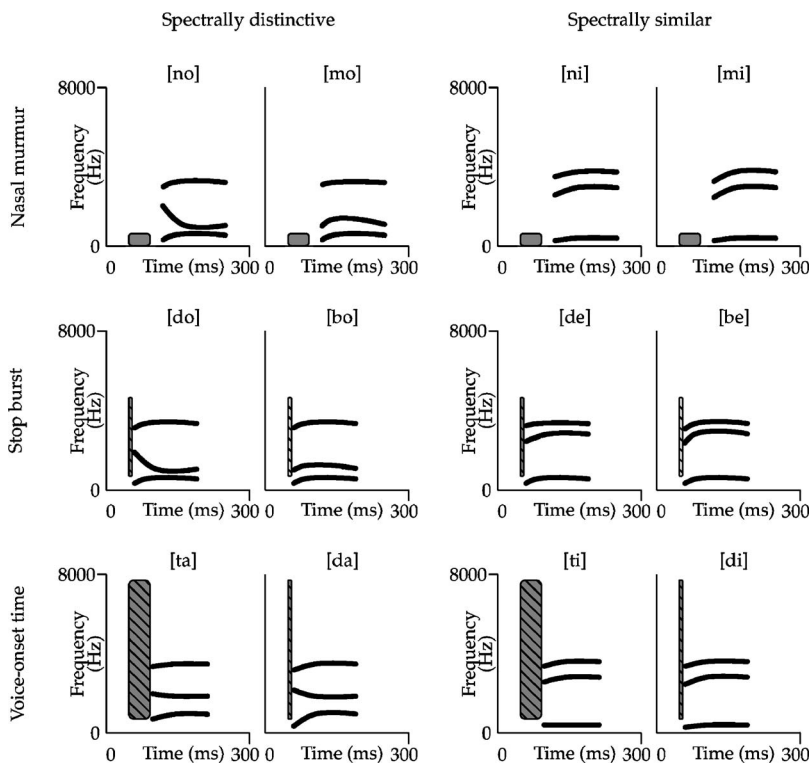
FIG. 1. Stylized spectrograms of prototypical tokens of the contrasts used in this study. The boxes represent: nasal murmur (/no/-/mo/, /ni/-/mi/), stop burst (/do/-/bo/, /de/-/be/), and burst+aspiration (/ta/-/da/, /ti/-/di/). The black lines represent vowel formants (F1, F2, F3).

Studdert-Kennedy's (1987) study showed that despite this difference in the spectral informativeness of the vowel formant transitions across the two contrasts, children gave more weight than did adults to the transitional cue for *both* /su/-/ʃu/ and /si/-/ʃi/.

It should be noted that the results of Nittrouer and Studdert-Kennedy (1987) do appear to favor a strategy-based explanation of adult-child cue weighting differences: The difference in spectral informativeness between the /su/-/ʃu/ contrast and the /si/-/ʃi/ contrast does not appear to have influenced children's cue weighting patterns. However, it remains possible that evidence for the influence of cue salience on adult-child differences in cue weighting in speech perception may emerge if a number of different segmental contrasts are examined. The current study therefore examined adults' and children's cue weighting for a range of consonantal contrasts.

A. The current study

The current study investigated adults' and children's (ages 3–4 years, 5 years, and 7 years) weighting of dynamic vowel-onset formant transition cues (compared to static steady-state cues) to one of three consonant contrasts: (*i*) the nasal contrast /n/-/m/, (*ii*) the stop burst contrast /d/-/b/, and (*iii*) the voice onset time (VOT) contrast /t/-/d/.

The focus in this study was on possible changes in listeners' transitional cue weighting patterns due to a change in degree of spectral information *across contrasts*. This is in contrast to Sussman's (2001) study, in which the focus was on differences in listeners' cue weighting patterns for two cues that appeared *within the same contrast*. This means that we did not specifically examine adults' or children's weighting of the steady-state, nontransitional cues to each contrast,

despite the fact that, for the identification of some contrasts, steady-state cues may be more important to both adult and child listeners than transitional cues. However, it is only listeners' behavior in response to the change in spectral information in the transitional cues that allows us to differentiate between the DWS theory and a more general auditory theory.

The study followed the methodology of Nittrouer and Studdert-Kennedy's (1987) /su/-/ʃu/ versus /si/-/ʃi/ study in that, for each listener, the segmental contrast (/n/-/m/ or /d/-/b/ or /t/-/d/) and the type of the cue of interest—here, vowel-onset formant transitions—were held constant, and only the salience of the cue of interest was varied systematically. Salience was varied by manipulating the transitions' spectral informativeness, by pairing each consonant with two different vowels, as in Nittrouer and Studdert-Kennedy's (1987) study. For example, the /n/-/m/ contrast was paired with the vowels /o/ and /i/ to give the contrasts /no/-/mo/ and /ni/-/mi/. The vowel-onset formant transitions in /no/ are spectrally quite distinct from those in /mo/ (particularly F2, which differs in both direction and extent of movement across the two syllables), making transitions spectrally informative in determining consonant identity for this contrast. The vowel-onset formant transitions in /ni/ are quite similar to those in /mi/, making transitions much less informative in identifying the consonants in this contrast. Each listener in the current experiment thus listened to two CV contrasts with identical consonants, but with different degrees of spectral information available in the transitional cues.

Figure 1 shows stylized spectrograms of the three contrast pairs used in this study: (*i*) /no/-/mo/ versus /ni/-/mi/, (*ii*) /do/-/bo/ versus /de/-/be/, and (*iii*) /ta/-/da/ versus /ti/-/di/.[2] It should be noted that when we refer to transitional cues for the two /t/-/d/ contrasts, we are in fact referring to *voiced* transitions: Unlike many studies of perception of

voicing contrasts (see, e.g., Kuhl and Miller, 1978), the synthetic stimuli used in the current study did not contain changing formant transitions for formants higher than F1 during aspiration. Instead, changing formant transitions for all three first formants began at voicing onset. Thus, although these stimuli sounded authentic, they were technically somewhat impoverished in comparison to natural speech.

In summary, this study aims to test the viability of Sussman's (2001) sensory explanation as an alternative, or an addition, to the DWS theory. Sussman's claim proposes that adult-child cue weighting differences are caused by children's heavier reliance on cues that are louder, longer, or more spectrally informative. The current study tests this claim by examining children's and adults' weighting of transitional cues in *both* a spectrally informative condition *and* a spectrally less informative condition. If the current study finds that children consistently give comparatively more weight than adults to the more spectrally informative transitions, and less weight than adults to the less spectrally informative transitions, then Sussman's claim will be supported.

It should be noted that Sussman's claim stems from a theory that children tend to rely on more informative cues due to some sort of sensory immaturity, possibly in terms of their central auditory processing. The current study did not attempt to address the issue of whether children and adults do in fact differ in their central auditory processing abilities. Additionally, this study did not directly test any sensory explanation for cue weighting behavior, but simply asks whether adult-child differences based on spectral distinctiveness or informativeness exist. Nevertheless, a positive result in the current study would be consistent with a sensory explanation of adult-child cue weighting differences.

II. METHOD

A. Participants

For the two /n/-/m/ contrasts, our testing criterion (described in Sec. II D) was met by ten out of ten adults tested (age range 21–29 years, average age 22 years), nine 7-year-olds (age range 7;1–7;11, average age 7;5 [year;month]) out of ten 7-year-olds tested, ten 5-year-olds (age range 5;0–5;9, average age 5;5) out of ten 5-year-olds tested, and four 3- to 4-year-olds (age range 4;4–4;10, average age 4;8) out of seven 3- to 4-year-olds tested.

For the two /d/-/b/ contrasts, our testing criterion was met by seven out of seven adults tested (age range 21–33 years, average age 26 years), nine 7-year-olds (age range 6;11–7;9, average age 7;5) out of ten 7-year-olds tested, seven 5-year-olds (age range 5;1–5;11, average age 5;5) out of seven 5-year-olds tested, and six 3- to 4-year-olds (age range 3;5–4;6, average age 3;11) out of eight 3- to 4-year-olds tested.

For the two /t/-/d/ contrasts, our testing criterion was met by eight out of eight adults tested (age range 21–49 years, average age 33 years), ten 7-year-olds (age range 7;0–7;11, average age 7;7) out of 11 7-year-olds tested, nine 5-year-olds (age range 5;1–5;8, average age 5;5) out of ten 5-year-olds tested, and nine 3- to 4-year-olds (age range 3;7–4;11, average age 4;1) out of 18 3- to 4-year-olds tested.

All children were in full-time preschool or primary (first and third year) education in Edinburgh (Scotland) and all were monolingual native speakers of Scottish Standard English (SSE). All of the children performed appropriately for their age on standardized tests of reading (Schonell Graded Word Reading Test, Schonell and Goodacre, 1971) and receptive vocabulary (BPVS, Dunn *et al.*, 1997). Parental questionnaires determined that all of the children and their siblings were free from speech/language disorders, hearing deficits, and histories of chronic otitis media (defined as more than three ear infections in the first 3 years, and/or the implantation of myringotomy tubes, see Nittrouer, 1996). No child was tested if he or she was suffering from, or had suffered from at any point in the weeks preceding the test session, any upper respiratory infection.

All adults were monolingual native speakers of English who had lived in the Edinburgh area for an average of 14 years. All of the adults reported themselves as being free from speech/language disorders, hearing deficits, and histories of chronic otitis media. Again, no adult was tested if he or she was suffering from, or had suffered from at any point in the weeks preceding the test session, any upper respiratory infection.

B. Stimuli

A trading-relations design was used for this study. For each individual contrast, two acoustic cues were manipulated: One vowel-onset transition cue and one nontransitional cue. The nontransitional cues were: Nasal murmur for /no/-/mo/ and /ni/-/mi/, frequency of stop burst for /do/-/bo/ and /de/-/be/, and voice onset time (VOT) for /ta/-/da/ and /ti/-/di/. The transitional cues were the changing formant frequencies (F1, F2, and F3) at the onset of the vowel (or at the onset of voicing in the case of the two /t/-/d/ contrasts). Two nine-point continua were created for each contrast by manipulating the two relevant acoustic cues for the contrast in question. The nontransitional cue was varied *along* the two continua from a value appropriate for the first consonant in the contrast to a value appropriate for the second consonant in the contrast. For example, for the /no/-/mo/ contrast, the nasal murmur was varied from a value appropriate for /no/ to one appropriate for /mo/ along both continua. The transitional cue, on the other hand, was varied *across* the two continua, with a value on one continuum that was appropriate for having followed the first consonant in the contrast, and a value on the other continuum that was appropriate for having followed the second consonant in the contrast. Thus, for the same /no/-/mo/ contrast, the value of the vowel formants at vowel onset were appropriate for having followed /n/ for the first continuum and appropriate for having followed /m/ for the second continuum. This design results in two continua which are identical in terms of the cue that varies along the continua and differ in terms of the cue that varies across the continua. This in turn allows for an investigation of the perceptual effect of the manipulated cues. A listener who is not influenced by the cue that changes across the continua—the vowel-onset transition information—will perceive the two continua as identical. On the other hand, a

listener who *is* influenced by the cue that changes across the continua will perceive the two continua differently. Following Nittrouer (1992), five different repetitions of the same vowel were created for each transition condition. Each of these was combined with each of the nine continuum values, resulting in 90 different stimuli per consonant contrast.

The stimuli used in this study were copy-synthesized versions of /no/-/mo/, /ni/-/mi/, /do/-/bo/, /de/-/be/, /ta/-/da/ and /ti/-/di/, created using the Sensyn (Sensimetrics Corp., n.d.) version of the Klatt cascade/parallel synthesizer (cascade for voiced and aspirated sounds, parallel for fricative and plosive burst sounds, Klatt, 1980). Copy synthesis is a method in which the values used to synthesize stimuli are closely modeled on detailed acoustic analysis of natural tokens of the target syllables (e.g., Hazan and Rosen, 1991). In this case, the targets were recorded by a male native speaker of Scottish Standard English, and acoustic formant measurements were taken at 10-ms intervals. It should be noted that as a result of this synthesis method, the change in formant frequencies in the synthetic vowel-onset formant transitions is neither stylized nor interpolated as a straight line change between vowel onset and vowel target values. Instead, the formants undergo a fairly rapid change in frequency immediately following the consonant (completed in 25–35 ms for bilabials, 30–40 ms for alveolars), followed by a slower change as the frequencies gradually reach the vowel target value. This gives formant transitions which resemble exponential curves. Additionally, because the durations of the transitional portions of the synthetic vowels are based on measurements of the whole natural vowel portion that changed in frequency over time (including the portion leading up to the vowel target which often contained more gradual changes), the durations of the synthetic transitions differ depending on the syllable being modeled. In particular, as will be noted below, contrasts with more extensively changing transitions (for example those in /no/-/mo/) will often show greater differences in total transition duration between the members of the contrast than will contrasts with less extensively changing transitions (such as those in /ni/-/mi/). A more detailed description of the parameters manipulated for each contrast follows, presented in pairs based on the initial consonants of the contrasts. All other parameters were left at the default values provided by the Klatt synthesizer (Sensimetrics Corp., n.d.). Note that the values used to synthesize the nontransitional cues were identical across consonant pairs (e.g., the nasal murmur values were the same for the /no/-/mo/ continua and the /ni/-/mi/ continua). Vowel-formant onset and target values for each individual vowel repetition can be found in Appendix A.

1. /ni/-/mi/ and /no/-/mo/ contrasts

Nasalization in the initial consonant and the first 25 ms of the following vowel was modeled by adding a single pole-zero pair to vowel-like spectra. Nine nasal portions were synthesized, ranging from the most /m/-like (F1: 207 Hz, F2: 1150 Hz, F3: 2200 Hz, nasal pole: 900 Hz, nasal zero: 1200 Hz) to the most /n/-like (F1: 207 Hz, F2: 1750 Hz, F3: 2832 Hz, nasal pole: 1000 Hz, nasal zero: 1800 Hz). For all the syllables, both the pole and the zero fell from their re-

spective values to 500 Hz over 20 ms after the onset of the vowel (at which point they merged, thus canceling each other out). The vowel from this point onward was non-nasal.

Two sets of vowels were created for each contrast, one with onset frequencies appropriate for a preceding /n/ and one with onset frequencies appropriate for a preceding /m/. The average /ni/-transition formant onset frequencies were F1: 252 Hz, F2: 1942 Hz, F3: 2632 Hz; the average /mi/-transition formant onset frequencies were F1: 262 Hz, F2: 1828 Hz, F3: 2494 Hz. The average vowel target values for all ten /ni/-/mi/ vowels were F1: 345 Hz, F2: 2228 Hz, F3: 2849 Hz. The average /no/-transition formant onset frequencies were F1: 267 Hz, F2: 1619 Hz, F3: 2299 Hz; the average /mo/-transition formant onset frequencies were F1: 257 Hz, F2: 801 Hz, F3: 2363 Hz. The average vowel target values for all ten /no/-/mo/ vowels were F1: 427 Hz, F2: 824 Hz, F3: 2421 Hz.

The total duration of each syllable was 485 ms, with 95 ms of nasal murmur, 25 ms of nasalized vowel, and 365 ms of oral vowel. The average duration of vowel formant transitions (including both rapidly and gradually changing portions) as measured from vowel onset to vowel steady state was 112 ms for /no/-transition stimuli, 74 ms for /mo/-transition stimuli, 50 ms for the /ni/-transition stimuli, and 58 ms for the /mi/-transition stimuli. F0 for each complete syllable began at 140 Hz at voicing onset, rose to 166 Hz 140 ms after onset, and fell to 97 Hz at vowel offset.

2. /de/-/be/ and /do/-/bo/ contrasts

Nine different complex bursts were synthesized. The spectral shape of the bursts was modeled by means of three spectral peaks, determined by the shape of the vocal tract at consonant release. The amplitude of these peaks ranged from 54 dB at 5550 Hz, 36 dB at 2700 Hz, and 20 dB at front cavity peak[3] (most /d/-like) to 6 dB at 4500 Hz, 0 dB at 2100 Hz, and 50 dB at front cavity peak (most /b/-like).

Two sets of vowels were created for each contrast, one with onset frequencies appropriate for having followed /d/ and one with onset frequencies appropriate for having followed /b/. The average /de/-transition formant onset frequencies were F1: 220 Hz, F2: 1809 Hz, F3: 2446 Hz; the average /be/-transition formant onset frequencies were F1: 257 Hz, F2: 1694 Hz, F3: 2247 Hz. The average vowel target values for all ten /de/-/be/ vowels were F1: 428 Hz, F2: 2116 Hz, F3: 2539 Hz. The average /do/-transition formant onset frequencies were F1: 240 Hz, F2: 1558 Hz, F3: 2389 Hz; the average /bo/-transition formant onset frequencies were F1: 236 Hz, F2: 722 Hz, F3: 2378 Hz. The average vowel target values for all ten /do/-/bo/ vowels were F1: 422 Hz, F2: 855 Hz, F3: 2469 Hz.

The total duration of each syllable was 400 ms, with 15 ms of burst, and 385 ms of vowel. The average duration of vowel formant transitions (including both rapidly and gradually changing portions) as measured from vowel onset to vowel steady state was 107 ms for /do/-transition stimuli, 37 ms for /bo/-transition stimuli, 100 ms for /de/-transition stimuli, and 110 ms for /be/-transition stimuli. F0 for each

complete syllable began at 140 Hz at onset of voicing, rose to 150 Hz 110 ms after onset of voicing, and fell to 90 Hz at vowel offset.

3. /ta/-/da/ and /ti/-/di/ contrasts

Nine different VOT values were synthesized, varying in 5-ms steps from 40 ms of aspiration (as generated by the AH parameter in the Klatt synthesiser, Sensimetrics Corp., n.d.) (most /t/-like) to 0 ms of aspiration (most /d/-like).

Two sets of vowels were created for each contrast, one with voiced onset frequencies appropriate for having followed /t/ and one with voiced onset frequencies appropriate for having followed /d/. The average /ta/-transition formant onset frequencies were F1: 537 Hz, F2: 1536 Hz, F3: 2551 Hz; the average /da/-transition formant onset frequencies were F1: 261 Hz, F2: 1642 Hz, F3: 2472 Hz. The average vowel target values for all ten /ta/-/da/ vowels were F1: 711 Hz, F2: 1433 Hz, F3: 2665 Hz. The average /ti/-transition formant onset frequencies were F1: 311 Hz, F2: 1924 Hz, F3: 2599 Hz; the average /di/-transition formant onset frequencies were F1: 221 Hz, F2: 1893 Hz, F3: 2569 Hz. The average vowel target values for all ten /ti/-/di/ vowels were F1: 309 Hz, F2: 2183 Hz, F3: 2819 Hz.

The total duration of each syllable ranged from 315 ms for the shortest VOT to 355 ms for the longest VOT, with 315 ms of voiced vowel in all cases. The average duration of vowel formant transitions (including both rapidly and gradually changing portions) as measured from voiced vowel onset to vowel steady state was 55 ms for /ta/-transition stimuli, 85 ms for /da/-transition stimuli, 110 ms for /ti/-transition stimuli, and 105 ms for /di/-transition stimuli. F0 for each complete syllable began at 124 Hz at onset of voicing, rose to 130 Hz 90 ms after onset of voicing, and fell to 60 Hz at vowel offset.

C. Procedure

Each participant listened to, and identified, stimuli from both the spectrally distinct and the spectrally similar conditions of a consonant contrast. Thus, listeners who listened to the /no/-/mo/ contrast, for example, also listened to the /ni/-/mi/ contrast. All participants were tested individually in a quiet room. The stimuli were presented over headphones (Sennheiser HD 490, frequency response 17–22 000 Hz), via a CD player. Volume was set at a comfortable listening level. Each participant was asked to indicate that the level was both comfortable and audible (for the children, the signal was split to two headphones and the chosen listening level was monitored by the experimenter); very few adjustments to the level were made by the participants. No adjustments to listening level were made within the presentation of a single contrast. Testing for the children took place over two or three days. Testing for the adults took place on one day, with a short break half-way through testing.

The listener's task was to identify individual stimuli as either one or the other half of a given contrast (e.g., as either "no" or "mow" for the contrast /no/-/mo/). The adult participants performed the task alone, by entering their responses on a form. The child participants provided their responses to the experimenter by saying the word aloud, and by placing a counter on a picture corresponding to the relevant word (see Appendix B for a description of the pictures used in the study).

Before testing, the children were given an opportunity to practice responding to natural productions of the target words. This ensured that the children were able to identify the targets in natural speech, and that they clearly associated each picture with the relevant target. The children received feedback throughout this practice, and did not proceed to the pretest with synthetic stimuli until they had, unprompted, correctly identified a complete set of ten randomly presented natural stimuli (five of each CV syllable).

A pretest was administered to both child and adult participants to ensure that they understood the task. This test consisted of the congruent endpoints of the continua, that is, the most extreme values of the nontransitional cue combined with the appropriate vowel-onset frequencies. For example, for the /ta/-/da/ contrast, the congruent endpoints were the 0-ms VOT plus vowels with /d/-transitions (the most /da/-like stimuli) and the 40-ms VOT plus vowels with /t/-transitions (the most /ta/-like stimuli). There were ten stimuli in the pretest (five per congruent endpoint), presented in random order. No feedback was given during this pretest.

Five different random orders of the 90 stimuli were created for each contrast. During the main test, the 5-year-old, 7-year-old, and adult participants heard the entire set of 90 stimuli twice, in two different random orders, resulting in 180 responses per participant and ten responses per vowel-onset type for each point on the continuum. The 3- to 4-year-old participants heard the set of 90 stimuli only once, resulting in five responses per vowel-onset type for each point on the continuum for this group. Although this smaller number of presentations may have led to noisier data than if ten responses per transition type had been collected, it was only practical to test a smaller number of responses for this age group because of limitations on the children's attention span. Subsequent examination of the data showed that the results from the 3-to 4-year-olds were not qualitatively different from those of the other child participants. Each randomization was split into blocks of ten stimuli for presentation. The interstimulus interval for the adults was 3 s, with an interblock interval of 10 s. Following Walley and Carrell (1983), there was no fixed interstimulus interval for the child participants. Instead, the presentation was paused briefly after every stimulus, allowing the children sufficient time to respond. At the end of each block, the children were allowed to choose a small prize (a sticker).

D. Analysis

Each consonant contrast pair had two sets of responses, one for each vowel-onset transition condition—i.e., spectrally similar or spectrally distinctive. It was important to maintain consistency of participants across the spectrally distinctive and spectrally similar conditions (that is, to ensure that if data from one participant were analyzed for the spectrally distinctive condition, data for that participant were also

analyzed for the spectrally similar condition). Therefore, only data from those listeners who responded correctly to 80% of the within-test congruent endpoints from the *spectrally distinctive* condition were included in analysis. Thus, if a participant was able to correctly identify 80% of the /no/-/mo/ congruent endpoints, for example, the responses of that participant for both the /no/-/mo/ and the /ni/-/mi/ contrasts were analyzed.

The responses for the two /t/-/d/ contrasts were normalized using a probit transformation. This transform extracts rate-of-change information from data appropriately modeled with an S-shaped curve and yields estimates of the slope and the mean of the curve (Cohen and Cohen, 1983). The slope is assumed to correspond to the degree of categoriality of the responses and the mean corresponds to the point on the continuum at which the responses reach 50% (i.e., 50% /t/ or /d/ responses). The responses for the two /n/-/m/ and two /d/-/b/ contrasts (which did not engender S-shaped response curves) were subjected to linear regression analysis. This yielded estimates of the slope and the *y* intercept of each regression line.

For each of the six individual contrasts (two per consonant contrast pair), the *degree of separation* of the two response curves/regression lines was calculated. For the probit-transformed responses, this was done by taking the difference of the two means; for the remaining responses, separation was calculated by taking the difference of the two *y* intercepts. Separation is thus a measure of the extent to which responses were affected by the change in vowel-onset transitional information across the two continua. Similarly, the *slope* of the response curves can be regarded as a measure of the degree to which listeners' responses were affected by the change in nontransitional information along the continua. Analyses of variance (ANOVAs) were calculated with either separation or slope of response curves as independent variables.

III. RESULTS

This section presents results for pairs of contrasts. As noted above, we did not examine the relative weighting of the two cues (transitional versus nontransitional) available within a single contrast (e.g., /no/-/mo/), but rather examined weighting of transitional cues *across* pairs of contrasts (e.g., transitional cues in /no/-/mo/ versus transitional cues in /ni/-/mi/). While it may be the case that the steady-state nontransitional cues are more or less heavily weighted than the transitional cues in any one individual contrast, our main interest was to identify any possible changes in weighting of transitions due to the change in spectral distinctiveness of that cue across the pairs of contrasts. The above comparison allows us to determine the influence of cue distinctiveness/informativeness on perceptual weighting.

Figures 2–4 show the average responses per age group for each contrast. It is clear from these graphs that there is an overall influence of degree of spectral informativeness on both adults' and children's responses. All listeners showed larger separation of response curves (indicating heavy weighting of the cue that changed across the continua,
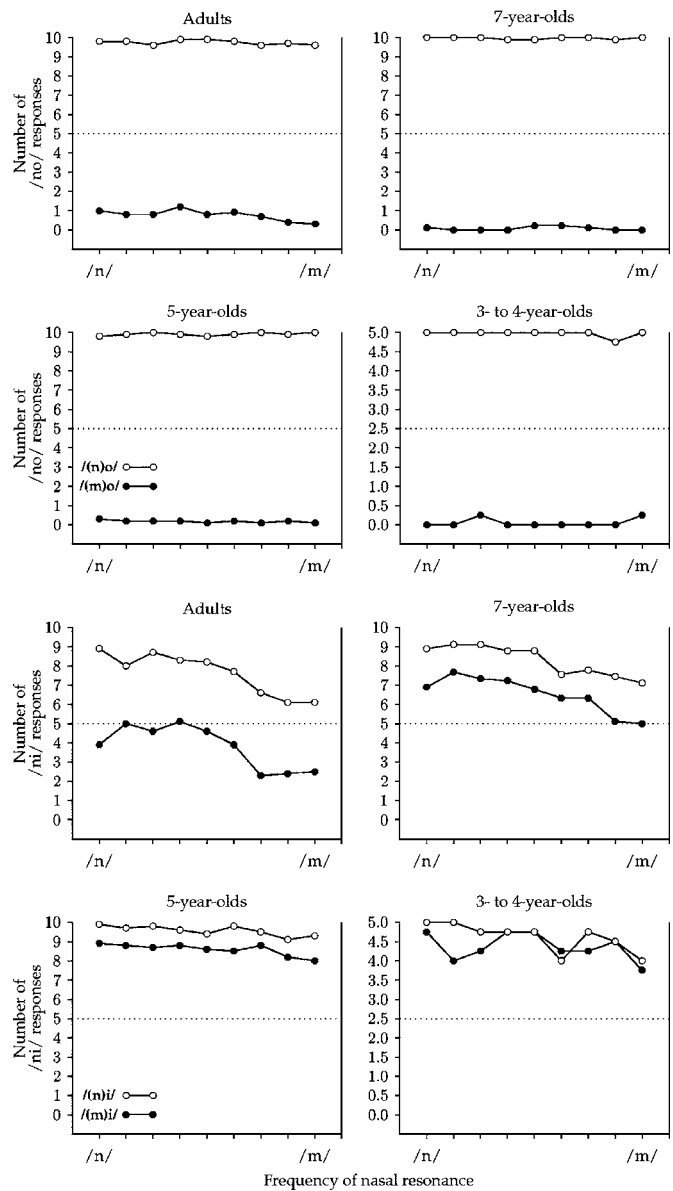
FIG. 2. Adults' and children's responses to /n/-transition stimuli (open circles) and /m/-transition stimuli (filled circles). The top four graphs represent responses to the /no/-/mo/ contrast; the bottom four graphs represent responses to the /ni/-/mi/ contrast. Responses are presented in terms of /n/-responses as a function of frequency of complex nasal murmur ranging from most /n/-like to most /m/-like (see text for details of frequency range). The dotted lines indicate the 50% /n/ response point. The *y* axis range for the 3- to 4-year-olds is half that of the other participants because this group heard half as many repetitions per point on the continuum.

namely the transitional cue) for those contrasts in which transitions were spectrally distinctive (/no/-/mo/, /do/-/bo/, /ta/-/da/) compared to those where transitions were spectrally similar (/ni/-/mi/, /de/-/be/, /ti/-/di/). ANOVAs with separation of response curves as the independent variables showed that these apparent differences were significant: /no/-/mo/ versus /ni/-/mi/ [$F(1,64)=163.00, p<0.001$]; /do/-/bo/ versus /de/-/be/ [$F(1,56)=85.00, p<0.001$]; /ta/-/da/ versus /ti/-/di/ [$F(1,70)=72.76, p<0.001$]. It should also be noted (although it is not the focus of this study) that there is some influence of context (specifically, the degree of spectral informativeness of the transitional cue) on adults' and children's weighting of the *non*transitional cues, particularly for
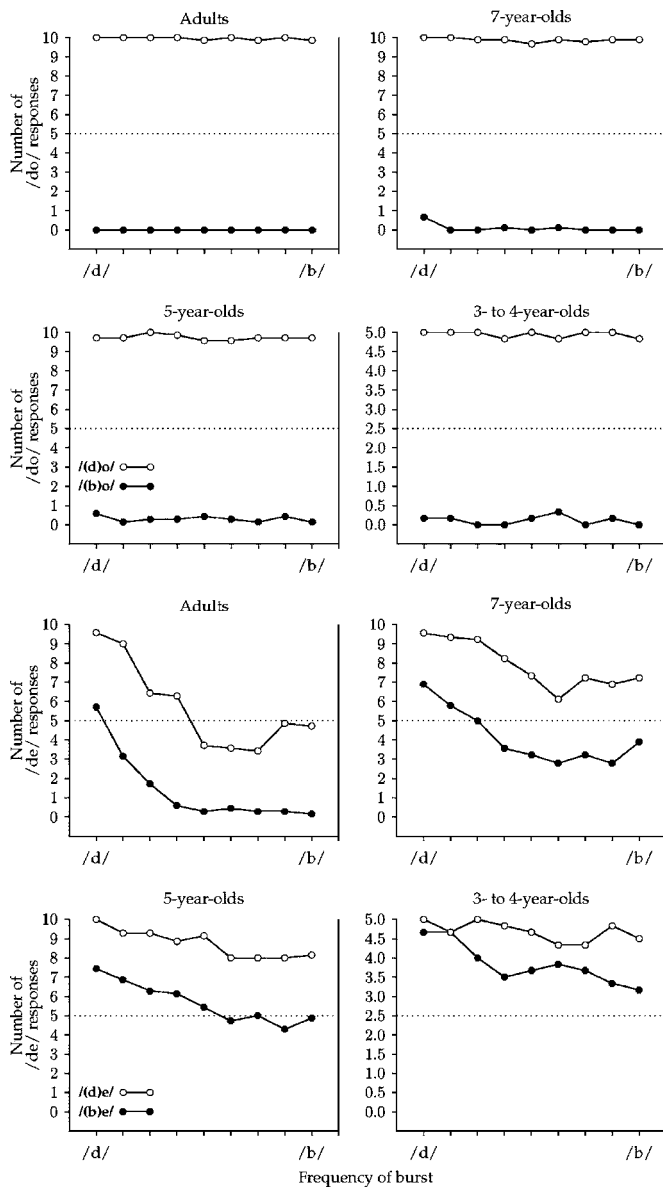
FIG. 3. Adults' and children's responses to /d/-transition stimuli (open circles) and /b/-transition stimuli (filled circles). The top four graphs represent responses to the /do/-/bo/ contrast; the bottom four graphs represent responses to the /de/-/be/ contrast. Responses are presented in terms of /d/-responses as a function of frequency of complex stop burst ranging from most /d/-like to most /b/-like (see text for details of frequency range). The dotted lines indicate the 50% /d/-response point. See Fig. 2 for more details.

FIG. 4. Adults' and children's responses to /t/-transition stimuli (open circles) and /d/-transition stimuli (filled circles). The top four graphs represent responses to the /ta/-/da/ contrast; the bottom four graphs represent responses to the /ti/-/di/ contrast. Responses are presented in terms of /t/-responses as a function of VOT ranging from most /t/-like (40 ms) to most /d/-like (0 ms). The dotted lines indicate the 50% /t/ response point. See Fig. 2 for more details.

the /n/-/m/ and /d/-/b/ contrasts. In spite of the fact that the acoustic characteristics of the nasal murmur continua were identical in the /no/-/mo/ and /ni/-/mi/ contrasts, and likewise the stop burst continua were identical in the /do/-/bo/ and /de/-/be/ contrasts, subjects were not able to use these cues when they were paired with the relatively distinctive /o/ transitions and steady states (as indicated by the lack of change in listeners' responses along the continua of steady-state cues). Given that the /ni/-/mi/ and /de/-/be/ results show that these nontransitional cues can be successfully used in categorization by all adult listeners and many older child listeners, it is clear that listeners' lack of successful use of these cues in the /no/-/mo/ and /do/-/bo/ contrasts is not simply due to their acoustic composition.

Turning to a comparison of adult and child transitional cue weighting behavior, the graphs indicate that for some consonantal contrasts, specifically /n/-/m/ and /d/-/b/, a change in spectral distinctiveness of transitional cues from more distinctive to less distinctive had a greater impact on children than on adults. For the spectrally distinctive /no/-/mo/ contrast, adults and children both had very widely separated response curves (again indicating heavy weighting of the formant transition cue). However, when the spectral distinctiveness of the transition was reduced in the /ni/-/mi/ contrast, children paid much less attention to the transitional cue than did adults. This is reflected in the fact that the separation of children's response curves was much smaller than that of adults for /ni/-/mi/. In fact the transitions in the

/ni/-/mi/ contrast did not cue a reliable difference between the two members of this contrast for most children. Similarly, for the /do/-/bo/ versus /de/-/be/ contrast pair, adults and children both showed widely separated curves in response to the spectrally distinctive transitions in /do/-/bo/, but children showed a smaller separation of curves than adults in response to the spectrally less distinct transitions in /de/-/be/. ANOVAs with separation of response curves as the independent variable again support these observations. There was no significant difference between adults and children in response curve separation for /no/-/mo/ but children showed significantly smaller response curve separation than adults for /ni/-/mi/ $[F(1,31)=16.85, p<0.001]$. A small difference was found between adults and children in response curve separation for /do/-/bo/ $[F(1,27)=4.67, p=0.04]$, however both groups appeared to make almost exclusive use of transitional cues in identifying this contrast. For /de/-/be/, the 7-year-olds patterned with the adults (there was no significant difference in response curve separation between these two groups) but the two younger groups of children (3- to 4-year-olds and 5-year-olds) again showed significantly smaller separation of response curves than the older listeners $[F(1,27)=6.40, p=0.02]$. These results thus lend support to Sussman's suggestion that children may have difficulty making use of acoustic cues—in this case transitional cues—that are less spectrally informative.

However, this pattern of children making less use than adults of less distinctive acoustic information did not hold for the two /t/-/d/ contrasts. First, there was *no* significant difference between adults and children in response curve separation for the contrast with spectrally similar transitions, /ti/-/di/. This seems inconsistent with Sussman's claim which would have predicted that children should have been less able to make use of this less informative cue than adults. However, it should be noted that the response curve graphs indicate that both adults and children appeared to make almost exclusive use of the nontransitional cue, VOT, and almost no use of the transitional cue. This exclusive use of VOT by both groups could explain a lack of difference in transitional cue use between the two groups. This could also suggest that, for this contrast, VOT was more informative to listeners than vowel formant transitions, although we cannot definitively establish degree of "informativeness" in this case, since VOT and vowel formant transitions are different types of acoustic cue. If VOT were the more informative cue, then we might expect to see a difference between adults and children in their weighting of this cue, as Sussman claims that children should make more use than should adults of more informative cues. However, an examination of the average slope of adults' and children's response curves (which reflects degree of weight given to the cue that changes *along* the continua, here VOT) for the /ti/-/di/ contrast shows that children's curves were less steep than were those of adults: An ANOVA with average response curve slope as the independent variable shows a significant difference between adults and children for this measure $[F(1,34)=29.53, p<0.001]$. This indicates that children gave *less* weight than adults to the VOT cue, counter to Sussman's predictions.[4]

Furthermore, this finding of adults giving more weight than children to a more informative cue is repeated in the results of the /ta/-/da/ contrast. Here, again counter to Sussman's predictions, adults were found to have response curves that were more separated than those of the children, indicating that they gave more weight than children to the spectrally distinctive transitional cue: An ANOVA with response curve separation as the independent variable shows a significant difference between adults and children for this measure $[F(1,34)=6.00, p=0.02]$. No difference was found between adults and children in weight given to VOT for this contrast. The results for the two conditions of the /t/-/d/ contrast are clearly problematic for Sussman's hypothesis, since adults, rather than children, were shown to give most weight to a spectrally more informative cue.

IV. SUMMARY AND DISCUSSION

The goal of the current study was to determine to what extent the spectral distinctiveness or informativeness of a given cue influences listeners' cue weighting patterns. In particular, this study aimed to test the effect of changing the spectral distinctiveness of vowel-onset formant transition cues on cue weighting in 3- to 7-year-old children as compared to adults. The motivation behind this test was a claim that possible sensory immaturities in their central auditory processing would cause children to be less able to use quieter, shorter, and less spectrally informative cues, and thus more likely to make heavier use than adults of louder, longer, and more spectrally informative cues (Sussman, 2001).

The results of the current study provide evidence that spectral informativeness of transitional cues (i.e., the degree of difference in spectral extent and direction of vowel-onset formant transitions) plays a role in speech perception tasks in general. All listeners, both adult and child, were more influenced by the spectrally distinctive vowel-onset formant transitions found in /no/-/mo/, /do/-/bo/, and /ta/-/da/ than they were by the spectrally similar transitions found in /ni/-/mi/, /de/-/be/, and /ti/-/di/.

This study also provides evidence that, for some consonantal contrasts, adults and children do differ in the degree to which their perception is affected by a change from spectrally distinctive to spectrally similar information. Although children and adults did not differ in their weighting of spectrally distinctive transitional information for /no/-/mo/, young children gave much less weight than did older children and adults to the spectrally similar transitional information in /ni/-/mi/. Similarly, children and adults differed only slightly in their weighting of spectrally distinctive transitions for /do/-/bo/, but children gave much less weight than did adults to the spectrally similar transitions in /de/-/be/. This provides some support for Sussman's claim that adult-child cue weighting differences could be caused by an inability in children to make use of shorter, quieter, or less informative cues.

However, the results for the /t/-/d/ contrasts, although providing further evidence of adult-child differences in cue weighting, do not support Sussman's sensory hypothesis. First, no difference was found between adults and children in

the weight given to the spectrally similar transitional cue in /ti/-/di/. A theory based on general auditory differences between adults and children, however, would predict that children should give less weight to this cue than adults. While this result could possibly be explained by the fact that it appears that *both* adults and children made very little use of the transitional cue for this contrast, adults were found to have made *more* use than children of the other available cue to the contrast, VOT. Again, this runs counter to a general auditory theory, which would predict that children should rely more heavily than adults on the more acoustically salient cue. Finally, children were also found to weight the spectrally distinctive transitions in /ta/-/da/ less heavily than adults; Sussman's sensory hypothesis predicts that children should weight more salient cues more heavily than should adults. The conclusion that can be drawn from the current study, therefore, is that while at least some adult-child cue weighting differences could be explained by a sensory hypothesis, this is not the case for all adult-child cue weighting differences.

Interestingly, the results for the two /t/-/d/ contrasts are also not predicted by Nittrouer's more strategy-based DWS hypothesis. Children were *not* found to weight transitional information more heavily than adults for either /ti/-/di/ or /ta/-/da/ (see also Mayo and Turk, 2004). It would seem, therefore, that any explanation for adult-child cue weighting which is based solely on sensory differences, or on a strong, transition-based interpretation of strategy differences, between adults and children, will not be able to account for all cue weighting differences between the two groups. Further research will be required to determine whether a combination of sensory and strategy explanations can account for all cue weighting differences, or whether these differences could be due to other general cognitive or perceptual factors.

ACKNOWLEDGMENTS

APPENDIX A: VALUES FOR SYNTHETIC CV STIMULI

Vowel-formant onset and target values for synthetic /ni/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	272	363	1935	2207	2614	2805
2	261	352	1931	2231	2632	2831
3	245	336	1938	2261	2636	2882
4	232	335	1922	2258	2632	2877
5	248	366	1986	2261	2645	2849

Vowel-formant onset and target values for synthetic /mi/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	255	344	1808	2220	2482	2837
2	276	342	1818	2223	2492	2868
3	238	357	1851	2208	2492	2829
4	283	334	1816	2202	2508	2872
5	260	325	1847	2211	2496	2835

Vowel-formant onset and target values for synthetic /no/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	270	412	1623	811	2293	2396
2	269	410	1579	808	2311	2388
3	246	414	1620	803	2217	2398
4	279	400	1614	849	2379	2440
5	270	438	1661	850	2293	2396

Vowel-formant onset and target values for synthetic /mo/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	259	428	890	805	2337	2415
2	234	443	757	833	2436	2462
3	261	432	765	837	2422	2461
4	259	440	792	842	2320	2424
5	271	452	802	801	2300	2429

Vowel-formant onset and target values for synthetic /de/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	219	425	1806	2096	2477	2554
2	231	436	1810	2116	2491	2542
3	221	416	1821	2093	2419	2536
4	221	443	1814	2109	2401	2544
5	206	426	1796	2112	2442	2520

Vowel-formant onset and target values for synthetic /be/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	249	420	1694	2100	2259	2535
2	247	429	1771	2114	2293	2540
3	271	427	1631	2094	2213	2524
4	247	429	1680	2117	2254	2540
5	273	429	1693	2104	2214	2553

Vowel-formant onset and target values for synthetic /do/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	258	423	1522	877	2426	2468
2	244	420	1545	876	2355	2453
3	219	427	1587	825	2400	2470
4	275	419	1560	878	2360	2470
5	206	424	1576	840	2403	2497

Vowel-formant onset and target values for synthetic /di/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	225	305	1857	2202	2560	2799
2	210	302	1904	2206	2574	2837
3	199	305	1923	2189	2573	2825
4	238	305	1870	2149	2534	2812
5	235	314	1911	2186	2605	2815

APPENDIX B: PICTURES USED TO ELICIT CHILD RESPONSES

CV syllable	Description of picture
[ni]	a knee
[mi]	a girl pointing to herself ("me")
[no]	a woman wagging her finger to say "no"
[mo]	a lawn-mower mowing a lawn
[de]	a street scene with the sun coming up ("day")
[be]	a typical Scottish bay
[do]	a person making cookie dough
[bo]	a ribbon tied in a bow
[ti]	a teapot and teacup ("tea")
[di]	a girl called "Dee"
[ta]	a girl receiving a present ("ta" is British English slang for "thank you")
[da]	a father ("da" short for "dada")

Vowel-formant onset and target values for synthetic /bo/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	258	422	724	880	2407	2498
2	231	409	729	836	2353	2469
3	219	422	725	837	2396	2473
4	244	429	720	854	2369	2442
5	228	421	721	847	2359	2451

Vowel-formant onset and target values for synthetic /ta/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	528	709	1530	1433	2560	2685
2	526	715	1524	1416	2536	2662
3	555	702	1555	1416	2513	2638
4	531	707	1541	1443	2577	2697
5	544	716	1528	1435	2564	2713

Vowel-formant onset and target values for synthetic /da/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	261	716	1631	1423	2498	2653
2	291	705	1629	1443	2490	2675
3	271	708	1667	1434	2442	2662
4	243	721	1643	1461	2496	2630
5	238	712	1639	1421	2433	2631

Vowel-formant onset and target values for synthetic /ti/ stimuli.

Stimulus no.	F1 onset	F1 target	F2 onset	F2 target	F3 onset	F3 target
1	324	324	1948	2192	2571	2831
2	306	306	1861	2171	2584	2752
3	316	316	1943	2184	2606	2832
4	300	300	1918	2153	2623	2831
5	310	310	1951	2197	2610	2856

[1]The exact physical location of the maturation has not yet been definitively established. However, most researchers accept that it is more likely to be in the central auditory system, as the auditory cortex continues to mature after birth (Eisenberg *et al.*, 2000), than in the peripheral auditory system, as the cochlea is fully developed at birth (Schneider and Trehub, 1992).

[2]The data for the /de/-/be/, /ta/-/da/, and /ti/-/di/ contrasts are also featured in Mayo and Turk (2004), where they are analyzed for a different purpose.

[3]This front cavity peak is modeled by the *frication excited bypass path* in Sensyn's Klatt synthesiser (Sensimetrics Corp.); no frequency value is given by Klatt for this parameter.

[4]The results seen for the /t/-/d/ contrasts could be due in part to the way in which the stimuli were constructed, namely that the aspiration noise did not contain transitional information. However, Howell *et al.* (1992) found similar results for edited natural /b/-/p/ stimuli (rather than wholly synthetic as in the current study). Additionally, although some of our adult listeners were native speakers of non-Scottish varieties of English, all of the adult listeners for the two /t/-/d/ contrasts were native speakers of SSE. Therefore, these results, which are inconsistent with both the DWS and a sensory hypothesis, cannot be due to any dialect differences between the adult and child listeners.

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

Berg, K. M., and Boswell, A. E. (**2000**). "Noise increment detection in children 1 to 3 years of age," Percept. Psychophys. **62**, 868–873.

Cohen, J., and Cohen, P. (**1983**). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd ed. (LEA, Hillsdale, NJ).

Dunn, L. M., Dunn, L. M., Whetton, C., and Burley, J. (**1997**). *British Picture Vocabulary Test*, 2nd ed. (NFER-NELSON, Berkshire, UK).

Eisenberg, L. S., Shannon, R. V., Schaefer, Martinez A., Wygonski, J., and Boothroyd, A. (**2000**). "Speech recognition with reduced spectral cues as a function of age," J. Acoust. Soc. Am. **107**, 2704–2710.

Elliott, L. L. (**1979**). "Performance of children aged 9–17 years on test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Elliott, L. L., and Katz, D. (**1980**). "Children's pure-tone detection," J. Acoust. Soc. Am. **67**, 343–344.

Elliott, L. L., Longinotti, C., Meyer, D., Raz, I., and Zucker, K. (**1981**). "Development differences in identifying and discriminating CV syllables," J. Acoust. Soc. Am. **78**, 669–677.

Elliott, L. L., Connors, S., Kille, E., Levin, S., Ball, K., and Katz, D. (**1979**). "Children's understanding of monosyllabic nouns in quiet and in noise," J. Acoust. Soc. Am. **66**, 12–21.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Hazan, V., and Rosen, S. (**1991**). "Individual variability in the perception of cues to place contrasts in initial stops," Percept. Psychophys. **49**(2), 187–200.

Howell, P., Rosen, S., Lang, H., and Sackin, S. (**1992**). "The role of F1 transitions in the perception of voicing in initial plosives," in *Speech, Hearing and Language: Work in Progress* (University College, London).

Jensen, J., and Neff, D. (**1993**). "Development of basic auditory discrimination in preschool children," Psychol. Sci. **4**, 104–107.

Klatt, D. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

Krause, S. E. (**1982**). "Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults," J. Acoust. Soc. Am. **71**, 990–995.

Kuhl, P. K., and Miller, J. D. (**1978**). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," J. Acoust. Soc. Am. **63**, 905–917.

Lacerda, F. (**1992**). "Young infants' discrimination of confusable speech signals," in *The Auditory Processing of Speech: From Sounds to Words* edited by M. E. H. Schouten (Mouton's-Gravenhage), pp. 229–238.

Malech, S. R., and Ohde, R. N. (**2003**). "Cue weighting of static and dynamic vowel properties in children versus adults," J. Acoust. Soc. Am. **113**, 2257(A).

Maxon, A. B., and Hochberg, I. (**1982**). "Development of psychoacoustic behaviour: Sensitivity and discrimination," Ear Hear. **3**, 301–308.

Mayo, C., and Turk, A. (**2004**). "Adult-child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (**2003**). "The influence of phonemic awareness development on acoustic cue weighting in children's speech perception," J. Speech Lang. Hear. Res. **46**, 1184–1196.

Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). "Trading relations in the perception of speech by five-year-old children," J. Exp. Child Psychol. **37**, 231–250.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.

Nittrouer, S. (**1996**). "The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM," J. Speech Hear. Res. **39**(5), 1059–1070.

Nittrouer, S., and Miller, M. E. (**1997**). "Predicting development shifts in perceptual weighting schemes," J. Acoust. Soc. Am. **101**, 2253–2266.

Nittrouer, S., and Studdert-Kennedy, M. (**1987**). "The role of coarticulatory effects in the perception of fricatives by children and adults," J. Speech Hear. Res. **30**, 319–329.

Nittrouer, S., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). "The effect of segmental order on fricative labeling by children and adults," Percept. Psychophys. **62**(2), 266–284.

Ohde, R. N., and Haley, K. L. (**1997**). "Stop-consonant and vowel perception in 3-and 4-year-old children," J. Acoust. Soc. Am. **102**, 3711–3722.

Parnell, M. M., and Amerman, J. D. (**1978**). "Maturational influences on perception of coarticulatory effects," J. Speech Hear. Res. **21**, 682–701.

Schneider, B., and Trehub, S. (**1992**). "Sources of developmental change in auditory sensitivity," in *Developmental Psychoacoustics*, edited by L. Werner and E. Rubel (American Psychological Association, Washington, DC).

Schonell, F., and Goodacre, E. (**1971**). *The Psychology and Teaching of Reading* (Oliver and Boyd, London).

Sensimetrics Corp. (n.d.), *SenSyn: Speech Synthesizer Package*, Cambridge, MA.

Simon, C., and Fourcin, A. J. (**1978**). "Cross-language study of speech pattern learning," J. Acoust. Soc. Am. **63**, 925–935.

Sussman, J. E. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?" J. Acoust. Soc. Am. **109**, 1173–1180.

Trehub, S., Schneider, B., Morrongiello, B., and Thorpe, L. (**1988**). "Auditory sensitivity in school-age children," J. Exp. Child Psychol. **46**, 273–285.

Trehub, S. E., Schneider, B. A., and L., H. J. (**1995**). "Gap detection in infants, children, and adults," J. Acoust. Soc. Am. **98**, 2532–2541.

Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.

Wardrip-Fruin, C., and Peach, S. (**1984**). "Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants," Lang Speech **27**(4), 367–379.

Watson, J. (**1997**). "Sibilant-Vowel Coarticulation In The Perception Of Speech By Children With Phonological Disorder," PhD thesis, Queen Margaret College, Edinburgh.

Werner, L. A., Marean, G. C., Halpin, C. F., Spetner, N. B., and Gillenwater, J. M. (**1992**). "Infant auditory temporal acuity-gap detection," Child Dev. **63**, 260–272.

Developing a speech intelligibility test based on measuring speech reception thresholds in noise for English and Finnish

Martti Vainio
Department of General Linguistics and Department of Speech Sciences, University of Helsinki, Helsinki, Finland

Antti Suni
Department of General Linguistics, University of Helsinki, Helsinki, Finland

Hanna Järveläinen
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Espoo, Finland

Juhani Järvikivi
Department of Psychology, University of Turku, Turku, Finland

Ville-Veikko Mattila
Multimedia Technologies Laboratory, Nokia Research Center, Tampere, Finland

A subjective test was developed suitable for evaluating the effect of mobile communications devices on sentence intelligibility in background noise. Originally a total of 25 lists, each list including 16 sentences, were developed in British English and Finnish to serve as the test stimuli representative of adult language today. The sentences, produced by two male and two female speakers, were normalized for naturalness, length, and intelligibility in each language. The sentence sets were balanced with regard to the expected lexical and phonetic distributions in the given language. The sentence lists are intended for adaptive measurement of speech reception thresholds (SRTs) in noise. In the verification of the test stimuli, SRTs were measured for ten subjects in Finnish and nine subjects in English. Mean SRTs were -2.47 dB in Finnish and -1.12 dB in English, with standard deviations of 1.61 and 2.36 dB, respectively. The mean thresholds did not vary significantly between the lists or the talkers after two lists were removed from the Finnish set and one from the English set. Thus the numbers of lists were reduced from 25 to 23 and 24, respectively. The statistical power of the test increased when thresholds were averaged over several sentence lists. With three lists per condition, the test is able to detect a 1.5-dB difference in SRTs with the probability of about 90%. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1993129]

I. INTRODUCTION

In mobile communications, the intelligibility of speech transmitted over a voice communication channel may vary considerably due to a number of interfering factors. Typical examples of such factors are transmission channel errors, electrical channel noise, and environmental noise at the talker. Many subjective test methods making use of speech as test materials have been developed to assess the intelligibility of speech. In telecommunications, subjective audio quality tests include human listeners who give their opinion of the performance of the telephone transmission system under evaluation used either for conversation or for listening to spoken material. In utilitarian assessment methods, a single quality scale is typically used to give the opinion of the overall quality of the system, whereas in analytic methods several quality scales, describing different aspects of quality, can be used to give a multidimensional view of the quality. However, none of the presently available tests are sufficiently applicable to evaluate speech processing algorithms, such as speech coders, channel coders, and noise suppressors, which are the essential components of any modern communication system. As these algorithms have been designed to process continuous speech, taking into account, e.g., temporal masking, their performance cannot be measured with short speech stimuli. Therefore the quality of the coders cannot be assessed by processing only a single word at a time or by measuring the intelligibility only for short segments of speech.

Different languages also impose different needs with respect to the normalization and balancing of the test materials. It can be argued that to be truly universal the tests should be multilingual and should also be built with common criteria with regard to the acoustic as well as the linguistic methodology. This, in turn, calls for additional care with regard to choosing the linguistic material for the test sentences: the material should be normalized and balanced so that it reflects the current adult language usage across the given languages and/or dialects. The current research introduces a reliable and efficient method—based on the measurement of speech reception thresholds in noise—to evaluate the intelligibility of speech in mobile communications for two languages, namely English and Finnish.

The earliest quantitative subjective speech assessment methods have typically focused on speech intelligibility. In addition to this, most early methods were designed to evaluate the intelligibility of single segments or phonemes. Moreover, the evolution of the testing methodology has proceeded from the use of purely nonsensical units such as monosyllables[1] towards a controlled use of real words, as in e.g., the family of rhyme tests[2–5] and whole sentences corresponding more closely to typical language use. In addition to this, a noise source is often used to simulate the interference in a real world situation beginning with the work of Egan.[6]

Kalikow, Stevens, and Elliot[7] introduced the first speech perception in noise test. This test used sentence-level materials to measure word-level intelligibility of the input at fixed speech and noise levels. The listeners were asked to repeat the final monosyllabic noun of a sentence. The length of the sentences was manipulated and both the key-word familiarity and predictability (high versus low predictability) were balanced within lists of 50 sentences. As a percent intelligibility measure the test is, however, limited by floor and ceiling effects.[8]

An alternative to percent intelligibility measures was developed by Plomp and Mimpen.[9] The speech reception threshold (SRT) is free of the above-mentioned limitations. The SRT is defined as the presentation level of test speech necessary for a listener to understand the speech correctly a specified percent of the time, usually 50%. In the SRT, the speech material consists of sentences that are presented either in silence or in the presence of a reference noise signal. In practice the SRT measurement requires an implementation of an adaptive listening test procedure where the intensity level of speech can be varied depending on the listener's responses, i.e., whether sentences are understood correctly or not. If a sentence is not understood correctly, the level of the next sentence is increased, whereas in the opposite case, the level is decreased. In this way, over a sequence of sentences, the level of speech should gradually be reaching a stable value, i.e., a listener's SRT, where speech is just understandable to the listener.

Recently Nilsson et al.[8] introduced a hearing in noise test (HINT) for sSRT (sentence speech reception threshold) measurements. Their starting point was the Bamford-Kowal-Bench (BKB)[10] sentences originally designed for use with British children and representative of children's speech. The sentences were revised by removing British idioms and controlling for sentence length and the number of present and past tense verbs, after which the sentences were normalized for naturalness, difficulty, and reliability.

While the methodology presented in Nilsson et al.[8] is very good for developing SRT tests in general, the sentences they used suffer from a number of shortcomings: (1) they do not represent current adult language usage and (2) they are not translatable to other languages without losing such critical features as phonetic and structural balance. The latter is naturally true for any set of sentences. Developing an SRT test for a new language, therefore, requires the development of a completely new set of sentences which is representative of that language alone. Furthermore, developing a more universal test requires the development of more than one language simultaneously. The situation clearly calls for methodology that uses common criteria for the languages in question. That is, the languages should be treated with criteria which are determined by those languages together.

The existence of large text corpora and high-quality linguistic tools for analyzing them offers new possibilities meeting the requirements listed above. The linguistic material can be balanced both phonetically and structurally in such a way that the end result is representative of current adult usage.

Our present study is largely based upon the HINT methodology, but intends to refine it by developing new test sentences for British English and Finnish, which would be representative of current adult speech. This, we hope, will reflect the real-world situations of the speech coding devices at work in a more realistic manner.

Following the procedure introduced in Nilsson et al.[8] we first created the new sentence lists, recorded and edited them, created speech-shaped masker noise spectra from the recordings, matched the recorded sentences for difficulty, tested the interlist reliability, as well as estimated the statistical power of the test itself.

In conjunction with this study, a fully computer-based test system which allows for a reliable sound reproduction, a fast test administration, and easy data collection, as well as an automated way to conduct statistical data analysis, was developed. With the system, the duration of a threshold measurement with a single list usually takes less than 2 min. The computerized system is not discussed further in this paper.

The following section outlines in more detail the development process of the test sentences in British English and Finnish. The rest of the article presents a study on the suitability of the sentences for SRT measurements followed by an experimental evaluation of the performance of the proposed intelligibility test.

II. DEVELOPMENT OF SENTENCE MATERIALS

Our goal was to create sentence sets that would be representative samples of the target language as used by adults in a relatively formal context with a minimal amount of phonological and phonetic reduction. Our aim was to produce a certain number of sentences (here 400, divided into 25 groups of 16 sentences each) rather than finding an unforeseen number of sentences meeting certain criteria from corpora. Nilsson et al.[8] report that list sizes of 10 and 12 are adequate for the adaptation in the SRT. As we could not know a priori what the list size for Finnish would be, we used 16 sentences per list to start with.

A. Source material

The source material for both languages consisted of different text corpora available for our use. For Finnish we used a selection of regional newspapers from the years 1998–2000 (Turun Sanomat, Hämeen Sanomat, Keskisuomalainen, and Alasatakunta) from SKTP-B archives.[11] The material contained approximately 20 million word tokens.

For English we used the Brown Corpus, Lancester-Oslo/Bergen Corpus, and Grolier Electronic Encyclopedia, containing about three million words. We also used a number of books from the Project Gutenberg (http://www.gutenberg.net) collection with approximately six million additional words to achieve the necessary number of suitable sentences for the sets. The Gutenberg collection includes a number of works dating from before the 20 century which have outdated and archaic word forms (see Sec. II B 2 below). Therefore these materials were excluded from the phone and word frequency calculations.

B. Selection procedure

1. Extraction of phonetic and lexical frequencies

The Finnish corpora were transcribed to phonetic form following the standards of carefully articulated spoken language. In Finnish this can be achieved with a fairly small set of rules as the grapheme to phoneme correspondence is very high. We did not take into account assimilation and coarticulation phenomena and indicated glottalization only in the case of a word starting with a vowel and the preceding word ending with one within a sentence.

The phonetic transcription of the English corpora was performed using the front end of an English version of the Festival speech synthesis system[12] (http://www.festvox.org/). The system uses decision trees trained with the CMU phonetic dictionary of approximately 100 000 words and a statistical parser to remove possible ambiguities.[13] The biphone and monophone frequencies were calculated from the transcriptions. Word frequencies and corresponding base forms for both languages were extracted from the corpora using a *functional dependency grammar* syntactic parser[14] (as implemented by Connexor Inc.) At this stage, we concluded that the possible errors introduced by the automatic syntactic analyses and letter-to-sound conversion would be so small as not to have any effect on the final results. That is, the phonetic balance is centered around the most common units and the grammatical well-formedness of the sentences in the final sets was verified by humans.

2. Filtering the sentences

A set of good candidates for the final set selection process was selected, in effect, by filtering the big corpora incrementally from millions of words into a set of approximately 1000 sentences per language. The first step in extracting suitable candidate sentences was to exclude the sentences of unwanted length from the large corpora. We measured the length of the sentences in syllables. Due to structural differences between English and Finnish, we could not use the same number of syllables for both languages. Due to suffixation, Finnish words tend to be much longer than English and a greater number of syllables is necessary to prevent ungrammatical and clipped sentences to be introduced into the sets. Therefore, we used 9 to 12 syllable sentences for Finnish, whereas for English 7 to 9 syllables were considered sufficient. For similar reasons, we omitted the most frequent words (such as articles and prepositions) from the base-form frequency counts in English. The Finnish material was syllabified with an algorithmic method whereas the English material was syllabified by using the CMU phonetic dictionary.[12] After the selection procedure approximately 30 000 Finnish and 20 000 English candidate sentences remained in the source material. At this stage the sentences were checked for grammatical correctness; the remaining sentences were analyzed with a syntactic parser[14] and incomplete sentences with respect to main syntactic roles were removed. This ensured that all of the sentences were syntactically complete and sentences lacking, for instance, a predicate verb were removed.

Cumulative biphone frequencies of candidates were then calculated based on the frequencies observed in the source corpora. Sentences which deviated the most from the average value were removed. Additionally, sentences with rare diphones (less than 1000 occurrences in Finnish corpora and less than 200 in English) were left out. The numbers reflect the size of the corpora as well as the size of the diphone inventories in each language, Finnish having a larger corpus and fewer diphones—ca. 700 as opposed to ca. 1500 in English.

A similar procedure was applied on the lexical level, removing sentences with extreme values with regard to cumulative frequencies of the base forms of words.

At this point we also used certain language-specific filtering. Finnish sentences containing words with nonpredictable pronunciation were removed i.e., words of foreign origin, as well as certain loan words, which could simply be identified by foreign letters (c, q, w, z, and x). The English material from the Gutenberg project included books from the late 19th and early 20th centuries and contained, therefore, archaic material such as, for instance, the words "thou," "hither," and "ye," which had to be accounted for. Although most of the problematic forms had already been removed on the grounds of low base-form frequencies of the archaic words, nevertheless, we still had a native English talker check the final 1000 candidates for naturalness.

3. Balancing of the sets

After reducing the number of candidates to approximately 1000 sentences per language, the final 25 groups of 16 sentences were created. Intergroup variation was minimized according to average sentence length in phonemes, average word base-form frequency and distribution of phones. Simultaneously the monophone frequencies of the groups were matched with the source corpora using χ^2 test for goodness-of-fit. Minimization was performed by first creating randomized sets containing 400 sentences, normalizing the observed variances for each variable (average sentence length, average word base-form frequency) and then using a brute-force algorithm to find balanced sets minimizing the total variance using the whole set of 1000 sentences. All pairwise changes of sentences were tested between the groups and between the sentences left out of the groups. The changes which reduced variance to a significant degree were kept. This process was allowed to continue until no improvement was observed.
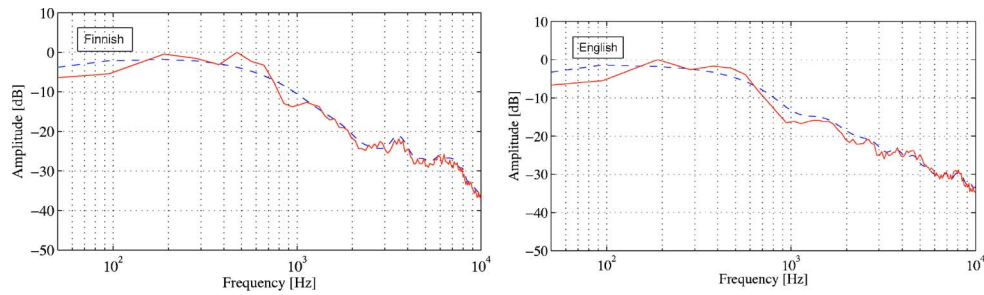
FIG. 1. (Color online) Magnitude spectra of the SRT test sentences in Finnish and English (solid lines), and the matching respective noise spectra (dashed lines).

C. Resulting materials

As a result 25 sets of 16 sentences for both languages were obtained. The χ^2 measure for the English sets varied from 9.45 to 17.55 (mean 13.27) ensuring that each set did not significantly differ from the whole corpora at chance level of 95% (i.e., $\alpha=0.005$, $df=50$); the respective values for Finnish were 10.39 to 16.74 (mean 13.11) for 41 degrees of freedom. The average length of the sentences in the sets varied between 23.3 and 24.0 phones for English and 31.9 and 32.9 phones for Finnish. Therefore, we could expect that the Finnish sentences would also be somewhat longer in duration than the English ones. The cumulative word base–form frequencies in the Finnish sets varied $\pm1.6\%$ from the grand average and the English sets varied between $\pm1.0\%$. The average lengths of the sentences (in syllables) were Finnish, 12.7 syllables (standard deviation 1.1), and English, 7.49 syllables (standard deviation 0.76). The resulting average utterance durations were 2.1 s (standard deviation 0.27 s) for Finnish and 1.7 s (standard deviation 0.32 s) for English.

III. RECORDING AND EDITING OF SENTENCE MATERIALS

The sentence materials in each language were recorded by four talkers. Two talkers in each language were male and two female. All were native speakers of British English or Finnish. Sentence lists 1–12 and 13–25 were spoken by one male and one female in each language. The Finnish talkers were voice professionals: two actors, a phonetician, and a speech therapist. All of the English talkers were teachers accustomed to public speaking but not specially trained to use their voice. None of the talkers represented an extreme voice type of any kind.

The recordings were made in an acoustically anechoic chamber. A Macintosh G3 computer with a 16-bit sound card (dynamic range 90 dB and frequency response ±0.5 dB between 30 Hz and 18 kHz) was connected to a Bruel & Kjaer (B&K) 2238 Mediator sound level meter, which includes a B&K 4188 condensator microphone and a BZ 7126 preamplifier. Headphone monitoring was provided for the talkers, and they were instructed to maintain a neutral speech style and volume. The microphone was always placed at about 0.50 m distance from the talker's mouth. The sentences were recorded directly to individual sound files by using the QuickSig measurement software,[15] which also allows filtering and amplitude monitoring. The sentences were checked

for clipping and they were linear phase high-pass filtered at 70 Hz to remove induction noise. The sampling rate was 44.1 kHz. The resulting signal-to-noise ratio of the recordings was approximately 50 dB.

The sentences were then edited by removing any extra silence at the beginning and end. Other unwanted sounds, such as inspiration and lip smacks, were also removed. The signals were then up-sampled to 48 kHz and their mean-squared (MS) amplitudes were equated to 60 dB (relative to one sample unit in a 16-bit digital representation).

At this point we created the speech-shaped masker noise spectra by computing a long-term spectrum of both Finnish and English speech material. We summed up all the sentences in both languages separately, computed short-term spectra from these samples in 512-point slices, and averaged to obtain an estimate of the long-term spectrum for each language. Infinite impulse response (IIR) filters were then designed with frequency responses matching the long-term spectra in 128 frequency points up to 10 kHz. The criteria produced a 64-zero, 2-pole filter for Finnish and a 64-zero, 16-pole filter for English. White noise was then generated and filtered with the designed filters to produce language-dependent, spectrally matched, masking noise for the speech samples. The thick lines in Fig. 1 show the long-term spectra of Finnish and English, and the dashed lines show the corresponding filter responses.

IV. MATCHING SENTENCE DIFFICULTY

It was expected that the individual intelligibility of the sentences would not be equal in spite of the equal MS amplitude and overall phonetic balance. Therefore, initial intelligibility of each sentence was tested and the sentences were then rescaled according to the intelligibility scores with the aim that reverse variations in the signal level would compensate for the observed variations in initial intelligibility.

A. Experimental setup and procedure

(a) *Participants*. Two groups of eight listeners participated in both the Finnish and English tests. The participants were native talkers of English and Finnish. The English group consisted of people of British, American, and Australian origin. Prior to the test, the participants were screened for normal hearing in the range of 125 Hz to 8 kHz. The participants were paid for their participation.

(b)   *Procedure*. The sentences were presented through headphones in an acoustically controlled reference listening room which satisfied the ITU-R BS.1116 Recommendation.[16] The speech and noise samples were mixed by the computer and played at −4 dB S/N ratio. The participants were instructed to type in what they had heard after each sentence. No feedback was given as to the correctness of the responses.

All the sentences were presented to each listener in four 75-min sessions. Prior to the actual SRT test material, a set of 20 practice sentences was presented. The sentences were similar to the actual test materials, but lacked the phonetic balance. The playback order of the lists as well as the sentences within lists was randomized. The materials from different talkers were presented to the participants in different order.

After the first test group, intelligibility scores were computed for each sentence of each individual talker. The score was based on the percentage of correct words in the responses to a given sentence averaged over all listeners. Each sentence that was less intelligible than all sentences on average was scaled up relative to 1 dB in MS level for each 10% difference. Sentences with higher intelligibility scores than average were scaled down in a similar manner. Thus, for instance, a 15% increase in the intelligibility score led to −1.5-dB decrease in MS level. The test was repeated for the next group of listeners with the rescaled sentences, and new level adjustments were made based on the new listeners' responses.

B. Results

After the first group of listeners, the overall mean intelligibility was 73% (standard deviation, std=32%) for the Finnish and 61% (std=34% correct) for the English sentences. After the second round, the mean scores were 70% and 71%, respectively, and the standard deviations had decreased to 29% correct and 28% correct, respectively. The distribution of the final MS adjustments are depicted in Fig. 2 for both the Finnish and English sentences. About 50% of the adjustments for Finnish and about 60% for English fell between ± 1 dB. However, the adjustments were not distributed evenly between the talkers. Sentences spoken by Finnish males were generally scaled down (−1.8 and −0.7 dB for male 1 and male 2 on average), whereas the sentences spoken by females were scaled up (+1.8 and +0.6 dB on average). For the English sentences, one of the male voices was generally scaled down (−1 dB in average), while the other male voice and one female voice were scaled up (+0.5 dB in average). The other female voice had a 0-dB mean scaling. Although the differences in intelligibility between talkers were mostly compensated by the sentence difficulty equalization, some talker effect could still be observed in the final adaptive SRT measurements. This will be discussed further below.
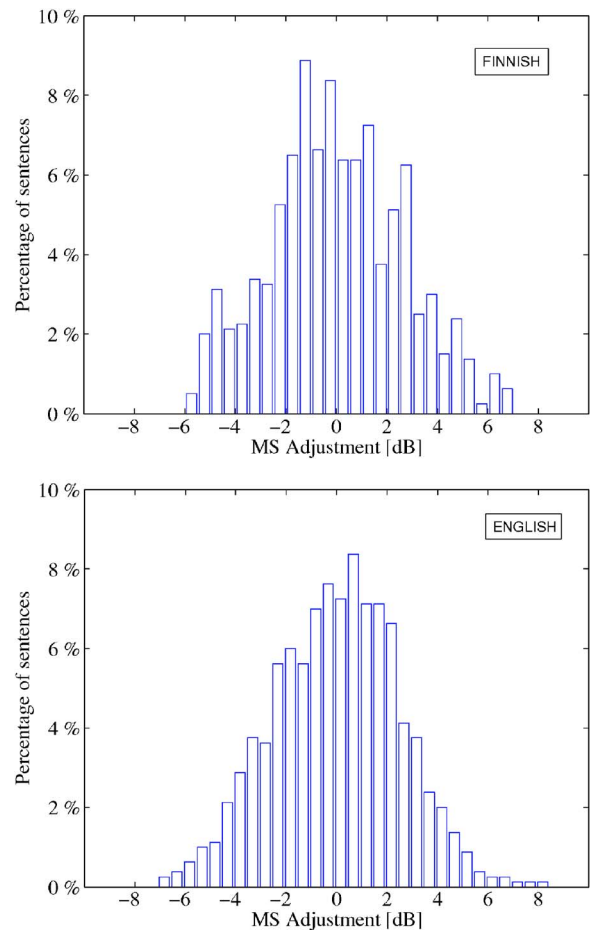
FIG. 2. (Color online) Distribution of MS amplitude adjustments for Finnish and English sentences (bar width=0.5 dB).

V. TESTING INTERLIST RELIABILITY

In order to determine whether each of the 25 sentence lists would produce a reliable and replicable measure of speech intelligibility, the SRTs were measured using an adaptive procedure and the results were analyzed statistically with analysis of variance (ANOVAs). The power of the test was also studied in order to determine the threshold differences.

A. Method

(a)   *Participants*. A total of ten native talkers participated in the SRT measurement test for Finnish and nine in English. Prior to the test, the participants were screened for normal hearing in the frequency range of 125 Hz to 8 kHz.

(b)   *Method*. The sentences were presented through headphones using the GuineaPig3,[17] computer-based, subjective test system in the same listening room as in the previous experiment.

The noise and the speech samples were mixed by the system at a S/N level that was dependent on subjects' performance in the previous trial. After a correct response, the S/N ratio of the following sentence was decreased by 2 dB, while an incorrect response increased it by the same amount. The S/N ratio in the first sentence of the test block was −5

dB relative to the overall rms level before scaling. The first sentence was repeated and its S/N ratio was increased in 2-dB steps until the subject had given a correct answer. All the other sentences were presented only once, regardless of whether the response was correct or not. No feedback was given to the subjects about their responses.

(c) *Procedure.* The SRT measurements were divided into two blocks of 12 and 13 lists each. Each subject was presented with the first 12 lists spoken by one talker and lists 13–25 spoken by another talker. The distribution of different talkers was balanced, so that the sentences from each talker were presented a total of five times to the subjects. The order of test blocks was varied between subjects and the list order was randomized by the test system.

The task was to type in the sentence the listener had heard using the GuineaPig3 test interface. This time minor variations were allowed in articles ("a/the") and verb tense ("is/was," "are/were," "has/had") for the English sentences, and in verb tense [as in "on/oli" ("is/was"), or "puhuu/puhui" ("speaks/spoke")] and the nearly identical singular versus plural forms [as in "syytä/syitä" ("reason/reasons"), or "maata/maita" ("land/lands")] in Finnish. No capital letters nor punctuation were required, nor was the system sensitive to added or missing spaces in compound words. Other than the exceptions mentioned above, a correct response required getting all words correct. The sentence presentation levels, the responses, and their correctness were recorded by the computer system. Since the sentence scoring was automatic, there was no way to treat spelling errors.

B. Results

The SRT was computed for each list as the average of the fifth and all the subsequent presentation levels within that list, including the determined level of the 17th trial that would be presented next. The mean S/N ratio at threshold over all lists was −2.47 dB in Finnish with a standard deviation of 1.61 dB, and in English −1.12 dB with a standard deviation of 2.36 dB. The current mean threshold for English was higher and the variability was larger than those measured by Nilsson *et al.*,[8] who obtained a mean threshold of −2.92 dB, with a standard deviation of 0.78 dB. The current mean threshold for Finnish was closer to the previous study. The use of four talkers instead of one as in Nilsson *et al.*[8] is a source of extra variability in both languages.

There are a number of possible explanations for the observed differences between the Finnish and English SRTs. The differences could be due to either the talkers, the listeners, or the sentence materials *per se* and, consequently, the languages themselves. While the Finnish talkers were either actors or phoneticians, the English talkers had had no professional voice training. That is, the differences could be explained simply by the quality of the speech and articulation. This can, however, only be a partial explanation. Another fact about the tests concerns the English speaking listeners; although the talkers were British, not all the listeners were. The results were therefore compared between British

and American subjects, and a significant difference was at first observed ($p=0.007$). However, we discovered that two out of the nine subjects scored poorly compared to all the others, and both of them happened to be British whereas the best performing subjects were not American. In fact, if those two listeners are removed from the results, the difference between British and American subjects becomes nonsignificant with $p=0.41$. Therefore, we can conclude that the language differences, again, cannot be explained by the listeners.

We next considered the sentence materials themselves as a source for the differences in SRT scores. The most conspicuous difference between the sentence materials is that the Finnish sentences are made up of fewer, although longer, words than the English ones. The Finnish sentences are also longer with respect to the number of syllables and, therefore, duration. This provides them with more redundancy which could make them easier to perceive in noise. However, the sentence length had no significant effect on the SRTs in either language. This was checked by comparing the intelligibility scores of the longest and shortest 10% of the sentences in both languages. No significant correlations were found.

The final, and perhaps the most crucial, difference has to do with the languages themselves and how certain sounds are distributed within them. Finnish has a much greater relative number of voiceless stops than English, 20.53% as opposed to 12.1% in the current materials, respectively. Thus, there are significantly more voiceless gaps in Finnish speech signals than in the English ones. The average power values that the Finnish signals were normalized with are, therefore, more affected by the voiceless gaps in Finnish than in English. Consequently, the Finnish signals have slightly better S/N ratios, which could explain the differences in SRTs. There are, thus, a number of factors which work additively in favor of the Finnish materials in the tests. Fortunately, these factors have no bearing on the final applicability of the sets themselves.

A reliable SRT measure should reveal true differences in intelligibility. The criteria for this are that the SRTs obtained from different lists should not differ statistically, and that the probability of observing a difference in SRT by chance should be reasonably low. The test results were therefore analyzed in terms of variance and statistical power.

(a) *List equivalence.* Differences were scored between the mean SRT for each list and the mean SRT over all lists and subjects. They are presented in Fig. 3. The error bar shows one standard deviation below and above the average. The mean differences in the Finnish sentences were less than 1.0 dB for all lists, except for lists 12 and 16. After the removal of these lists, the one-way ANOVA did not reveal significant differences with respect to the list means ($p=0.14$). Having only list 16 removed, the ANOVA remained nonsignificant at the 0.05 level with $p=0.07$. None of the English lists differed significantly from the overall mean, the ANOVA being nonsignificant with $p=0.64$, although for lists 10, 16, and 22 the mean deviation exceeded 1.0 dB.
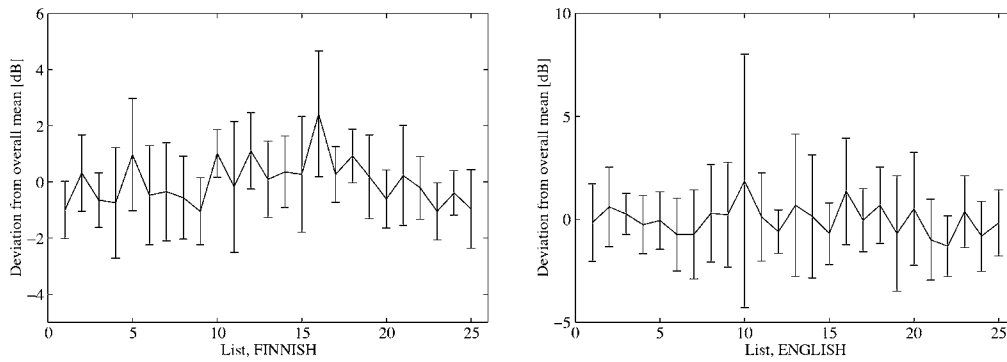
FIG. 3. Differences between mean SRTs of each list against the overall mean SRT in Finnish and English. The error bar shows one standard deviation above and below the average.

(b) *Statistical power*. The probability of correctly detecting a threshold difference can be estimated by studying the power of the test. Intuitively, it is clear that using a larger number of lists per condition reveals a true change in SRT more reliably than using only one list.

Let us first study the difference between two SRTs obtained from repeated measures within subjects, i.e., SRTs from lists spoken by the same talker and heard by the same listener. The standard deviation of the difference scores from all listeners was 1.61 dB for the Finnish material and 1.15 dB for the English material, resulting in 95% confidence intervals (±1.96 standard deviations) of ±3.15 and ±2.27 dB, respectively. Averaging two difference scores decreased the standard deviation to 1.40 and 0.97 dB, and by averaging three difference scores it was decreased to 1.01 and 0.81 dB. Thus, the confidence interval is also reduced when more lists are used per condition, and the threshold can be estimated more reliably.

Statistical power means the probability of correctly rejecting the null hypothesis that the true means of SRTs measured from different conditions are equal. The power measure is obtained by first finding the acceptable region for the observed sample mean of differences between the repeated measures, assuming the true mean difference to be 0 dB. This was done by a one-tailed Student's t test on a 5% risk level. Using this critical mean difference as decision criterion, it was computed how probably an even greater difference between two conditions would be observed, assuming now that the true mean difference is greater than zero. This was done by using the cumulative t distribution. Figure 4 shows the statistical power as a function of the mean difference in SRTs between conditions. It is seen that a 0.5-dB difference can be detected with no higher than 25% probability, while a difference of 1.5 dB can be detected with about 80% probability in the Finnish test and with about 95% probability in the English test. It is also evident from Fig. 4 that calculating the SRT as an average of two or three lists per condition increases the power of the test. It is thus recommended to use three lists per condition.

C. Required list length

The SRTs were recalculated for the reduced list lengths of 14, 12, and 10 sentences in addition to the original length of 16 sentences per list. The mean SRTs between the different list lengths varied within 0.25 dB in Finnish and 0.07 dB in English. For the Finnish sentences, the standard deviation of SRTs increased from the original value of 1.61 dB to 1.69, 1.78, and 1.91 dB, respectively. For the English sentences, it increased from the original value of 2.36 dB to 2.45, 2.55, and 2.69 dB. As the increase in standard deviation was relatively small, it was concluded that the list length could well be reduced to 10 sentences per list without a significant difference in the thresholds.

D. The talker effect

The material was recorded using two male and two female talkers in both languages. Based on the differences in
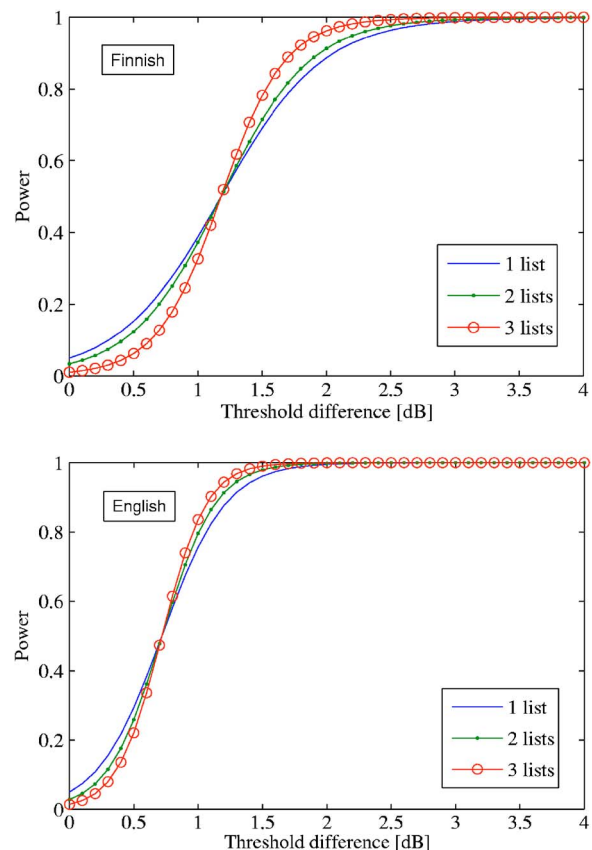
FIG. 4. (Color online) Statistical power of the SRT test in Finnish and English for one, two, and three lists.

1748   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Vainio *et al.*: Speech intelligibility test: English and Finnish
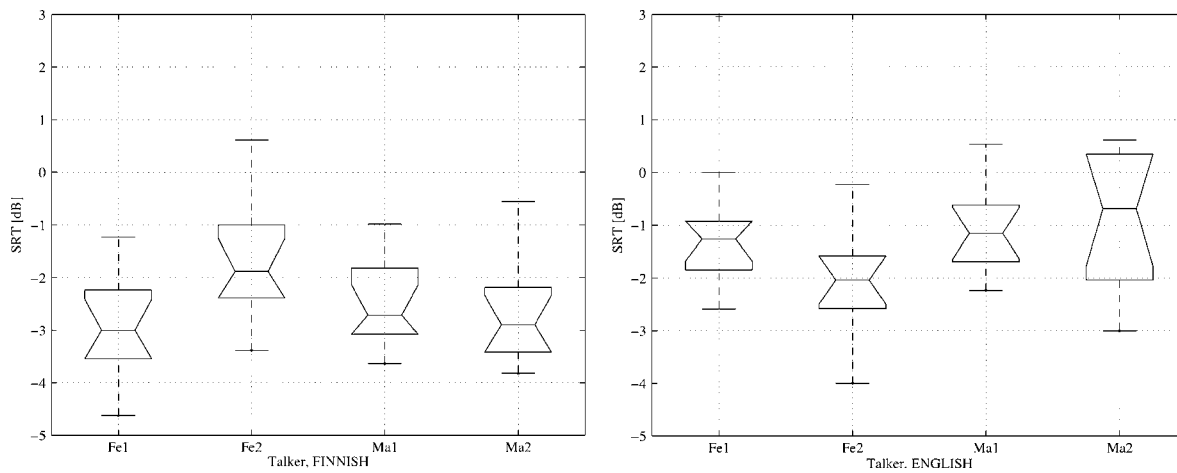
FIG. 5. Box plots of the SRT results for the different talkers in Finnish (left panel) and English (right panel) including all lists. The box plot shows the median, the upper, and the lower quartiles of the data.

the distribution of the sentence rescaling operations between talkers, it was assumed that some talker-induced variation may have been present in the SRT test. Therefore, the results were grouped according to the talker and an ANOVA was carried out between groups, revealing a significant effect in Finnish ($p=0.047$) and a marginally significant effect in English ($p=0.074$) (see Fig. 5). The Tukey HSD (honestly significant difference[18]) posthoc test was used to find significant differences in mean SRTs over each talker. The critical difference at the 0.05 uncertainty level was 0.68 dB for Finnish and 0.80 dB for English. In both Finnish and English, the second female talker differed slightly from the other talkers. All other differences in pairs of talker means were nonsignificant.

For female 2, the individual lists, having a mean SRT differing the most from the general mean, were lists 19 and 13 in English and lists 15 and 16 in Finnish. Removing these lists from both female 2 and male 2 diminished the talker effect in both languages. After the removal of the lists, the ANOVA was nonsignificant ($p=0.174$) in Finnish. In English, removing list 19 alone resulted in a nonsignificant ANOVA ($p=0.213$) and the removal of list 13 did not have a significant effect ($p=0.386$).

VI. SUMMARY AND CONCLUSIONS

A subjective test suitable for evaluating the effect of mobile communications devices on the sentence intelligibility in background noise was developed. A total of 400 sentences were carefully developed and recorded in British English and Finnish to serve as the test stimuli representative of adults' language use. The sentences, produced by two male and two female talkers for each language, were matched for naturalness, length, arid intelligibility. The sentence sets were balanced with regard to the expected lexical as well as phonetic distributions in the given language and the grammaticality of the sentences was checked. This resulted in 25 balanced sets of 16 sentences for both British English and Finnish.

The sentence lists were designed to be used in adaptive measurement of speech reception thresholds (SRT) in noise.

The adaptive procedure together with the sentence material will make it possible to compare devices with a high resolution and accuracy.

SRTs were measured for both Finnish and English speaking subjects using the currently developed system and sentence lists. Some differences were found between the language groups which could be explained by linguistic differences, differences between listeners in the experiments, and differences between the talkers who read the sentences. The differences do not, however, impede the use of the test in any way.

The reliability of the SRTs was determined by statistical analysis of the results. The mean thresholds did not vary significantly between lists with the exception of two of the Finnish lists. The list length required to reach a stabilized SRT could be reduced to ten sentences per list without any significant difference in the obtained thresholds. A significant talker effect was found in both languages. A posthoc analysis verified that in both Finnish and English, the voice of one female produced slightly different SRTs than with the other three talkers. This problem was overcome by removing two lists from the Finnish set and one list from the English set. Thus the final material consists of 23 Finnish and 24 English lists.

The sensitivity of the test stimuli has been verified through the use of within-subject repeated measurements of SRTs with different lists. The 95% confidence intervals for the difference scores were $\pm$ 3.15 and $\pm$ 2.27 dB for Finnish and English, respectively. The confidence interval describes the sensitivity of the SRT test: a greater threshold difference can be expected to be caused by a true difference in the intelligibility. When the difference scores were averaged over three repeated measurements, the confidence interval reduced to $\pm$ 1.97 dB in English and $\pm$ 1.58 dB in Finnish. Thus, by using multiple lists per condition, smaller differences in SRT can be reliably detected.

The SRT test presented here provides an accurate, reliable, and efficient method for measuring speech intelligibility in noise. Based on the results, the test is directly appli-

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Vainio *et al.*: Speech intelligibility test: English and Finnish   1749

cable to testing the intelligibility of various speech communication devices such as those used for mobile communications.

ACKNOWLEDGMENTS

[1] H. Fletcher and J. C. Steinberg, "Articulation testing methods," Bell Syst. Tech. J. **8**, pp. 806–854 (1929).

[2] G. Fairbanks, "Test of phonemic differentiation," J. Acoust. Soc. Am. **30**, 596–600 (1958).

[3] A. S. House, C. E. Williams, M. H. L. Hecker, and K. D. Kryter, "Articulation testing methods: Consonantal differentiation with a closed response set," J. Acoust. Soc. Am. **37**, 158–166 (1965).

[4] W. D. Voiers, *Speech Intelligibility and Speaker Recognition* (Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1977), Vol. **2**, Chap. 32.

[5] W. D. Voiers, "Evaluating processed speech using the Diagnostic Rhyme Test," Speech Technol. **1**, 30–39 (1983).

[6] J. P. Egan, "Articulation testing methods," Laryngoscope **58**, 955–991 (1948).

[7] D. N. Kalikow, K. N. Stevens, and L. L. Elliot, "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351 (1977).

[8] M. Nilsson, S. D. Soli, and J. A. Sullivan, "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099 (1994).

[9] R. Plomp and A. Mimpen, "Improving the reliability of testing the speech reception threshold of sentence," Audiology **18**, 43–52 (1979).

[10] J. Bench and J. Bamford, *Speech-Hearing Tests and the Spoken Language of Hearing-impaired Children* (Academic, London, 1979).

[11] CSC-Center for Scientific Computing, "Finnish language bank," http://www.csc.fi/kielipankki/ (2003).

[12] A. Black and P. Taylor, *Festival Speech Synthesis System: system documentation (1. 1.1)*, Tech. Rep., Human Communication Research Centre Technical Report HCRC/TR-83 (1997).

[13] S. DeRose, "Grammatical category disambiguation by statistical optimization," Comput. Linguist. **14**, 31–39 (1988).

[14] P. Tapanainen, "Parsing in two frameworks: finite-state and functional dependency grammar," Ph.D. thesis, University of Helsinki, 1999, URL citeseer.nj.nec.com/tapanainen99parsing.html

[15] M. Karjalainen, T. Altosaar, and P. Alku, "QuickSig—An object-oriented signal processing environment," in *Proc. IEEE (ICASSP)* (IEEE, New York, 1988), pp. 1682–l685.

[16] A. Järvinen, L. Savioja, H. Möller, V. Ikonen, and A. Ruusuvuori, "Design of a reference listening room—A case study," in *Proceedings of the Audio Engineering Society; 103th International Convention* (Audio Engineering Society, New York, 1997), preprint number 4559.

[17] J. Hynninen and N. Zacharov, "GuineaPig-A generic subjective test system for multichannel audio," in *Proceedings of the Audio Engineering Society; 106th International Convention* (Audio Engineering Society, New York, 1999), preprint number 4871.

[18] R. Lehman, *Statistics and Research Design in the Behavioral Sciences* (Wadsworth, Pacific Grove, CA, 1991), pp. 369–370.

Contextual variation in the acoustic and perceptual similarity of North German and American English vowels

Winifred Strange
Ph.D. Program in Speech and Hearing Sciences, City University of New York—Graduate School & University Center, 365 Fifth Avenue, New York, New York 10016-4309

Ocke-Schwen Bohn
English Department, Aarhus University, DK-8000 Aarhus C, Denmark

Kanae Nishi
Speech and Hearing Sciences, Indiana University, Bloomington, Indiana

Sonja A. Trent
Psychology Department, University of South Florida, Tampa, Florida 33620

Strange *et al.* [J. Acoust. Soc. Am. **115**, 1791–1807 (2004)] reported that North German (NG) front-rounded vowels in hVp syllables were acoustically intermediate between front and back American English (AE) vowels. However, AE listeners perceptually assimilated them as poor exemplars of back AE vowels. In this study, speaker- and context-*independent* cross-language discriminant analyses of NG and AE vowels produced in CVC syllables (C=labial, alveolar, velar stops) in sentences showed that NG front-rounded vowels fell within AE back-vowel distributions, due to the "fronting" of AE back vowels in alveolar/velar contexts. NG [ɪ, e, ɛ, ɔ] were located relatively "higher" in acoustic vowel space than their AE counterparts and varied in cross-language similarity across consonantal contexts. In a perceptual assimilation task, naive listeners classified NG vowels in terms of native AE categories and rated their goodness on a 7-point scale (very foreign to very English sounding). Both front- and back-rounded NG vowels were perceptually assimilated overwhelmingly to back AE categories and judged equally good exemplars. Perceptual assimilation patterns did not vary with context, and were not always predictable from acoustic similarity. These findings suggest that listeners adopt a context-independent strategy when judging the cross-language similarity of vowels produced and presented in continuous speech contexts. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1992688]

I. INTRODUCTION

While early research on cross-language speech perception using synthetic speech stimuli led to the conclusion that non-native vowel contrasts did not pose a perceptual problem for second language (L2) learners (Stevens *et al.*, 1969), more recent research with natural speech materials has documented that both inexperienced and experienced L2 listeners have perceptual difficulty with some, but not all, non-native vowel contrasts (e.g., Best, Faber, and Levitt, 1996; Flege, 1995; Gottfried, 1984; Levy and Strange, 2002; Polka, 1995; Yamada, Yamada, and Strange, 1995). Current theories of L2 speech perception (Best, 1995; Flege, 1995) invoke the concept of cross-language *phonetic similarity* in predicting relative difficulty listeners will have in perceptually distinguishing non-native contrasts.

In Best's Perceptual Assimilation Model (PAM), non-native phonetic segments are perceptually assimilated as more or less "good" exemplars of native categories by inexperienced listeners, unless they are so phonetically disparate from any native phone that they are "uncategorizable." Predictions of relative discrimination problems are then based on whether the contrasting non-native segments are both assimilated as equally good (or poor) exemplars of the same native catgory (single category pattern—poor discrimination), both are assimilated to a single native category, but with different perceived "goodness" (category goodness pattern—intermediate discrimination), or each is assimilated to a different native category (two category pattern—excellent discrimination). In Flege's L2 Speech Learning Model (SLM), the perceptual similarity of native and non-native segments is also characterized along a continuum from "identical" through "similar" to "new." According to this theory, L2 learners will assimilate similar non-native segments to native categories (via a perceptual equivalence categorization process) and fail to establish separate L2 perceptual representations; this will lead to persistent perception and production problems, even after years of L2 experience. In contrast, new non-native segments, which are very disparate from native categories, will eventually come to be represented as separate from L1 categories, and will therefore be perceived and produced more accurately than similar segments. Thus, cross-language perceptual similarity is used in both these models to predict initial and continuing perception (and production) difficulty.

In order for the concept of cross-language perceptual similarity as a predictor of relative discrimination accuracy

to be noncircular, it must be established using measures that are independent of non-native segment identification and discrimination performance. Recent studies have employed a *perceptual assimilation task* as a direct measure of perceived phonetic similarity. In this task, listeners are presented the non-native segments and are asked to categorize them in terms of their similarity to (mental representations of) native categories and rate their "category goodness" on a Likert scale (e.g., Bohn and Flege, 1990; Guion *et al.*, 2000; Harnsberger, 2000; Strange *et al.*, 1998; Strange *et al.*, 2001; Strange *et al.*, 2004). In the studies by Strange and her colleagues, perceptual assimilation patterns for non-native vowels produced in CVC(ə) (di)syllables have been shown to vary with both the prosodic context (citation-form lists vs sentences) in which non-native vowels are produced and presented (Strange *et al.*, 1998, 2004), and the immediate phonetic context (preceding and following stop consonants) in which the vowels are produced and presented (Strange *et al.*, 2001).

The present study investigated the perceptual assimilation of North German (NG) vowels by native American English (AE) listeners with no German-language experience. The primary concern was the extent to which acoustic and perceptual similarity varied as a function of the phonetic context in which the vowels were produced and presented. Both AE and NG vowels produced in stop consonant–vowel–stop consonant syllables have been shown to vary systematically in their spectral and temporal structure due to the coarticulatory influences of preceding and following consonants (Stevens and House, 1963; Steinlen, 2005; Steinlen and Bohn, 1999; Hillenbrand, Clark, and Nearey, 2001; Strange and Bohn, 1998; Strange *et al.*, 2002). To the extent that these coarticulatory variations in acoustic structure *differ* across languages, cross-language perceptual assimilation patterns might also vary as a function of the immediate phonetic context in which the vowels are produced and presented. That is, perceived similarity of AE and NG vowels may differ across contexts if the acoustic variation of NG vowels differs from the variation AE listeners expect on the basis of the native-language patterns. Previous research on Japanese listeners' perceptual assimilation of AE vowels showed that some AE vowels were actually assimilated to *different* Japanese vowel categories as a function of consonantal context (Strange *et al.*, 2001). Bohn and Steinlen (2003) also reported that some British English vowels were assimilated to different Danish categories as a function of consonantal context.

In the present study, 14 NG monophthongal vowels (7 long vowels /i, e, ɑ, o, u, y, ø/; 7 short vowels /ɪ, ɛ, a, ɔ, ʊ, ʏ, œ/) were produced in CVC syllables—bVp, bVt, dVt, gVk, gVt—which occurred medially in a short carrier sentence. Thus, the place-of-articulation of preceding and following consonants varied; however, since voicing contrasts are neutralized in final position in German, all final consonants were voiceless. It was assumed that the spectral characteristics of the medial vowels would be more variable than their temporal characteristics. This NG corpus was analyzed and compared with an analogous portion of an existing corpus of 11 AE vowels produced in similar CVC syllables embedded in a carrier sentence—bVb, bVp, dVd, dVt, gVg, gVk (Strange *et al.*, 2001). Thus, the phonetic contexts were not identical to those of the NG corpus, but both corpora contained vowels produced between labial, alveolar, and velar initial and final consonants. Since place-of-articulation differences in surrounding consonants appear to have greater effects on midsyllable formant frequencies of AE vowels than voicing or manner differences (Hillenbrand *et al.*, 2001), a comparison of these two corpora was considered valid. Based on this acoustical comparison of AE and NG vowels, predictions were made about AE listeners' perceptual assimilation of NG vowels to native categories. These predictions were tested by asking AE listeners with no German language experience to categorize the NG vowels in terms of their similarity to AE vowels and rate their "goodness" as exemplars of those AE categories.

II. ACOUSTIC SIMILARITY OF NG AND AE VOWELS

In a previous study (Strange *et al.*, 2004), NG vowels produced by the same speakers in hVp syllables in citation form (lists) and embedded in a carrier sentence revealed several interesting cross-language patterns of phonetic similarity. Of special interest was the relationship of the NG front-rounded vowels to NG and AE front-unrounded and back-rounded vowels. While the front-rounded vowels were spectrally more similar to front-unrounded than to back-rounded NG vowels, they were, in general, acoustically intermediate between front-unrounded and back-rounded AE vowels. This was due to the fact that AE so-called back vowels were acoustically "fronted" relative to NG back vowels (higher $F2$ and $F3$ values). A second finding was that NG mid- and mid-low vowels [e, o, ɛ, ɔ] were acoustically higher (lower $F1$ values) than their transcriptional AE counterparts in citation-form utterances. Cross-language discriminant analyses, in which the AE vowels served as the input corpus and NG vowels the test corpus, revealed that approximately half of the tokens were classified as more similar to AE high, midhigh, and mid vowels [i, u, ɪ, oʊ], respectively. For hVp syllables spoken in sentence contexts, this tendency was less marked, with the majority of NG tokens classified as more similar to their AE counterparts in the cross-language discriminant analysis. The present study was concerned with how these within- and across-language acoustic patterns varied with consonantal context. Specific questions were as follows:

(1) To what AE vowel categories are the NG front-rounded vowels acoustically most similar; how do the patterns of acoustic similarity vary with consonantal context? It was expected that front-rounded NG vowels in dVt context would be most similar to AE so-called back vowels, given that the latter are allophonically "fronted" to an even greater extent in this context than in others (Hillenbrand *et al.*, 2001; Strange *et al.*, 2002).

(2) Do back-rounded NG vowels show less variation across labial, alveolar, and velar contexts (i.e., less fronting) than AE back-rounded vowels? If so, then NG back-

rounded vowels would be more similar to their AE counterparts in labial (and perhaps velar) context than in alveolar contexts.

(3) To what AE vowel categories are NG mid- and mid-low vowels most similar, and do acoustic classification patterns vary with consonantal context? It was predicted that some tokens of the NG vowels might be classified as more similar to "higher" AE vowels in all contexts.

A. Speakers and stimulus materials

Four young adult male speakers of North German produced the stimulus corpus (the same speakers as were used in Strange *et al.*, 2004). The CVC syllables, embedded in the carrier sentence "Ich habe CVC gesagt," were printed on index cards in German orthography. Speakers were instructed to produce the sentences "as if they were speaking to a native speaker" and not to emphasize the target syllable. After a practice period, speakers produced three blocks of 15 utterances containing the 14 vowels (the 15th utterance was a repeat of the first, and omitted from the final corpus) in each of the five consonantal contexts; during recording, the consonantal context was fixed within blocks, but varied from block to block. The final corpus consisted of two instances of each vowel spoken in each consonantal context (usually from the second and third blocks) for a total of 140 stimuli for each speaker (2 repetitions × 5 contexts × 14 vowels). Stimuli were recorded on a DAT recorder, transferred to computer, and downsampled to 22.05 kHz for acoustical analysis and presentation to listeners in perceptual assimilation tests.

Four native speakers of North German from the same region as the speakers were asked to identify the vowels in the sentence utterances and rate their goodness on a 7-point Likert scale (7=best). Overall identification accuracy (collapsed over all speakers, contexts, and listeners) was 97% (93%–100% across vowels) and median goodness ratings were 6 or 7 for all 14 vowels. Thus, all instances of the NG vowels were highly identifiable and judged to be excellent exemplars of native vowel categories by speakers of the same dialect of German.

The AE comparison corpus was the same as used in previous studies (Strange *et al.*, 2001). Four young adult male native speakers of AE produced the CVC syllables embedded in the sentence ("I say the CVC on the tape") with instructions to speak as if you were talking to a native speaker of English. The corpus consisted of three tokens of each of 11 vowels produced in each of six contexts: b-b, b-p, d-d, d-t, g-g, g-k.

B. Acoustical analysis procedures

Analysis procedures were identical to those described in previous studies (Strange *et al.*, 1998, 2001, 2004). Temporal and spectral measurements were performed using SOUNDSCOPE/16 1.44 (ppc)™ speech analysis software designed by GW Instruments, Inc. (© 1992 GWI, Somerville, MA 02143). Using time-synchronized waveform and wideband spectrogram displays (300-Hz) filter, 6-dB pre-emphasis) with LPC formant tracks superimposed, vocalic

TABLE I. Average formant frequencies ($F1/F2/F3$) and vocalic durations of North German (NG) and American English (AE) vowels in CVC syllables (b-p, d-t, g-k contexts). NG Long/Short Vowel Ratio=1.6, AE Long/Short Vowel Ratio=1.4.

NG vowels	$F1$ (Hz)	$F2$ (Hz)	$F3$ (Hz)	Duration (ms)	AE vowels	$F1$ (Hz)	$F2$ (Hz)	$F3$ (Hz)	Duration (ms)
i	323	2063	3001	89	i	316	2250	2832	95
ɪ	395	1853	2509	61	ɪ	450	1811	2555	80
e	372	2039	2704	113	e	455	2034	2601	122
ɛ	530	1784	2457	82	ɛ	599	1668	2506	98
					æ	700	1654	2472	144
y	320	1609	2003	100					
ʏ	410	1396	2160	64					
ø	382	1370	2034	118					
œ	532	1388	2187	83					
u	334	733	2043	104	u	361	1318	2324	99
ʊ	433	928	2235	68	ʊ	479	1287	2443	88
o	401	740	2334	124	o	506	1125	2456	123
ɔ	571	1003	2279	79	ɔ	666	1087	2525	146
ɑ	710	1231	2408	149	ɑ	754	1323	2499	143
a	679	1335	2330	81	ʌ	629	1327	2480	96

duration was defined as the interval between the onset of voicing (first pitch period) and the beginning of consonant closure (indicated by a decrease in waveform amplitude and cessation of energy in the upper formants). Frequencies of the first three formants were measured from a 25-ms Hamming window centered at the 50% point of the vocal nucleus, using LPC analysis (28 coefficients, autocorrelation). When computed LPC peak values reflected spurious peaks or missed formants (based on an inspection of the spectrogram and formant history), manual placement of the marker on wideband FFT spectra or, in rare cases, narrow-band FFT, were used to estimate formant values.

C. Results

Table I presents the average formant frequencies and durations for the portion of the NG corpus (on the left) and the portion of the AE corpus (on the right) that were common across the two corpora: bVp, dVt, gVk. Since Hillenbrand *et al.* (2001) showed that initial consonants had greater effects than final consonants on AE vowel formant frequencies, it was considered appropriate to balance the number of tokens having labial, alveolar, and velar initial consonants in both corpora. Similar NG and AE vowels are aligned in the table to facilitate comparisons. Figure 1 presents the entire NG corpus plotted as points in $F1/F2$ bark[1] space; ellipses surround all 40 tokens of each vowel (4 speakers × 5 contexts × 2 repetitions). Ellipses were rotated by eye to include all tokens while encompassing the smallest area.

Looking first at within-language patterns for the full NG vowel corpus shown in Fig. 1, it appears that the front-rounded NG vowels were spectrally more similar to front-unrounded than back-rounded NG vowels, and that the back vowels were realized as back vowels in all contexts. There was some fronting of back vowels (higher $F2$ values) in d-t context (mean $F2$: [u]=794 Hz; [ʊ]=1098 Hz; [o] =768 Hz) relative to the same vowels in b-p context (mean

14 North German Vowels in 5 Consonantal Contexts
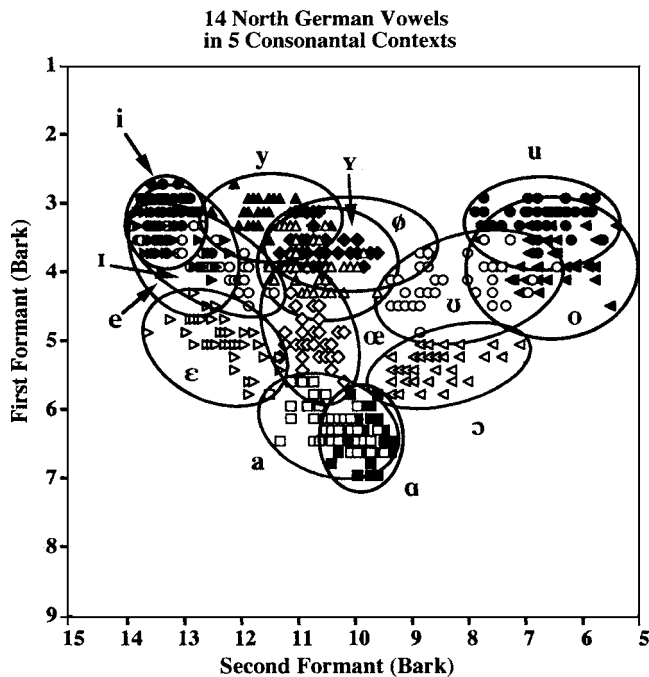
FIG. 1. Formant 1/formant 2 (bark) plots of North German vowels produced by four adult male speakers in b-p, b-t, d-t, g-t, and g-k contexts in a carrier sentence: 40 tokens/vowel. Long vowels are plotted as solids symbols; short vowels as open symbols.

$F2$: [u]=695 Hz; [ʊ]=830 Hz; [o]=742 Hz), especially for the short vowel [ʊ], with vowels in g-k context intermediate except for [o] (mean $F2$: [u]=711 Hz; [ʊ]=857 Hz; [o] =709 Hz). There was considerable spectral overlap among the high, midhigh, and mid vowels [i, ɪ, e], [y, ʏ, ø], [u, ʊ, o] and between the short and long low vowels [ɑ, a] as a function of speaker and context. Table I shows that these clusters of spectrally overlapping vowels tended to be differentiated by duration.

To quantify the spectral overlap of NG vowels and the extent to which duration acoustically differentiated spectrally overlapping spectral categories, two linear discriminant analyses were performed (Klecka, 1980). For the first, only syllable-midpoint $F1/F2/F3$ bark values were used as input parameters; all 40 instances of each vowel category (4 speakers×5 contexts×2 repetitions) were included as the input set (to establish parameter weightings and centers of gravity) and as the test set. Overall correct classification of vowels was 81% (50%–98% across individual vowels), with acoustic confusions occurring primarily between midlong and midhigh short vowel pairs [e/ɪ, ø/ʏ, o/ʊ] and low long/ short [ɑ/a]. Front-rounded vowels were spectrally differentiated from both front-unrounded and back-rounded vowels, with only 3 of 160 tokens (1 each for ʏ, ø, œ) classified as something other than a front-rounded vowel. A second discriminant analysis which included $F1/F2/F3$ bark values plus vocalic duration as input parameters yielded an overall correct classification rate of 94% (88%–100% across individual vowels). Most of the remaining acoustic ambiguities were confusions between high- and midlong vowel pairs [i/e, y/ø, u/o] and a few confusions of unrounded short vowels with rounded short vowels [ɪ/ʏ, a/œ, ɛ/œ] or short front-rounded vowels with each other [œ/ʏ]. That is, duration

served to differentiate spectrally overlapping midlong and midhigh short NG vowels [e/ɪ, ø/ʏ, o/ʊ] and the long/short low vowels [ɑ/a] across speakers and contexts.

Similar within-language, speaker-, and context-independent discriminant analyses were performed for the AE corpus, using all 72 tokens of each of 11 vowels as both input and test sets. The first analysis used $F1/F2/F3$ bark values as parameters. Overall correct classification was 82% (63% to 100% across individual vowels), with most errors on spectrally adjacent long/short vowel pairs [eɪ/ɪ, æ/ɛ, oʊ/ʊ] and on [ɑ/ɔ]. A second discriminant analysis using duration as well as formant bark values as input parameters yielded an overall correct categorization rate of 90% (81%–99% for individual vowels). Remaining acoustic ambiguities included long–short confusions, as well as a few between spectrally adjacent long vowel pairs and short vowel pairs. Because of the fronting of back-rounded vowels, there were also a few back–front confusions. Long–short confusions were due to smaller average duration differences (relative to NG vowels) for long/short contrasting pairs, and to speaker differences and variability in syllable durations due to the voicing contrasts in the final consonants. This reinforces the conclusion from previous research that vowel duration is a less robust acoustic cue for AE vowel contrasts than for NG contrasts.

Of greater interest is the cross-language comparison of spectral structure, which is depicted in Figs. 2(A) and 2(B). For clarity of presentation, the 7 long NG vowels (all 5 contexts) are presented in Fig. 2(A), superimposed on the ellipses for the 11 AE vowels (all 6 contexts). Individual AE vowel tokens are not plotted. Figure 2(B) superimposes the 7 short NG vowels over the same AE vowel ellipses. From these figures and the average data presented in Table I, it is readily apparent that the back-rounded vowels [u, ʊ, o/oʊ] differed considerably across languages, with NG vowels having much lower $F2$ values than their AE counterparts. In general, three of the *front*-rounded NG vowels [y, ʏ, ø] appeared to be more similar in $F2$ values to AE back high-and mid vowels, while [œ] appeared to be acoustically intermediate between front-and back AE vowels. Other cross-language differences are also apparent; as before, many tokens of the midlong NG [e, o] had lower $F1$ values than their AE counterparts, while midlow short NG vowels [ɛ, ɔ] appeared to overlap with their AE counterparts, but with some tokens of [ɛ] having lower $F1$ values.

A *cross-language* discriminant analysis was used to quantify these relationships between NG and AE vowels. This was accomplished by using the $F1/F2/F3$ bark values for the entire AE vowel corpus as the input set (to establish parameter weightings and centers of gravity for the 11 categories), and then assessing classification patterns for the entire NG vowel corpus, using Bark formant values as parameters for this test set. Therefore, this analysis established cross-language spectral similarities and differences, using the speaker- and context-*independent* domains of the AE vowels. Table II presents the NG classification data; the most frequent AE vowel classification for each NG vowel and the number of tokens so classified are given in the second and third columns, while the remaining AE classifications are listed in the fourth and fifth columns. The NG vowels are

A. North German Long Vowels

B. North German Short Vowels
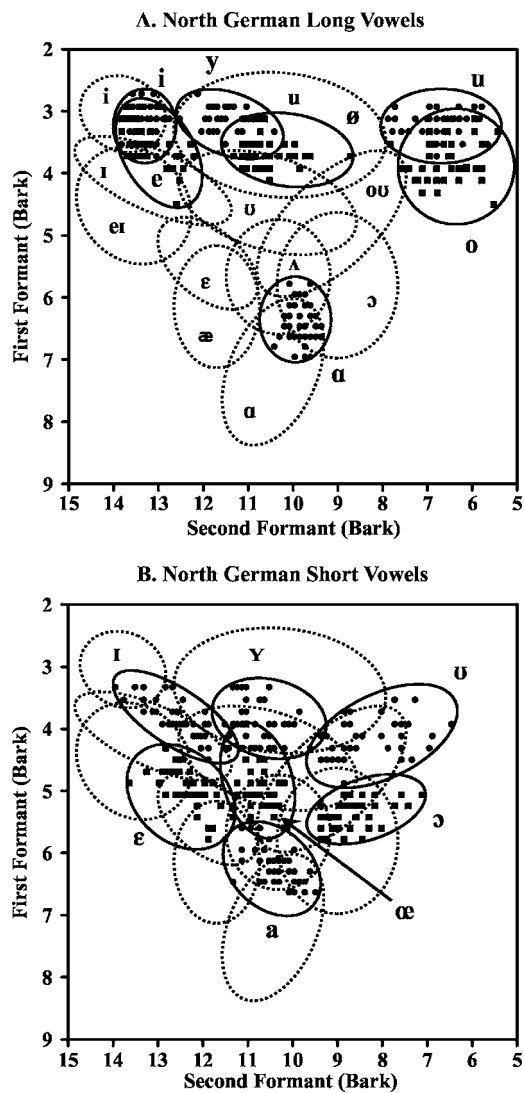
FIG. 2. Formant 1/formant 2 (bark) plots of NG long vowels (A) and short vowels (B) superimposed on ellipses indicating American English vowel distributions from productions of four adult male speakers in b-b, b-p, d-d, d-t, g-g, and g-k contexts in a carrier sentence (56 tokens/vowel). Individual AE tokens are not plotted.

TABLE II. Spectral similarity ($F1/F2/F3$ in barks) of NG vowels (five contexts) and AE vowels (six contexts): Context-independent discriminant analysis.

	Modal classification			Other vowels	
	NG Vowel	AE Vowel	# Stimuli	AE Vowel	# Stimuli
Point vowels	i	i	39	ɪ	1
	ɑ	ɑ	24	ʌ	9
				ɔ	7
	u	u	39	oʊ	1
Front rounded	y	u	39	i	1
	ø	u	40		
	ʏ	u	29	ʊ	10
				ɪ	1
	œ	ʊ	27	ʌ	9
				u, ɛ	2 each
Midhigh, mid, and midlow	ɪ	ɪ	33	i	6
				ʊ	1
	e	i	24	ɪ	13
				eɪ	3
	ɛ	ɛ	25	ɪ	9
				eɪ	6
	ʊ	oʊ	25	u	13
				ɪ	2
	o	oʊ	29	u	11
	ɔ	oʊ	33	ɔ	7
Low short	a	ʌ	24	ɑ	13
				ɔ, ɛ, æ	1 each

divided into four groups: (a) the point vowels; (b) the front-rounded vowels; (c) the midhigh, mid-, and midlow vowels, and (d) the low short vowel.

Looking first at the point vowels [i, ɑ, u], it can be seen that NG [i, u] were classified almost unanimously as most similar spectrally to their AE counterparts, while NG [ɑ] tokens were assimilated to the spectrally overlapping AE categories [ɑ, ɔ, ʌ]. Given the acoustic similarity across AE and NG speakers in the highest front [i] and lowest back [ɑ] vowels, we speculated that there were no great differences in the acoustic vowel spaces attributable to speaker differences across language groups.[22] Thus, classification of other NG vowels as AE vowels other than their transcriptional counterparts was interpreted as reflecting cross-language differences in their relative locations in vowel space. While NG [u] was classified as most similar to AE [u], Fig. 2(A) illustrates that the NG tokens differed markedly from their AE counterparts (nonoverlapping distributions of $F2$ values);

however, they were not more similar to any *other* AE vowel category.

In this context-independent analysis, the four NG front-rounded vowels were classified as spectrally most similar to AE back-rounded vowels (97% of the tokens). This was due to the allophonic fronting of AE so-called back-rounded vowels in alveolar and (to a lesser extent) velar contexts, which shifted the centers of gravity for AE so-called back vowels forward in vowel space. This result differs markedly from earlier results (Strange *et al.*, 2004) comparing NG and AE vowels produced in hVp/hVb(a) (di)syllables. In this noncoronal context, NG front-rounded vowels were acoustically intermediate between front and back AE vowels. In citation-form context, only 37% were classified as back AE vowels; in sentence context, 53% were classified as back AE vowels.

Turning to the midhigh, mid-, and midlow NG vowels, the discriminant analysis revealed considerable variability in cross-language spectral similarity (from 83% "correct" classification to AE transcriptional counterparts for [ɪ] to only 5% for [ʊ]). The majority of tokens of NG [ɪ, ɛ, o] was classified as most similar to their AE counterparts (83%, 63%, 73%, respectively), while the majority of tokens of NG [e, ʊ, ɔ] was classified as similar to different AE vowels ([i, oʊ, oʊ], respectively). There was an overall tendency for these vowels to be classified as similar to higher AE vowels (lower $F1$ values), except for [ʊ]. These results generally replicate those reported earlier for NG vowels produced in hVp context. Finally, the low short NG [a] was acoustically

TABLE III. Spectral similarity (from context-dependent discriminant analyses) of NG front-rounded vowels and AE back vowels.

NG front vowel	AE back vowels	b-p % of tokens	d-t % of tokens	g-k % of tokens
y	u	0%	100%	75%
ø	u, oʊ	87	100	100
ʏ	u, ʊ	75	100	100
œ	ʊ, ʌ, u	13	100	100
All 4 NG vowels	AE Back vowels	48%	100%	94%

most similar to the AE short [ʌ] (60% of the tokens), although other tokens were classified as more similar to AE long vowels [ɑ, ɔ] (35%) or to front AE vowels [ɛ, æ] (5%).

In summary, cross-language context-independent acoustic comparisons of AE and NG vowels produced in labial, alveolar, and velar contexts revealed that the distributions of four NG vowels [i, u, y, ø] were close matches to the distributions of a particular AE vowel, or at least did not overlap with a *different* AE vowel category. Note that NG [u, y, ø] were acoustically most similar to the same AE vowel [u]. The other ten NG vowels were produced with distributions of spectral values which spanned at least two AE vowel distributions. For NG [ɑ, a] this was because the AE [ɑ, ɔ, ʌ] distributions overlapped spectrally. Similarly, classification of NG [ʊ] to AE [oʊ] and NG [e] to AE [ɪ] was due in part to the spectral overlap of AE [oʊ, ʊ] and AE [eɪ, ɪ], respectively. For the remaining NG vowels, [o, ɪ, ɛ, ɔ, œ, ʏ], it was because the NG vowel distribution spanned (at least) two relatively distinct AE vowel distributions.

In addition to this context-independent cross-language analysis, three additional cross-language context-*dependent* discriminant analyses were performed in which AE vowels in each of three contexts (bVp, dVt, gVk) served as separate input sets (pooled over all four speakers) and NG vowels in the same three contexts (also pooled over all four speakers) were used as the test sets, respectively. While these analyses employ small numbers of tokens of each vowel (12 and 8, respectively), they are instructive in interpreting the variability in cross-language categorization of NG vowels in the

context-independent analysis described above.

Table III gives the results for the classification of NG front-rounded vowels to back AE vowel categories. Classification percentages are summed over back AE vowels with categories listed in column 2. In b-p context, NG [y] was classified unanimously as front AE vowels [i, eɪ], and [œ] was also classified most often as front vowels [ɛ, ɪ]. In contrast, the majority of tokens of NG [ø, ʏ] was more similar to AE back vowels. In g-k context, only two tokens of [y] were classified as front vowels [i, ɪ], while in d-t context, all NG front-rounded vowel tokens were spectrally more similar to AE back vowels. The similarity pattern for vowels in labial context generally replicated that reported for vowels in hVp context in the earlier study (Strange *et al.*, 2004). These contextual differences in cross-language similarity patterns are attributable in large part to contextual variation of AE back-rounded vowels, i.e., to fronting (raising of *F*2) in alveolar and velar contexts.

Table IV shows context-dependent cross-language spectral similarity patterns for the remaining ten NG vowels. First, it can be seen that NG [u] showed very large variations in spectral similarity to AE [u] across contexts. Whereas all eight tokens in b-p context were classified as most similar to AE [u], all eight tokens in d-t context and all but two tokens in g-k context were acoustically more similar to AE [oʊ]. This was due to the fact that *F*2 values for AE [u] were higher than for AE [oʊ] in these contexts. Three other NG vowels produced context-specific differences in modal classification patterns; for [ɪ, ɛ, ɔ], acoustic similarity to AE counterparts was greatest in the d-t context. Of the remaining five vowels, NG [e, ʊ] were spectrally more similar to different AE vowels [i, oʊ], respectively, in all three contexts; NG [i, o, ɑ] were quite consistently classified as most similar to their AE counterparts, or were classified into two or more spectrally overlapping AE vowels in all contexts. Finally, NG [a] was variably classified as spectrally overlapping [ɑ, ʌ] except in g-k context, where some tokens were classified as front vowels.

In summary, context-dependent discriminant analyses indicated that several NG vowels differed in spectral similarity to AE vowels as a function of preceding and following

TABLE IV. Spectral similarity (context-dependent discriminant analyses) of NG vowels and their AE counterparts.

NG vowel	AE vowel counterpart	bVp % tokens	(Other vowels)	dVt % tokens	(Other vowels)	gVk % tokens	(Other vowels)
A. Point vowels							
i	i	100		87	(eɪ)	100	
ɑ	ɑ	50	(ʌ)	75	(ɔ)	50	(ʌ)
u	u	100		0	(oʊ)	25	(oʊ)
B. Midhigh, mid, midlow vowels, and short low vowel							
ɪ	ɪ	63	(eɪ)	87	(eɪ)	38	(i, eɪ)
e	eɪ	38	(i, ɪ)	38	(ɪ, i)	0	(i, ɪ)
ɛ	ɛ	50	(ɪ)	63	(ɪ)	25	(ɪ, eɪ)
ʊ	ʊ	0	(oʊ, u)	0	(oʊ)	0	(oʊ)
o	oʊ	87	(u)	87	(ɔ)	100	
ɔ	ɔ	0	(oʊ)	63	(oʊ)	13	(oʊ)
a	ʌ	50	(ɑ)	38	(ɑ)	75	(ɛ, æ)

consonantal context. The front-rounded NG vowels were less similar to back-rounded vowels in labial contexts, while the back-rounded NG [u] was more similar to the AE back vowel in that context. The midhigh and midlow vowels were acoustically more similar to their AE counterparts in alveolar context than in the other contexts.

III. PERCEPTUAL ASSIMILATION OF NG VOWELS BY AE LISTENERS

In a previous study (Strange *et al.*, 2004), native AE listeners' perceptual assimilation of NG vowels spoken in hVp syllables differed considerably from the patterns predicted on the basis of spectral similarity. In particular, NG front-rounded vowels were perceived as more similar to back-rounded AE vowels (categorized as back vowels on 77% of trials for isolated syllables, 96% for syllables in sentence context). Thus, even in this noncoronal context, NG front-rounded vowels were assimilated to AE back vowel categories. However, AE listeners judged three of them to be relatively poor exemplars of back vowel categories on a 7-point Likert scale (1=very foreign-sounding; 7=very English-sounding); median ratings for [y, ʏ ø, œ] in syllables were 2, 1, 3, 5, respectively; in sentence context, 2, 3, 2, 4, respectively. In contrast, NG high-and midback rounded vowels [u, o] were rated as somewhat better exemplars of back-rounded AE vowels, especially in citation-form context (median ratings=4, 5, respectively, in syllable context; 3, 5, respectively, for sentence context). We interpreted these results as indicating that both front- and back-rounded NG vowels were perceptually assimilated as variants of back AE vowels, due to the contextual variability of these native categories which subsumes the front rounded portion of vowel space. However, listeners were aware that the front-rounded vowels were phonetically less appropriate than were the back vowels in noncoronal context. That is, listeners were able to make context-dependent judgments about these non-native vowels as reflected by the differences in goodness ratings, especially when syllables were produced and presented in isolation.

Second, some mid- and midlow vowels were assimilated to their AE counterparts more consistently than was predicted from their cross-language acoustic similarity patterns, especially in the citation-form syllable condition. For example, midlong NG [e, o] were both acoustically intermediate between mid [eɪ, oʊ] and high [i, u] AE vowels. However, while NG [e] was perceptually assimilated more often to AE [i] (66% responses) than AE [eɪ] (23%), NG [o] was assimilated quite consistently as most similar to AE [oʊ] (89%). Many tokens of NG [ɛ, ɔ] were classified as acoustically more similar to higher AE vowels [ɪ, oʊ], respectively, but perceptually, were consistently judged to be relatively good tokens of their AE counterparts [ɛ, ɔ] (97% and 90% consistency; median ratings 6 and 5, respectively).

Given these discrepancies between acoustic and perceptual similarity of NG and AE vowels found in the previous study, it was important to directly test perceptual similarity patterns for the present corpus in which NG vowels were produced and presented in CVC syllables in sentence context. In this experiment, both consonants and vowels varied

randomly from trial to trial, while the speaker was held constant across trials. This begins to approach the situation confronted by L2 learners when listening to foreign utterances of particular speakers. An overall question of interest was whether perceptual assimilation patterns would reflect context-dependent or context-independent acoustic similarity patterns. That is, would categorization consistency and/or goodness ratings differ across contexts for those vowels which showed systematic contextual variations in cross-language spectral similarity? If listeners were able to make context-dependent judgments, we predicted the following patterns of contextual variation:

(1) Front-rounded NG vowels (especially [y, œ]) will be perceived as less similar to back-rounded AE vowels in labial context than in velar and alveolar contexts.
(2) Back-rounded NG vowels (especially [u]) will be considered poorer exemplars of back AE vowels in alveolar and velar contexts than in labial context.
(3) NG [ɪ, ɛ, ɔ] will be perceived as better exemplars of their AE counterparts in alveolar context than in labial or velar contexts.
   Finally, based on spectral similarity patterns, we expected that:
(4) NG [e, ʊ] will be perceptually assimilated to spectrally more similar AE [i, oʊ], respectively, than to their AE counterparts in all contexts.

A. Listeners and procedures

Because of the very large number of test trials required, a mixed design was employed, with consonantal context as a repeated-measures variable and speakers as a between-subjects variable. Forty-eight native English-speaking undergraduate and graduate students from Communication Sciences & Disorders who had completed one semester course in basic phonetics and IPA transcription served as subjects; 12 subjects were assigned randomly to each NG speaker condition. None had had any German language experience.

Categorization and rating responses were obtained using an interactive HYPERCARD program that displayed the 11 AE vowel category choices represented by key words *heed, hid, hayed, head, had, hud, hod, hawed, hoed, hood, who'd*. These response alternatives appeared on the computer screen with the IPA symbol for each vowel category listed above and a response "button" beside each word. Both stimulus presentation and response collection were controlled by computer. Thus, subjects could proceed at their own pace. Each trial consisted of two presentations of the same stimulus; after the first presentation, subjects selected one of 11 response alternatives. The second presentation occurred immediately afterward and the 7-point rating scale appeared on the computer screen; subjects rated the goodness of the NG vowel as an exemplar of the selected AE category. Instructions were to click on the 1 if the vowel was "very foreign-sounding;" 7 if the vowel was "very English-sounding," and to choose an appropriate number between 1 and 7 if the vowel was somewhere in between. Subjects were told that they could change their identification response after this second presentation, but were discouraged from doing so. (After

TABLE V. Perceptual assimilation patterns. Overall modal responses (pooled over listeners, speakers and contexts) in columns 2–4; modal responses in b-p (columns 5–6), d-t (columns 7–8), and g-k (columns 9–10) syllable contexts (pooled over listeners and speakers).

		Overall modal responses			Context-specific modal responses				
		All five contexts		b-p context		d-t context		g-k context	
NG	AE	%	(Mdn)	%	(Mdn)	%	(Mdn)	%	(Mdn)
i	i	99	(7)	99	(7)	99	(7)	99	(7)
ɑ	ɑ	75	(5)	81	(5)	79	(5)	65	(5)
u	u	88	(5)	90	(5)	92	(5)	82	(5)
y	u	93	(5)	93	(4)	95	(5)	89	(5)
ʏ	ʊ	74	(4)	68	(4)	72	(4)	73	(4)
ø	u	43	(5)	47	(4)	42	(5)	41	(5)
œ	ʌ	74	(4)	77	(4)	75	(4)	71	(4)
ɪ	ɪ	84	(6)	98	(6)	94	(6)	52	(5)
e	e	64	(6)	63	(5)	60	(6)	60	(5)
ɛ	ɛ	92	(6)	98	(6)	97	(6)	70	(6)
ʊ	ʊ	45	(4)	38	(3)	36	(3.5)	31	(4)
o	oʊ	69	(5)	65	(5)	72	(5)	68	(5)
ɔ	ɔ	53	(4)	36	(4)	48	(4)	72	(4)
a	ɑ	55	(5)	69	(5)	64	(5)	46	(4)

familiarization trials, very few changes were made by any subject.)

Prior to testing on NG vowels, subjects completed task and response alternative familiarization procedures using AE vowel stimuli presented in CVC syllables in a carrier sentence. Listeners completed 55 trials (5 blocks of 11 AE vowels each) and were given feedback following each of the first three blocks. The last two blocks were used as an assessment of task mastery. Listeners who misidentified more than three AE vowel tokens or misidentified both presentations of any one vowel (except for [ɑ/ɔ] confusions) were not included in the study.

For NG vowel tests, each listener completed 4 blocks of 140 trials in which consonantal context and vowels were sequenced randomly (4 different randomizations). Thus, each listener completed a total of 8 judgments of each token, 16 judgments on each NG vowel in each context. Testing was completed in two sessions scheduled on separate days within a few days of each other.

B. Results

For each vowel in each consonantal context, categorization responses were tallied (summing across trials and listeners/speakers) and median ratings for each response alternative were computed. The modal response alternative was determined for each vowel in each consonantal context; in addition, an overall modal response category was determined. Table V presents these data, with the overall modal response category (column 2), the overall response consistency (expressed as a percentage of total opportunities, column 3), and the overall median rating (column 4) for each NG vowel. In the right-hand columns are the data for three of the five contexts: b-p (columns 5–6), d-t (columns 7–8), g-k (columns 9–10) for which acoustic similarity data were reported. The NG vowels are clustered in four groups, with the

three point vowels first, the four front-rounded vowels next, the remaining midhigh, mid-, and midlow vowels third, and the low short vowel [a] last.

Looking first at the overall modal assimilation data, it is immediately apparent that some NG vowels were judged to be highly similar to particular AE vowels, as indicated by high levels of categorization consistency across contexts, speakers, and individual listeners (with relatively high median goodness ratings in some cases), while other vowels showed highly inconsistent categorization patterns (and lower median goodness ratings in some cases). Thus, NG vowels could be said to range from perceptually excellent exemplars of native AE categories (e.g., [i]) to uncategorizable as any AE vowel (e.g., [ø, ʊ]). Looking first at the point vowels (upper cluster), NG [i] stands out as most similar to its AE counterpart, with [u] somewhat less consistently categorized, especially in velar contexts, and rated as a poorer exemplar. NG [ɑ] was most often assimilated to its AE counterpart, but also categorized as [ɔ] (24% of trials). This inconsistency is due to the fact that many of the listeners did not differentiate these two AE vowels in their own dialect, as found in previous studies. When [ɑ/ɔ] responses were collapsed into a single response category, response consistency for NG [ɑ] (99%) was the same as for NG [i]. However, listeners' goodness ratings reflected that this vowel was not as good an exemplar of the native category in any of the five contexts (median ratings for all five contexts=5).

The front-rounded NG vowels were overwhelmingly categorized as more similar to back AE vowels (97%, collapsing over all back vowel response alternatives). There was inconsistency in the selection of the particular AE back vowel category for all but [y], which was consistently categorized as most similar to AE [u] in all contexts (range =89%–95% across the five contexts). Summing over all contexts, NG [ʏ] was assimilated to AE [ʊ] (74%), [ʌ] (13%), or [u] (11%); NG [ø] was assimilated to AE [u]

(43%), [ʊ] (33%), or [oʊ] (21%), and NG [œ] was assimilated to AE [ʌ] (74%) or [ʊ] (19%). (Response alternatives receiving fewer than 10% of total responses are not given—thus, percentages do not add to 100%.) NG [ø] was not assimilated to any one AE vowel on a majority of trials in any of the five contexts (categorization consistency range =36%–50%). This vowel can thus be considered as uncategorizable in terms of native categories. NG [ʏ, œ] were categorized more consistently to AE [ʊ, ʌ], respectively (range across five contexts=68%–83%; 69%–78%, respectively).

The context-dependent analyses indicated no marked differences across the five contexts either in categorization consistency or in rated goodness for the front-rounded vowels. Median ratings (4–5) suggested that these vowels were heard as neither very foreign sounding nor very native sounding. (Since median ratings for minority response selections ranged as low as 2, we can assume that, in general, listeners were using almost the full range of goodness ratings in making their judgments.) In summary, then, perceptual assimilation of all four front-rounded NG vowels suggested that listeners were responding on the basis of a context-independent mapping of these non-native vowels onto native categories. The four vowels differed in their fit to particular AE vowels, with [y] being consistently assimilated to AE [u], [ʏ, œ] being assimilated less consistently to [ʊ, ʌ], respectively, and [ø] being uncategorizable as any particular back AE vowel.

Categorization and goodness judgments for NG back-rounded long vowels also showed context-independent patterns of perceptual assimilation to AE back vowels. As described above, NG [u] was considered a fair exemplar of its AE counterpart in all five contexts. Note that perceived similarity of high front- and back-rounded NG vowels [y, u] were very similar in categorization and rating pattern. NG [o] was most often assimilated to its AE counterpart (65%–72% across contexts), but was sometimes heard as more similar to AE [u] in all five contexts (21% [u] responses overall; median rating=5 for all five contexts). This corresponds to the context-independent spectral similarity patterns for this vowel, which indicated that it was located higher in vowel space relative to AE [oʊ], and close to NG [u] (see Fig. 2).

AE listeners' perceptual assimilation of NG [ɪ, ɛ, ɔ] to their AE counterparts reveals somewhat larger differences across consonantal contexts in categorization consistency. In fact, the modal response category actually changed across contexts for each of these vowels. NG [ɪ, ɛ] were less consistently categorized as AE [ɪ, ɛ] in g-k context (and also g-t context for [ɪ]), than in labial and alveolar contexts. For NG [ɔ], contextual variation consisted of more or less categorizations to AE [ɑ]; when AE [ɑ/ɔ] responses were collapsed, categorization consistency did not differ markedly across contexts (range=68%–85%; median ratings=4–5 across contexts) and in no context was this vowel categorized as any other AE vowel a majority of the time. Thus, while NG [ɔ] was spectrally more similar to AE [oʊ] in labial and velar contexts, it was perceptually assimilated to that vowel only 4% of the time.

Turning to the two NG vowels [e, ʊ], which were predicted, on the basis of spectral similarity, to be perceptually assimilated to AE vowels other than their counterparts, we see that, indeed, these vowels were among the least consistently categorized to their AE counterparts in all contexts ([e]=60%–71% across five contexts, median ratings=5–6 across contexts; [ʊ]=32%–61%; median ratings=3–4 across contexts). Collapsing across contexts, NG [e] was assimilated to AE [i] on 31% of trials, while NG [ʊ] was assimilated to AE [oʊ] (32%), or [ʌ] (12%). There were variations in assimilation pattern for NG [ʊ] as a function of consonantal context; categorization as AE [ʊ] was more consistent in b-t and g-t contexts (59%, 61%, respectively) than in the contexts listed in Table V. However, NG [ʊ] can generally be considered uncategorizable as any one AE vowel and received the poorest goodness ratings. NG [e] was perceptually ambiguous, spanning the AE categories [i, eɪ], but was judged to be relatively "native sounding" even when categorized as more similar to AE [i].

To summarize the perceptual assimilation data, NG vowels varied in their perceptual similarity to particular AE vowels. The point vowels were assimilated to their AE counterparts quite consistently in all consonantal contexts, with [i] considered the best fit to a native category, while [ɑ, u] showed slightly more variability within and across contexts in categorization consistency and were considered less good exemplars of native categories. All four front-rounded NG vowels were assimilated as relatively poor to fair exemplars of back AE vowels. While [y] was consistently assimilated to AE [u], NG [ʏ, œ] were less consistently heard as [ʊ, ʌ], respectively, and NG [ø] was uncategorizable as any one AE back vowel. NG [a] was assimilated as a relatively poor to fair exemplar of either AE [ɑ/ɔ] or [ʌ]. Of the remaining NG vowels which have transcriptional equivalents in AE, NG [ɪ] was consistently assimilated as a good exemplar of its counterpart except in g-t and g-k context, and NG [ɛ] was consistently assimilated as a good exemplar of its AE counterpart except in g-k context. NG [ɔ] was perceived as intermediate between AE [ɑ/ɔ] and [oʊ], while NG [e, o] were perceived as intermediate between their counterparts and high AE vowels [i, u], respectively, reflecting the fact that these three NG vowels were located higher in the acoustic vowel space, relative to the point vowels of NG and their AE counterparts. Finally, NG [ʊ] was uncategorizable as any one AE vowel and received the poorest overall goodness judgments of all the NG vowels.

In general, we can conclude that listeners' perceptual assimilation performance did not reflect much consideration of context-dependent variations in spectral similarity. That is, when listeners had to judge vowels embedded in CVC syllables in sentence-length utterances with consonantal context varying from trial to trial, they adopted a context-independent strategy in selecting which AE vowels the non-native vowels were most like and in making judgments about their category goodness. Perceptual assimilation of three vowels [ɪ, ɛ, ʊ] varied systematically with context such that modal categories actually changed across contexts for some speakers' utterances. However, contextual variations in perceived similarity predicted from differences in spectral similarity patterns were generally not borne out.

IV. GENERAL DISCUSSION

In comparing the context-*independent* discriminant analysis results for NG vowels (Table II) and the perceptual assimilation patterns shown by native AE listeners (Table V), it can be seen that spectral similarity predicted the perceptual assimilation of front-rounded NG vowels to back-rounded AE vowels. In fact, the judged goodness of the front-rounded vowels and back-rounded NG vowels as exemplars of back AE vowels were equivalent. Neither set was considered an excellent fit to a native category in any of the contexts. We would expect, therefore, that perceptual differentiation of front- and back-rounded NG vowels would be highly problematic for AE listeners when the vowels were produced and presented in varying consonantal contexts in sentence length utterances. That is, these vowel contrasts constitute single-category assimilation patterns according to Best's classification scheme (Best, 1995).

In the previous study (Strange *et al.*, 2004), when vowels were produced and presented in a single consonantal context in sentences, subjects assimilated NG front/back-rounded vowel contrasts in a category goodness pattern, i.e., back-rounded NG vowels were better exemplars of back-rounded AE vowels than were front-rounded NG vowels. It is not clear whether the differences across studies were due to differences in task demands (variable vs fixed consonantal context), differences in vowel productions, or individual differences in listeners' strategies. Further research is needed in which vowels produced in labial, alveolar, and velar contexts in sentences are presented using a blocked design in which the context is fixed from trial to trial. If subjects can adopt a context-dependent strategy in this case, then we would predict that front- and back-rounded vowel contrasts in labial contexts might show a category-goodness pattern (with NG front-rounded vowels poorer exemplars of AE back vowels) or even a two-category pattern (front-rounded vowels assimilated as front AE vowels—back-rounded vowels assimilated as back AE vowels), while in alveolar and velar contexts, the contrasts might produce single-category assimilation patterns. These differences in perceptual similarity across studies in which the task and materials differed might explain the discrepancies in studies of discrimination of front- vs back-rounded vowels in German and French by AE listeners (Gottfried, 1984; Levy and Strange, 2002; Polka and Bohn, 1996; Rochet, 1995).

As in the previous study (Strange *et al.*, 2004), several of the midhigh, mid-, and midlow NG vowels, which are traditionally transcribed as identical or similar to AE vowels, were produced relatively higher in the vowel space than their AE counterparts. These cross-language differences in phonetic realization were reflected in perceptual assimilation patterns as well. Thus, we can conclude that some NG vowel contrasts which are also phonologically distinctive in AE will nevertheless be problematic for AE-speaking learners of German. Confusions between high and midlong vowels [e/i], [u/o] should be expected. Gottfried (1984) reported that AE listeners had difficulty discriminating French [e/i]. The fact that NG (and French) [e, o] are monophthongs compounds

the problem of perceptual differentiation from high vowels for AE speakers, and may also lead to accented production of these vowels.

In contrast, the acoustic raising of the midhigh front and midlow short vowels [ɪ, ɛ, ɔ] did not appear to affect their perceptual assimilation to their AE counterparts, except perhaps in velar contexts. Thus, we would not expect perceptual problems in differentiating NG [ɪ/ɛ] or [ɔ/o]. However, the fact that these NG vowels are perceptually assimilated to native vowel categories which are phonetically realized as lower in vowel space might lead to accented productions of [ɪ, ɛ, ɔ] by AE speaking L2 learners of German (Flege, 1995). In addition, NG [ɔ] is produced as a short vowel, whereas it is phonetically long in AE.

The NG contrast between low vowels [ɑ/a], differentiated primarily by duration, might also lead to some perception (and production) problems for AE learners of German, since both vowels were assimilated to long AE vowels in a category goodness pattern. Explicit training in listening for (and producing) the duration contrast might help learners to differentiate these vowels. Finally, two NG vowels [ʊ, ø] were uncategorizable for AE listeners. They are thus predicted to be relatively easily discriminated from categorizable NG vowels [o, u], respectively, according to Best's PAM. According to Flege's SLM, these vowels would also come to be differentiated perceptually and productively with German language experience, since they would not be assimilated into existing L1 categories.

Future studies are needed to determine the relationship between perceptual assimilation patterns and discrimination of non-native vowels contrasts by inexperienced and more experienced L2 learners of German. In these studies, it is important that perceptual performance be assessed using materials in which vowels are produced and presented in multiple consonantal contexts and contain the acoustic variability normally encountered in language learning situations (e.g., Bohn and Steinlen, 2003). However, previous research on French vowels (Levy and Strange, 2002), the results of Strange *et al.* (2004), and the present study for German vowels suggest that presenting vowels in multiple consonantal contexts in random sequences may not optimize perceptual differentiation by beginning learners. That is, if listeners are forced to adopt a context-independent strategy of processing non-native vowels, they might confuse exemplars of contrasting non-native vowels which are acoustically quite disparate. Thus, in training studies, it might be efficacious to introduce vowel contrasts in fixed consonantal contexts (at least within blocks of training trials) starting with those contexts which, at a context-dependent phonetic level of processing, might be perceived as category goodness contrasts (see Jamieson and Morosan, 1986; 1989 for a similar strategy for optimizing training of consonantal contrasts). Data on the cross-language acoustic and perceptual similarity of vowels such as those presented here will help in the design of such training materials.

V. SUMMARY

The results of this study provide some answers to questions about the acoustic and perceptual similarity of NG and

AE vowels

(1) Analysis of the acoustic similarity of NG and AE vowels showed that NG vowels range from excellent cross-language matches to their AE counterparts to vowels which span two spectrally distinct AE vowel categories.

(2) NG front-rounded vowels vary in their spectral similarity to back-rounded AE vowels as a function of consonantal context. This variation is due primarily to the large allophonic variation in the production of AE back-rounded vowels (i.e., fronting in alveolar, and to a lesser extent, velar contexts). NG back-rounded [u] also differed in spectral similarity to its AE counterpart, again, due to the large allophonic variation in the AE vowel. In addition, three other NG vowels [ɪ, ɛ, ɔ] varied in their spectral similarity to AE counterparts as a function of consonantal context.

(3) In general, there were no striking context-dependent variations in perceptual assimilation of NG vowels to AE vowels by listeners when the consonantal context in which the vowels were embedded varied randomly from trial to trial. That is, in this task, listeners appeared to adopt a context-independent strategy for judging the perceptual similarity of NG vowels to native categories.

(4) Patterns of cross-language spectral similarity computed using context-independent parameter weightings and centers of gravity for AE vowel distributions predicted patterns of perceptual assimilation of front rounded NG vowels. Discriminant analyses also verified that some midhigh, mid, and midlow NG vowels tended to be produced as "higher" vowels (lower $F1$ values) than comparable AE vowels. For some of these vowels, perceptual assimilation patterns also reflected these cross-language differences in phonetic realization. However, for other NG vowels, perceptual assimilation to AE counterparts was better than was predicted from spectral similarity patterns. Thus, it is necessary to assess perceptual similarity directly.

(5) Together, cross-language comparisons of the acoustic and perceptual similarity of NG and AE vowels can be used to predict the relative difficulty that AE speakers may have in learning to perceive (and produce) NG vowel contrasts. Structured training materials in which the contrasts are presented in contexts graded from easier to more difficult might optimize performance and lead to robust L2 phonetic category formation.

ACKNOWLEDGMENTS

[1] The formula used to compute barks was as follows: $13 \times \arctan(0.76 \times \text{Hz}/1000) + 3.5 \times \arctan[(\text{Hz}/1000/7.5)^2]$.

[2] Strange *et al.* (2004), using the same four speakers, also reported similarity in the point vowels that indicated speaker normalization across languages was not a problem.

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, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (**1996**). "Assimilation of non-native vowel contrasts to the American English vowel system," J. Acoust. Soc. Am. **99**, 2602.

Bohn, O.-S., and Flege, J. E. (**1990**). "Interlingual identification and the role of foreign language experience in L2 vowel perception," Appl. Psycholing. **11**, 303–328.

Bohn, O.-S., and Steinlen, A. K. (**2003**). "Consonantal context affects cross-language perception of vowels," Proceedings of the 15th International Congress of Phonetic Sciences, 2289–2292.

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, Timonium, MD), pp. 233–277.

Gottfried, T. L. (**1984**). "Effects of consonantal context on the perception of French vowels," J. Phonetics **12**, 91–114.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: the case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

Harnsberger, J. D. (**2000**). "A cross-language study of the identification of non-native nasal consonants varying in place of articulation," J. Acoust. Soc. Am. **108**, 764–783.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Jamieson, D. G., and Morosan, D. E. (**1986**). "Training non-native speech contrasts in adults: Acquisition of the English /∂/ - /θ/ contrast by francophones" Percept. Psychophys. **40**, 205–215.

Jamieson, D. G., and Morosan, D. E. (**1989**). "Training new, non-native speech contrasts: A comparison of the prototype and perceptual fading techniques." Can. J. Psychol. **43**, 88–96.

Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage, Newbury Park, CA).

Levy, E., and Strange, W. (**2002**). "Effects of consonantal context on perception of French rounded vowels by American English adults with and without French language experience," J. Acoust. Soc. Am. **111**, 2361.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Polka, L., and Bohn, O.-S. (**1996**). "A cross-language comparison of vowel perception in English-learning and German-learning infants," J. Acoust. Soc. Am. **100**, 577–592.

Rochet, B. L. (**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, Timonium, MD), pp. 379–410.

Steinlen, A. K. (**2005**). *The Influence of Consonantal Context on Native and Non-native Vowel Production: A Cross-language Study* (Gunter Narr, Tübingen).

Steinlen, A. K., and Bohn, O.-S. (**1999**). "Acoustic studies comparing Danish vowels, British English vowels, and Danish-accented British English vowels," J. Acoust. Soc. Am. **105**, 1097.

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., and Öhman, S. (**1969**). "Cross-language study of vowel perception," Lang Speech **12**, 1–23.

Strange, W., and Bohn, O.-S. (**1998**). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.

Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–1807.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–170.

Strange, W., Weber, A., Levy, E., Shafiro, V., and Nishi, K. (**2002**). "Within- and across-language acoustic variability of vowels spoken in different phonetic and prosodic contexts: American English, North German, and Parisian French," J. Acoust. Soc. Am. **112**(2), 2384.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). "Perceptual assimilation of American English vowels by Japanese listeners," J. Phonetics **26**, 311–344.

Yamada, T., Yamada, R. A., and Strange, W. (**1996**). "Perception of Japanese vowels by native speakers of American English," Proceedings of the Acoustical Society of Japan **1**, 383–384.

Simulations of string vibrations with boundary conditions of third kind using the functional transformation method

L. Trautmann[a),b)]
Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, FIN-02015 HUT, Finland

S. Petrausch[c)]
Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, D-91058 Erlangen, Cauerstr. 7, Germany

M. Bauer[d)]
Department of Electronic & Electrical Engineering, University College London, London WC1E 7JE, Torrington Place, United Kingdom

The functional transformation method (FTM) is an established mathematical method for accurate simulation of multidimensional physical systems from various fields of science, including optics, heat and mass transfer, electrical engineering, and acoustics. It is a frequency-domain method based on the decomposition into eigenvectors and eigenfrequencies of the underlying physical problem. In this article, the FTM is applied to real-time simulations of vibrating strings which are ideally fixed at one end while the fixing at the other end is modeled by a frequency-dependent input impedance. Thus, boundary conditions of third kind are applied to the model at the end fixed with the input impedance. It is shown that accurate and stable simulations are achieved with nearly the same computational cost as with strings ideally fixed at both ends. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992649]

I. INTRODUCTION

Physical modeling of musical instruments simulates vibrations by calculating a model derived from the physical analysis of the vibrating structure. The physical model is described by partial differential equations (PDEs) that can be solved with a number of approaches including finite difference methods,[1,2] mass spring models,[3] modal synthesis,[4] the digital waveguide method,[5,6] and the functional transformation method.[7] For a review of the physical modeling methods see, e.g., Refs. 8 and 9. These methods differ in the discretization principle of the continuous-time and continuous-space physical model. The discretization approaches vary in simulation accuracy and computational complexity. The development of the discretization approaches and combinations thereof is subject to current intensive research.[10,11]

The functional transformation method (FTM) is a discretization approach for continuous physical models that provides high simulation accuracy and inherent stability. The FTM has first been applied to sound synthesis of vibrating string models.[12] Extensions to the basic model of a vibrating string using the FTM are given in Refs. 13 and 9.

Simulation accuracy and stability are ensured because the continuous time- and space-dependent model, also referred to as initial-boundary-value problem (IBVP), is first solved analytically with appropriate transformations before it is discretized. Multirate realizations provide for computational efficiency.[14,15] One of the drawbacks of the FTM so far was the limitation to idealized termination of the vibrating object which was either fixed, loose, or had a purely resistive fixing.[9]

This article tackles this shortcoming by extending the FTM to arbitrary frequency-dependent boundary conditions of third kind. Thus, terminations can be considered which describe real musical instruments more precisely. The extension to arbitrary boundary conditions is developed for the example of a vibrating string terminated with boundary conditions of third kind at the bridge end. Boundary conditions of third kind model complex bridge impedances in string instruments. This realistic string fixing also allows vibrational coupling of two or more strings and different string polarizations via the bridge.

Two realizations are presented and discussed in terms of simulation accuracy, computational accuracy, and applicability to different models. These realizations are the frequency-shift approach (FSA) and the time domain boundary excitation approach (TDBEA). Compared to the eigenmode expansion presented in earlier publications (e.g., in Ref. 16) the FSA used here is calculated from a vector form of the wave equation with only first-order derivations. This allows us to include the impedance boundary conditions in a straightforward way. An approximating solution by linearization of the FSA realization is given in Refs. 17 and 18 and will be compared to the FSA as presented here.

[a)]Electronic mail: lutz.trautmann@acoustics.hut.fi
[b)]Also at University of Erlangen-Nuremberg, Multimedia Communications and Signal Processing, D-91058 Erlangen, Cauerstr. 7, Germany; Electronic mail: traut@LNT.de; URL: http://www.LNT.de/~traut
[c)]Electronic mail: stepe@lnt.de
[d)]Electronic mail: m.bauer@ee.ucl.ac.uk
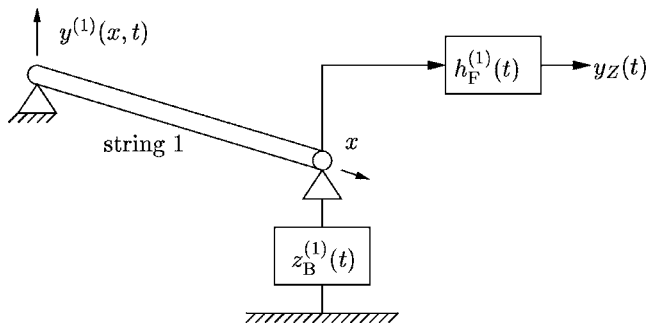
FIG. 1. String terminated with an impedance function $z_B(t)$ simulating the impedance at the bridge position seen by the string.

This article is organized as follows. In Sec. II, the physical model of an ideally transversal vibrating string is described by a vector PDE and its corresponding initial and boundary conditions. The physical model is solved with the FTM in two ways: in Sec. III A, the initial boundary value problem is solved with the FSA and in Sec. III B the solution is approximated with solution-dependent boundary excitations, the TDBEA. A comparison of both approaches is given in Sec. IV.

II. PHYSICAL MODEL

The construction of stringed musical instruments are modeled as shown in Fig. 1 where a string is clamped between nut and bridge. The string amplitude vibration $y(x,t)$ is given in form of a PDE. The string is fixed at the nut end ($x=0$) and terminated with an impedance function $z_B(t)$ at the bridge end ($x=l$). The bridge impedance of a classical guitar has actually larger effects on the string vibration than the nut impedance except for some single frequencies.[19] Since the principle derivation of the FTM is not affected by an additional impedance function at the nut end the model is kept as simple as possible.

The resulting bridge vibrations are filtered with the impulse response $h_F(t)$ simulating the linear and time-invariant (LTI) transmission path between the bridge vibration and the receiver of the sound. This impulse response mainly incorporates the instrument body and the effects of the surrounding room. This article focuses on modeling the string vibrations.

A. String model

The physical model of the vibrating strings is given as a vector PDE and its corresponding initial and boundary conditions. This IBVP results from the physical analysis of the forces acting on the vibrating string. A detailed derivation is given in, e.g., Ref. 9. In this article, the strings are assumed to be homogeneous and isotropic. Furthermore, the smoothness of their surfaces may not permit stress concentrations. The deflection of both strings is assumed to be small enough to change neither the cross section area nor the tension on the strings so that the strings behave linearly. Besides, string stiffness caused by Young's modulus and damping effects due to air damping and vibrational losses within the string are neglected.

Characterizing a string by its cross section area A, its mass density ρ, and the tension T_s applied to it gives the IBVP as the following spatially 1-D wave equation with appropriate initial and boundary conditions:

$$\mathbf{B}D_x\mathbf{y}(x,t) + \mathbf{C}D_t\mathbf{y}(x,t) = \mathbf{v}(x,t), \tag{1a}$$

$$\mathbf{f}_i^H\mathbf{y}(x,t) = \mathbf{I}_2\mathbf{y}(x,t)|_{t=0} = \mathbf{0}, \tag{1b}$$

$$\mathbf{f}_{b,B}^H\mathbf{y}(x,t) = \phi_B(x_B,t). \tag{1c}$$

with

$$\mathbf{B} = \mathbf{I}_2, \quad \mathbf{C} = \begin{bmatrix} 0 & -\frac{1}{T_s} \\ -\rho A & 0 \end{bmatrix},$$

$$\mathbf{v}(x,t) = \begin{bmatrix} 0 \\ f_{e1}(x,t) \end{bmatrix} = \begin{bmatrix} 0 \\ f_e(x,t) \cdot \delta(x) \end{bmatrix}. \tag{1d}$$

In PDE (1a), the operator D_x denotes first-order spatial derivative and D_t denotes first-order temporal derivative. The output vector $\mathbf{y}(x,t) = [v_y(x,t), f_y(x,t)]^T$ comprises the string's particle velocity $v_y(x,t) = \dot{y}(x,t)$ and the elastic restoring force $f_y(x,t) = T_s y'(x,t)$. Excitation vector $\mathbf{v}(x,t)$ contains the external excitation force density $f_{e1}(x,t) = f_e(x,t)\delta(x)$ as defined in (1d), where $\delta(x)$ denotes the Dirac impulse. Writing $\mathbf{y}(x,t)$ directly in temporal and spatial derivatives of the deflection $y(x,t)$ in (1a)–(1d) and inserting both coupled equations in (1a) into each other results in the well-known scalar wave equation

$$T_S y''(x,t) - \rho A \ddot{y}(x,t) = f_{e1}(x,t). \tag{2}$$

The initial conditions in Eq. (1b) are homogeneous. Here, $\mathbf{I}_2$ is the identity matrix of order two and the superscript H denotes the Hermitian operation, i.e., 5 conjugate transpose matrix. Thus, initial conditions are given here in terms of initial elastic restoring force $f_y(x,0)$ and initial velocity $v_y(x,0)$. They can easily be rewritten in the more common terms of initial displacement and velocity. Especially in the homogeneous case they are both the same since $f_y(x,0)$ only contains spatial derivatives of the displacement. The PDE (1a) and the initial conditions (1b) are only valid on the string when x is limited to $0 \leqslant x \leqslant l$.

The boundary conditions (1c) define the fixing of the string at both ends, $x=0$ and $x=l$. Given by the operator $\mathbf{f}_{b,B}^H$, the boundary conditions result in (time-dependent) boundary functions $\phi_B(x_B,t)$. Due to the different fixings of the string, the boundary conditions are different at $x=0$ (nut end) and at $x=l$ (bridge end). At the nut end the string is assumed to be ideally fixed with no boundary excitation. At the bridge end an elastic fixing is realized and, furthermore, excitations from other strings should be possible. This leads to homogeneous boundary conditions of first kind at the nut end and to nonhomogeneous boundary conditions of third kind at the bridge end as defined in (3a) and (3b), respectively:

TABLE I. Physical parameters of a typical idealized nylon "B" guitar string.

Variable	Name	Value	Unit
cross section area	$A = 0.5188\ mm^2$	0.5188	mm^2
length	$l=$	0.65	m
mass density	$\rho = 1140\ kg/m^3$	1140	kg/m^3
string tension	$T_s=$	60.97	N

$$\mathbf{f}_{b,0}^H \mathbf{y}(x,t) = \begin{bmatrix} 1 & 0 \end{bmatrix} * \mathbf{y}(0,t) = 0, \tag{3a}$$

$$\mathbf{f}_{b,l}^H \mathbf{y}(x,t) = [z_B(t)1] * \mathbf{y}(l,t) = \phi_l(l,t). \tag{3b}$$

The impulse response of the bridge is given by $z_B(t)$ and $*$ denotes temporal convolution. The bridge impulse response, respectively the bridge impedance, is discussed in more detail in Sec. II B.

Example: Physical parameters of typical idealized nylon "B" guitar strings as employed in Eq. (1d) are listed in Table I.[9] These string parameters are used throughout this article.

B. Bridge model

Frequency-dependent bridge impedances $Z_B(s)$ $= \mathcal{L}\{z_B(t)\}$ or bridge admittances $Y_B(s) = 1/Z_B(s)$ have been measured in previous publications, e.g., in Refs. 20, 21, and 19. In Refs. 20 and 19, an approximating model is derived for the bridge admittance $Y_B(s)$ consisting of several superposed modes. In Ref. 19 the lowest six most significant eigenmodes of the bridge are realized with damped sinusoids. This model neglects higher-frequency effects as well as nonprominent low-frequency modes. To show the principle effects of the bridge admittance to the string vibration this simplified model is used throughout this article. Thus, the bridge admittance $Y_B(s)$ is given by

$$Y_B(s) = \sum_{\kappa=1}^{6} \frac{s/m_\kappa}{s^2 + s a_\kappa + b_\kappa^2}. \tag{4}$$

The temporal frequency variable is denoted by s; a_κ and b_κ are real coefficients and the m_κ are effective input masses of the bridge seen from the string at the distinct frequencies. The modal frequencies and their damping coefficients are given by $\omega_\kappa = \sqrt{b_\kappa^2 - (a_\kappa/2)^2}$ and $\sigma_\kappa = a_\kappa/2$, respectively. The admittance functions of the bridge are position and direction dependent and thus differ for different strings and for both transversal string polarizations.[20]

Example: The approximation of the input admittance of a high-quality guitar bridge with six prominent modes as derived by measurements in Ref. 19 is shown in Fig. 2. The admittance is seen by the string vibrating perpendicular to the bridge. It is also used in this article, sometimes with modifications, to highlight effects of this nonideal string termination.
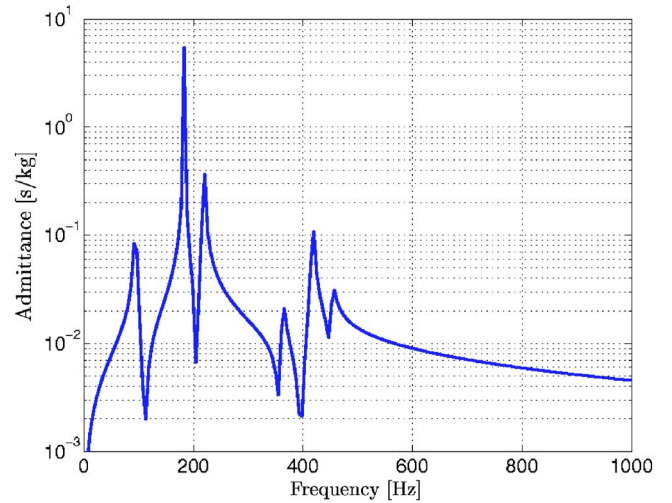
FIG. 2. (Color online) Input admittance of a guitar bridge approximated with six prominent modes as derived in Ref. 19.

III. SOLUTION WITH THE FUNCTIONAL TRANSFORMATION METHOD

The FTM solves IBVPs, as given in (1a)–(1d), analytically by using appropriate functional transformations. For a detailed description of the FTM see, e.g., Refs 22 and 9; here only an overview is given.

The principle procedure is shown in Fig. 3. The vector PDE with its initial and boundary conditions (ICs, BCs) is first transformed by the Laplace transformation ($\mathcal{L}\{\cdot\}$) with respect to time. The Laplace transformation removes the temporal derivatives and includes the initial conditions as additive terms. Thus, the vector PDE results in an ordinary differential equation(ODE). To transform the spatial variable into its frequency domain, the Sturm-Liouville transformation (SLT) ($\mathcal{T}\{\cdot\}$) is applied to it. The SLT is an integral transformation similar to the Laplace transformation with the difference that the transformation kernels and the spatial frequency variables are determined for each boundary-value problem separately. Therefore, the SLT is a key element of the FTM and the correct determination of the transformation kernels and the spatial frequency variables is the most sophisticated part. The application of the SLT to the boundary-value problem (ODE, BC) results in an algebraic equation which can be reordered to a multidimensional (MD) transfer function model (TFM). The MD TFM is the continuous solution of the IBVP in the temporal and spatial frequency domain. It is discretized with established discretization schemes. Application of the inverse SLT ($\mathcal{T}^{-1}\{\cdot\}$) and the inverse z transformation ($\mathcal{Z}^{-1}\{\cdot\}$) gives the discretized solution
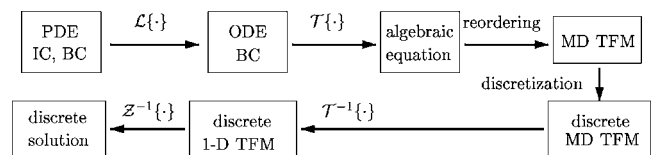
FIG. 3. FTM solving initial-boundary-value problems defined in the form of PDEs and initial conditions (ICs) and boundary conditions (BCs).

in the time and space domain. The solution consists of the modal data of the string depending directly on the physical parameters given in Table I.

This procedure is used in the next sections. Two different approaches will be introduced that are extending the FTM to include boundary conditions of third kind at the bridge end.

A. Boundary conditions simulated with frequency-shifts of the eigenfrequencies

The FSA realization of solving IBVPs with boundary conditions of third kind with the FTM consists of calculating the analytic eigenfunctions and eigenvalues of the spatial operator with respect to its original boundary conditions of third kind. The eigenfunctions are used as transformation kernels for the SLT while the eigenvalues define the discrete spatial frequency variable. It will be shown that the eigenvalues are directly connected to the eigenfrequencies of the vibrating string. For this problem, a solution using a linear approximation was presented in Ref. 18. There, an eigenfrequency shift proportional to the imaginary part of the bridge admittance and a damping constant proportional to the real part of the bridge admittance was used. Therefore, the analytical FSA derived here is compared to the method proposed in Ref. 18.

Since it is the most crucial step in this procedure, the SLT is derived first in Sec. III A 1. Then the MD TFM is presented in Sec. III A 2 and the inverse SLT and the inverse Laplace transformation lead to the continuous solution in Secs. III A 3 and III A 4, respectively. The discretization of the analytic solution and a simulation example are given in Sec. III A 5.

1. Sturm-Liouville transformation

As stated above, the key element of the FTM is the proper derivation of the SLT to transform the spatial variable into the spatial frequency domain. The SLT and its inverse are defined for the Laplace-transformed output vector $\mathbf{Y}(x,s)$ by

$$\mathcal{T}\{\mathbf{Y}(x,s)\} = \bar{Y}(\mu,s) = \int_0^l \widetilde{\mathbf{K}}^H(\mu,x)\mathbf{C}\mathbf{Y}(x,s)\,\mathrm{d}x, \tag{5a}$$

$$\mathbf{Y}(x,s) = \mathcal{T}^{-1}\{\bar{Y}(\mu,s)\} = \sum_\mu \frac{1}{N_\mu} \bar{Y}(\mu,s)\mathbf{K}(\mu,x). \tag{5b}$$

The SLT uses the eigenfunctions $\widetilde{\mathbf{K}}(\mu,x)$ and the eigenvalues $\widetilde{\beta}_\mu$ of the spatial operator $\widetilde{\mathbf{L}}$ with respect to the matrix $\mathbf{C}$ [compare (1a)–(1d)] as transformation kernels and discrete spatial frequency variables, respectively.[9] The index variable μ is an integer, $\mu \in \mathbb{N}$. The inverse SLT uses the eigenfunctions $\mathbf{K}(\mu,x)$ and the eigenvalues β_μ of the operator $\mathbf{L}$ as transformation kernels and spatial frequency variables.

In the example given in (1a) the operator $\mathbf{L}=\mathbf{B}\mathrm{D}_x$ equals $\widetilde{\mathbf{L}}=\mathbf{B}^H\mathrm{D}_x$, since both are given by $\mathbf{I}_2\mathrm{D}_x$. $\widetilde{\mathbf{L}}$ is the adjoint operator of $\mathbf{L}$.[22] Thus, the eigenvalue problems to be solved for the determination of the SLT are

$$\widetilde{\mathbf{L}}\{\widetilde{\mathbf{K}}(\mu,x)\} = \widetilde{\beta}_\mu \mathbf{C}\widetilde{\mathbf{K}}(\mu,x), \tag{6a}$$

$$\widetilde{\mathbf{f}}_{\mathrm{b},B}^H \widetilde{\mathbf{K}}^*(\mu,x) = \mathbf{0} \tag{6b}$$

for the forward transformation and

$$\mathbf{L}\{\mathbf{K}(\mu,x)\} = \beta_\mu \mathbf{C}\mathbf{K}(\mu,x), \tag{7a}$$

$$\mathbf{f}_{\mathrm{b},B}^H \mathbf{K}(\mu,x) = \mathbf{0} \tag{7b}$$

for the inverse SLT. Note that despite the equality of the operators $\mathbf{L}$ and $\widetilde{\mathbf{L}}$, (6) and (7) form adjoint eigenvalue problems since the adjoint boundary conditions $\widetilde{\mathbf{f}}_{\mathrm{b},B}^T$ are used for the forward transformation. Therefore the eigenvalue problem (6) is the adjoint eigenvalue problem to the inverse SLT (7).[23] The adjoint boundary operators are given by (compare with the derivation in Ref. 9)

$$\widetilde{\mathbf{f}}_{\mathrm{b},0}^H = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad \widetilde{\mathbf{f}}_{\mathrm{b},l}^H = \begin{bmatrix} 1 & -z_{\mathrm{B}}^*(t) \end{bmatrix} \tag{8}$$

at the nut and the bridge end, respectively. The eigenfunctions of these boundary-value problems are calculated using (6a) and (7a) and the boundary conditions (6b) and (7b) at the nut end $(x=0)$ to

$$\mathbf{K}(\mu,x) = \begin{bmatrix} -\frac{c}{T_s}\sinh\left(\frac{\beta_\mu}{c}x\right) \\ \cosh\left(\frac{\beta_\mu}{c}x\right) \end{bmatrix} K_2(\beta_\mu,0), \tag{9a}$$

$$\widetilde{\mathbf{K}}(\mu,x) = \begin{bmatrix} \cosh\left(\frac{\widetilde{\beta}_\mu}{c}x\right) \\ \frac{c}{T_s}\sinh\left(\frac{\widetilde{\beta}_\mu}{c}x\right) \end{bmatrix} \widetilde{K}_1(\widetilde{\beta}_\mu,0) \tag{9b}$$

(see Ref. 22). Here, $K_2(\beta_\mu,0)$ and $\widetilde{K}_1(\widetilde{\beta}_\mu,0)$ are scaling factors. The real-valued coefficient $c=\sqrt{T_s/(\rho A)}$ is the wave velocity of the string. Inserting the boundary conditions at the bridge end (3b) and (8) into (9) and solving for the eigenvalues β_μ and $\widetilde{\beta}_\mu$ leads to

$$\beta_\mu = \widetilde{\beta}_\mu^* = -c\frac{\ln r(s)}{2l} + jc\frac{\mu\pi}{l} = -c\frac{\ln|r(s)|}{2l}$$
$$+ jc\frac{\mu\pi - 0.5\angle(r(s))}{l}. \tag{10}$$

The argument of $r(s)$ is given by $\angle(r(s))$ where $r(s)$ denotes the frequency-dependent reflection coefficient, given by

$$r(s) = \frac{Z_{\mathrm{B}}(s) - R_{\mathrm{w}}}{Z_{\mathrm{B}}(s) + R_{\mathrm{w}}}. \tag{11}$$

The eigenvalues β_μ of (7) are the conjugate complex counterparts of the eigenvalues $\widetilde{\beta}_\mu$ of the adjoint eigenvalue problem (6) as described in Ref. 22. The characteristic impedance of the string is denoted by $R_{\mathrm{w}}=T_s/c$. Due to the frequency dependence of $r(s)$ the spatial eigenvalues $[\beta_\mu(s)$ and $\widetilde{\beta}_\mu(s)]$ and the spatial eigenfunctions $[\mathbf{K}(\mu,x,s)$ and $\widetilde{\mathbf{K}}(\mu,x,s)]$ also depend on s. Thus, the independent treatment of time t and and space x in the Laplace transformation and the SLT, used in previous publications for the derivation of FTM solutions, is not valid with the application of frequency-dependent boundary conditions of third kind. This is an extension of the

SLT described in Ref. 9. However, the SLT (5) is still a valid transformation and inserting (5b) into (5a) results in the same functions $\bar{Y}(\mu,s)$. Despite the frequency dependencies the fundamental bi-orthogonality of the eigenfunctions with respect to the matrix $\mathbf{C}$ is still ensured since the temporal frequency variable s must be seen as a fixed value for the proof of bi-orthogonality similar to the spatial variable x; for further details see Ref. 9. The norm factor N_μ used for the inverse SLT (5b) equals to $N_\mu = -l/T_s$ with $K_2(\beta_\mu,0) = \tilde{K}_1(\tilde{\beta}_\mu,0) = 1$.

2. Multidimensional transfer function model

The application of the Laplace transformation with respect to time and the SLT with respect to space to the IBVP (1a)–(1d) yields an algebraic equation. It can be solved for the transformed output variable $\bar{Y}(\mu,s)$ resulting in a MD TFM

$$\bar{Y}(\mu,s) = \frac{\bar{V}(\mu,s) + \Phi_B(\mu,s)}{s + \tilde{\beta}_\mu(s)}. \tag{12}$$

The transformed excitation function $\bar{V}(\mu,s)$ and the transformed boundary conditions $\bar{\Phi}_B(\mu,s)$ are given by[9]

$$\bar{V}(\mu,s) = \int_0^l \tilde{\mathbf{K}}^H(\mu,x,s)\mathbf{V}(x_e,s)\,\mathrm{d}x$$

$$= [0 \quad 1]\tilde{\mathbf{K}}^*(\mu,x_e,s)F_e(x_e,s), \tag{13a}$$

$$\bar{\Phi}_B(\mu,s) = [1 \quad 0]\tilde{\mathbf{K}}^*(\mu,x_B,s)\Phi(x_B,s). \tag{13b}$$

A prerequisite of (13a) is that the excitation force $F_e(x_e,s)$ only acts on one spatial position x_e of the string.

3. Inverse Sturm-Liouville transformation

Inserting the MD TFM (12) into the inverse SLT (5b) without first discretizing the model results in

$$\mathbf{Y}(x,s) = \mathcal{T}^{-1}\{\bar{Y}(\mu,s)\} = \sum_\mu \frac{1}{N_\mu}\bar{Y}(\mu,s)\mathbf{K}(\mu,s,x) \tag{14a}$$

$$= -\frac{T_s}{l}\sum_\mu \frac{\bar{V}(\mu,s)}{s + \tilde{\beta}_\mu(s)}\mathbf{K}(\mu,s,x) \tag{14b}$$

$$= -\frac{T_s}{l}F_e(x_e,s)\sum_\mu \frac{1}{s + \tilde{\beta}_\mu(s)}$$

$$\times ([0 \quad 1]\tilde{\mathbf{K}}^*(\mu,s,x_e)\mathbf{K}(\mu,s,x)) \tag{14c}$$

$$= -\frac{T_s}{l}F_e(x_e,s)\sum_\mu \frac{1}{s - c\{[\ln r(s)]/2l\} - jc(\mu\pi/l)}$$

$$\times \mathbf{H}(\mu,s,x) \tag{14d}$$

with

$$\mathbf{H}(\mu,s,x) = \begin{bmatrix} -\left(\frac{c}{T_s}\right)^2 \sinh\left(\frac{\beta_\mu(s)}{c}x_e\right)\sinh\left(\frac{\beta_\mu(s)}{c}x\right) \\ \frac{c}{T_s}\sinh\left(\frac{\beta_\mu(s)}{c}x_e\right)\cosh\left(\frac{\beta_\mu(s)}{c}x\right) \end{bmatrix}. \tag{14e}$$

For simplification, only the excitation function $F_e(x_e,s)$ is included here, while the boundary excitation $\Phi(x_B,s)$ is omitted. The continuous solution $\mathbf{Y}(x,s)$ of the IBVP in the temporal frequency domain consists of the excitation function $F_e(x_e,s)$ weighted with different transfer functions. These transfer functions contain an elaborate denominator since the reflection coefficient $r(s)$ depends on s. In the next section, the application of the inverse Laplace transformation to this complicated system is discussed in more detail.

4. Inverse Laplace transformation

The MD transfer functions

$$\frac{1}{s - c\{[\ln r(s)]/2l\} - jc(\mu\pi/l)} = \frac{1}{D_\mu(s)} \tag{15}$$

given in (14d) are meromorph functions with isolated singular points $s_{\infty\mu,i}$. Since $D_\mu(s)$ are analytical in $s_{\infty\mu,i}$ and $s_{\infty\mu,i}$ are singular roots of $D_\mu(s)$, $1/D_\mu(s)$ can be developed into its residuals with

$$\frac{1}{D_\mu(s)} = \sum_i \frac{\mathrm{Res}(s_{\infty\mu,i})}{s - s_{\infty\mu,i}}, \quad \mathrm{Res}(s_{\infty\mu,i})\frac{1}{(\partial/\partial s)D_\mu(s)|_{s=s_{\infty\mu,i}}}. \tag{16}$$

The first-order derivative of $D_\mu(s)$ is transferred with (11) and (15) to

$$\frac{\partial}{\partial s}D_\mu(s)|_{s=s_{\infty\mu}} = 1 + \frac{c}{2l}\frac{R_w(\partial/\partial s)Z(s)|_{s=s_{\infty\mu}}}{(Z(s_{\infty\mu}) - R_w)(Z(s_{\infty\mu}) + R_w)}. \tag{17}$$

The poles $s_{\infty\mu}$ of the transfer functions defined in (15) are not given explicitly since $r(s)$ depends on s. However, they can be calculated iteratively for example from

$$s_{\infty\mu}^{(n)} = c\frac{\ln r(s_{\infty\mu}^{(n-1)})}{2l} + jc\frac{\mu\pi}{l}. \tag{18}$$

This iteration converges to the zeros of $D_\mu(s)$ as far as the starting point is chosen within the conversion area of the complex reflection coefficient, i.e., close enough to the zeros. This is fulfilled for physical problems of this kind since the deviations are small. Another iterative method that proved to be good in practice is the normalized least mean squares (NLMS) algorithm in Ref. 24. It is a steepest descent method using the first-order derivatives of $D_\mu(s)$ calculated in (17). The required number of iterations for both methods depends on the smoothness of the reflection factor and on the desired accuracy of the result. In practice, between 5 and 15 iterations offer sufficient results. Note that, depending on the mobility of the bridge, the boundary conditions of third kind can split one eigenfrequency into two such that for one value of μ two different poles $s_{\infty\mu,i}$ exist with $i = 1, 2$.[25] Finding both poles can be ensured by setting the starting positions $s_{\infty\mu}^{(1)}$ to both sides of the closest eigenfrequency of the bridge as defined in (4).

Once the poles are identified, the inverse Laplace transformation can be applied to the single transfer functions as given in (16). For the entire system, as defined in (14a)–(14e), the situation is more complex since also the combination of spatial transformation kernels in $\mathbf{H}(\mu,s,x)$ depends on s. Applying the inverse Laplace transformation to $\mathbf{H}(\mu,s,x)$ would result in hypergeometrical functions that have to be determined by numerical integration for each parameter set and for each spatial position separately. This would decrease the computational efficiency of the FTM. Fortunately, the complicated spectra of $\mathbf{H}(\mu,s,x)$ are multiplied with the spectrum of the transfer function given in (15). Since this spectrum only consists of single discrete eigenfrequencies, the steady-state solution in the time domain is simply a weighting of the single spectral lines with the respective values of $\mathbf{H}(\mu,s,x)$ at the spectral positions $s_{\infty\mu}$. This simplification only affects the transiental behavior of the simulated string vibration; the spectra remain unchanged.

Using the procedure explained above, the system to be transformed back into the time domain as described by (14a)–(14e) simplifies to

$$\mathbf{Y}(x,s) = -\frac{T_{\mathrm{s}}}{l}F_{\mathrm{e}}(x_{\mathrm{e}},s)\sum_{\mu,i}\frac{\mathrm{Res}(s_{\infty\mu,i})}{s-s_{\infty\mu,i}}\mathbf{H}(\mu,s_{\infty\mu,i},x). \quad (19)$$

A real-valued continuous system in the time and space domain is derived by taking advantage of a number of properties. The poles $s_{\infty\mu}$ form complex conjugate pairs $s_{\infty\mu}=\sigma_\mu+j\omega_\mu$, the residuals $\mathrm{Res}(s)$ are symmetric with respect to frequency as seen in (16) and (17), and the first component of $\mathbf{H}(\mu,s_{\infty\mu},x)$ is symmetric while the second component is antisymmetric. Hence, the real-valued system results in

$$\mathbf{y}(x,t) = -\frac{2T_{\mathrm{s}}}{l}f_{\mathrm{e}}(x_{\mathrm{e}},t) * \sum_{\mu\geq 0,i}\mathrm{Res}(s_{\infty\mu,i})\mathbf{H}(\mu,s_{\infty\mu,i},x)e^{\sigma_{\mu,i}t}$$

$$\times\begin{bmatrix}\cos(\omega_{\mu,i}t)\\\sin(\omega_{\mu,i}t)\end{bmatrix}. \quad (20)$$

Here, * denotes temporal convolution. The continuous solution consists of the superposition of several eigenfrequencies with both frequencies and damping coefficients directly defined by the poles $s_{\infty\mu,i}$ of the MD TFM as calculated iteratively in (18). Corresponding to (10) and (15) the modal frequencies are shifted compared to the ideally fixed string $[r(s)\equiv 1]$ proportional to the argument of the reflection factor $r(s)$. Furthermore, some damping proportional to the logarithm of the absolute value of $r(s)$ is introduced with this kind of fixing.

In Ref. 18, an eigenfrequency shift proportional to the imaginary part of the bridge admittance and a damping constant proportional to the real part of the bridge admittance is applied such that

$$s_{\infty\mu} = \frac{T_{\mathrm{S}}}{l}\mathrm{Re}\{Y_{\mathrm{B}}\} + j\left(c\frac{\mu\pi}{l} - \frac{T_{\mathrm{S}}}{l}\mathrm{Im}\{Y_{\mathrm{B}}\}\right). \quad (21)$$

Comparing this approximation (21) used in Ref. 18 with the linearization of the eigenvalue problem, first derived in Ref. 17, with (18), the solution derived here is on one hand more
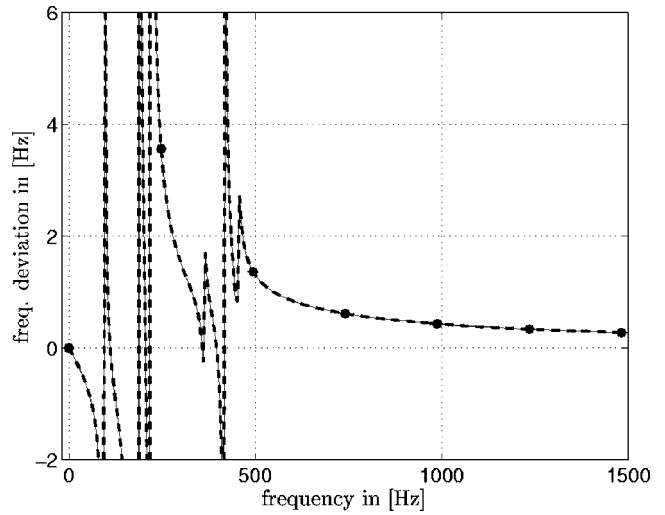
FIG. 4. Frequency shifts due to the boundary conditions of third kind for a vibrating string. The solid line shows the weighted phase angle of the reflection function $-0.5c\angle(r(s))/l$, the dashed line (covered by the solid line at the shown frequencies) shows the weighted imaginary part of the bridge admittance $\mathrm{Im}\{Y_{\mathrm{B}}(s)\}T_{\mathrm{s}}/l$. The dots are the eigenfrequencies and their deviation from an ideally fixed string.

accurate but, on the other hand computationally less efficient. The linear approximation (21) is only accurate for bridge admittances that are small compared to the characteristic string admittance. This is shown in the next examples.

a. Examples. In fact, the bridge admittance $Y_{\mathrm{B}}(s)$ is small compared to the characteristic string admittance $1/R_{\mathrm{w}}$ for almost all frequencies using the parameters given in Table I and Ref. 19. The only frequency where this is not true is at the second bridge mode which has a similar magnitude as the characteristic string admittance. However, the approximation propsed in Ref. 18 works well using the string parameters given in Table I since there is no eigenfrequency of the string at this particular frequency. This is shown in Fig. 4.

The frequency deviation due to the nonideal fixing of the string can either be calculated with (18) or (21) for this parameter set. Both, the solid line representing the frequency deviation derived in (18) and the dashed line representing the frequency deviation derived in (21), have only minor discrepancies. The discrete eigenfrequencies of the string are intersections of both trajectories. The maximum frequency deviation from the ideally fixed string is less than 4 Hz for the lowest eigenfrequency, the frequency deviations of the higher eigenfrequencies converge to 0 Hz since the bridge is almost a rigid fixing for the string at these frequencies.

However, the more accurate model from (18) should be used for string or bridge parameters with string eigenfrequencies close to the prominent bridge modes. This is shown in Fig. 5 where the bridge parameters from Ref. 19 and (4) are modified to $b_\kappa^{(m)}=10b_\kappa$, $m_\kappa^{(m)}=0.001m_\kappa$, and $a_\kappa^{(m)}=0.1a_\kappa$.

Due to the ten times higher bridge modes more string eigenfrequencies are within the region of interest. Since the effective mass is lowered, the bridge is not as rigid as when using the original parameter set. Therefore, the frequency deviations caused by the nonideal fixing are much higher and up to 700 Hz. As the bridge admittance is in this case no
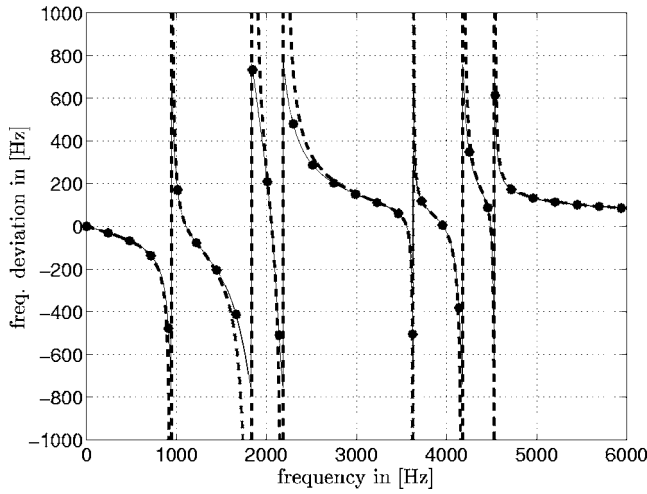
FIG. 5. Frequency shifts due to the boundary conditions of third kind for a vibrating string. The lines show the same functions as in Fig. 4 for another bridge parameter set.

FIG. 6. Discrete structure for the simulation of the elastic restoring force $f_y^d(x_a, k)$ derived with the FTM in a transversal vibrating string. Boundary conditions of third kind at the bridge position are realized with the correct eigenfrequencies of the complete system.

longer small compared to the characteristic string admittance the deviation between the iterative analytical solution (18) and the first-order approximation (21) becomes much higher. For the eigenfrequency at 2300 Hz the deviation is almost 250 Hz which cannot be further neglected. Moreover, the split of the eigenfrequencies of the string caused by the bridge modes cannot be reproduced with (21) due to the linearization of the problem.

In the previous examples only the frequency deviations caused by the boundary conditions of third kind are discussed. A similar discussion with the same results can be performed for the loss coefficients of the string modes. As a result, depending on the parameters of the string and the bridge, either the computationally efficient first-order approximation (21) is sufficient for the derivation of the eigenfrequencies of the string or the more accurate but computationally more expensive iterative method (18) must be used. With these derived eigenfrequencies we can construct a discrete system. The efficient implementation of such a system is shown in the next section.

5. Discretization

The discretization of the continuous MD TFM (12) as shown in Fig. 4 can be performed with standard s-to-z mapping. Assuming a temporal sampling interval T, such that $t = kT$, $k \in \mathbb{N}$, and using the impulse-invariant transformation with $z = e^{-sT}$, the temporal discretization is a straightforward operation.[9]

Applying the inverse SLT (5b) to this discretized model results in

$$\mathbf{Y}^d(x_a, z) = -\frac{T_s}{l} F_e^d(x_e, z) \sum_{\mu_T} \frac{\mathrm{Res}(s_{\infty\mu_T})z}{z - e^{s_{\infty\mu_T}T}} \mathbf{H}(\mu_T, s_{\infty\mu_T}, x_a).$$

(22)

Due to the discretization with sampling interval T the summation for the inverse SLT in (22) is limited to μ_T such that the simulated frequencies are below the Nyquist frequency with $|\mathrm{Im}\{s_{\infty\mu_T}\}T| = |\omega_{\mu_T}T| \leq \pi$. In addition, the spatial vari-

able x is discretized to a limited number of arbitrary spatial positions x_a. Spatial positions can be, for instance, the pick-up positions or the bridge position on the guitar. The discretized variables are denoted with superscript d. For details of the discretization procedure see Ref. 9.
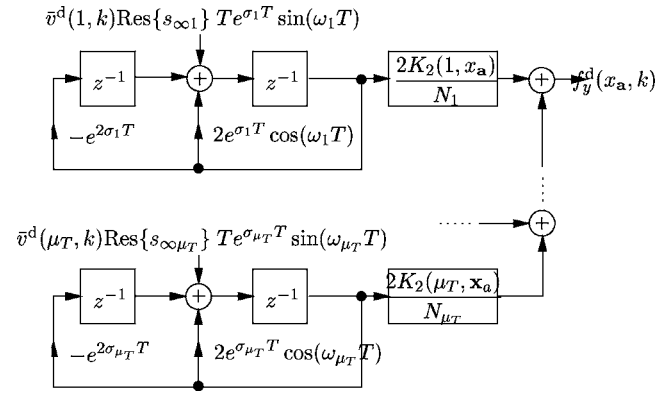
Conjugate complex pole pairs can also be combined in the discretized domain to a second-order system to avoid complex operations. The inverse z-transformation using the shifting theorem results in a number of parallel second-order recursive systems, each one realizing one vibrational eigenfrequency. This is shown for the calculation of the elastic restoring force $f_y^d(x_a, k)$ in Fig. 6.

This system can be implemented in a straightforward fashion as it only contains delays, multipliers, and adders. Assuming an average number of 70 simulated eigenfrequencies per string for a six-string guitar, this simulation needs about 231 multiplications per output sample (MPOS) and string. For more efficient implementations also filter bank approaches can be used as described, e.g., in Refs. 14 and 15. This saves up to almost 80% of the computational cost for a six-string guitar.[14]

Thus, from the real-time simulation point of view the system derived above is quite efficient. However, calculating the modal frequencies and the corresponding damping coefficients can become demanding for more complicated systems in comparison to systems terminated only with a fixed bridge impedance. In the next section an alternative method for the realization of boundary conditions of third kind is introduced.

B. Boundary conditions simulated with boundary excitations

In contrast to calculating the eigenfrequencies of the complete system explicitly as performed in Sec. III A, this section derives a method which avoids these difficulties. The idea is to simplify the boundary conditions of third kind in the IBVP (1a)–(1d) to boundary conditions of first kind, i.e., an ideal fixing of the string, and adding a solution-dependent boundary excitation function. The desired output signal $\mathbf{y}(x,t)$ is realized by the sum of an idealized output signal $\mathbf{y}^{ideal}(x,t)$ and a superposed signal $\mathbf{y}^{super}(x,t)$. The idealized

output signal solves the IBVP with boundary conditions of first kind while the superposed signal is excited by a solution-dependent boundary function.

In Sec. III B 1 the time-domain boundary excitation approach (TDBEA) is derived. The resulting structure is discretized in Sec. III B 2, giving an example with the parameters as employed in the second example in Sec. II.

1. Time-domain approach

It has been discussed in Sec. III A that typical bridge admittances are small compared to the characteristic string admittances. Therefore, the string appears almost fixed at the bridge end and a straightforward simplification of this model is to assume that the string is ideally fixed at both ends. Therefore, the analytical eigenfrequencies of the vibrating string, ideally fixed at both ends, are calculated first to derive the idealized signal $\mathbf{y}^{\text{ideal}}(x,t)$. The idealized eigenfrequencies are derived as shown in Fig. 3 and as described in Sec. III A but with boundary conditions of first kind (3a) at both ends. The eigenvalues and eigenfunctions used for the SLT and its inverse result in

$$\beta_\mu = jc\frac{\mu\pi}{l}; \quad \mathbf{K} = \begin{bmatrix} -j\frac{c}{T_s}\sin\left(\frac{\mu\pi}{l}x\right) \\ \cos\left(\frac{\mu\pi}{l}x\right) \end{bmatrix}, \tag{23a}$$

$$\widetilde{\beta}_\mu = -jc\frac{\mu\pi}{l}; \quad \widetilde{\mathbf{K}} = \begin{bmatrix} \cos\left(\frac{\mu\pi}{l}x\right) \\ j\frac{c}{T_s}\sin\left(\frac{\mu\pi}{l}x\right) \end{bmatrix}, \tag{23b}$$

see Ref. 22. By setting the impedance function $Z_B(s)$ to infinity, (23) can directly be obtained from (9)–(11). The application of the SLT to the idealized string generates the output vector $\mathbf{y}^{\text{ideal}}(x,t)$. At $x=l$ only the output $f_y^{\text{ideal}}(l,t)$ can contain nonzero values. The velocity $v_y^{\text{ideal}}(l,t)$ of this idealized model at the bridge position is always zero due to the idealized fixing of the string. The relationship between $v_y^{\text{ideal}}(l,t)$ and $f_y^{\text{ideal}}(l,t)$ does not fulfill the boundary condition of a string terminated with the impedance $z_B(t)$ as defined in (3b) since the desired termination of the string is not realized in this idealized model.

To satisfy the boundary conditions of third kind at the bridge position (3b) a second signal $\mathbf{y}^{\text{super}}(x,t)$ is superposed to the idealized output. In a first step, the superposed signal is excited via the nonzero boundary function of the ideally fixed string $f_y^{\text{ideal}}(l,t)$. It then excites the boundary of the ideally fixed string via $\phi_l^{\text{super}}(l,t)=(\mathbf{f}_{\text{b},l}^{\text{ideal}})^H \mathbf{y}^{\text{super}}(x,t) = v_y^{\text{super}}(l,t)$. The superposed velocity excitation of the string via the boundary also generates a force wave from the boundary which is related to the velocity by the characteristic string impedance due to the definition of the characteristic string impedance. For a consistent notation with the equations to follow it is transformed into the Laplace domain:

$$F_y^{\text{super}}(l,s) = R_w V_y^{\text{super}}(l,s). \tag{24}$$

By adding the ideal vibration and the superposed vibration in the Laplace domain, the desired boundary conditions of third kind (3b) are modified to

$$Z_B(s)(V_y^{\text{ideal}}(l,s) + V_y^{\text{super}}(l,s)) = -(F_y^{\text{ideal}}(l,s) + F_y^{\text{super}}(l,s)). \tag{25}$$

The assumption is made that there is no external excitation force $f_e(x_e,t)$ at the bridge position l. Nevertheless, there might be excitations elsewhere on the string. The superposed velocity signal $V_y^{\text{super}}(l,s)$ is excited by the idealized string output $F_y^{\text{ideal}}(l,s)$ with

$$V_y^{\text{super}}(l,s) = A(s)F_y^{\text{ideal}}(l,s). \tag{26}$$

Inserting (26) into (25) and solving for $A(s)$ results in $A(s) = -1/(Z_B(s)+R_w)$ since $V_y^{\text{ideal}}(l,s) \equiv 0$. In the time domain the multiplication of both frequency functions in (26) corresponds to a temporal convolution such that

$$v_y^{\text{super}}(l,t) = a(t) * f_y^{\text{ideal}}(l,t) \quad \text{with } a(t) = \mathcal{L}^{-1}\{A(s)\}. \tag{27}$$

Thus, the procedure for calculating the output of the string terminated with boundary conditions of third kind can be described by the following steps:

(1) Calculate the output $f_y^{\text{ideal}}(l,t)$ of the ideally terminated string at the boundary $x=l$.
(2) Generate the superposed signal $v_y^{\text{super}}(l,t)$ with (27).
(3) Use the superposed velocity signal $v_y^{\text{super}}(l,t)$ as a boundary excitation $\phi_l^{\text{super}}(l,t)$ for the idealized string.
(4) Add both signals, the idealized and the superposed.

Since the calculations are performed in the spatial frequency domain, all other positions $x \neq l$ on the string can be calculated by just other weightings with the transformation kernels $\mathbf{K}(\beta_\mu,x)$ as given in (23a). This is shown in Ref. 9.

2. Discretization

The discretization is performed similar to the procedure presented in Sec. III A 5. An s-to-z mapping with the impulse-invariant transformation is applied to the MD TFM (12). Then the inverse SLT (5b) and the inverse z transformation are used by applying the shifting theorem. To keep the system simple the conjugate complex pole pairs are not combined to a second-order system as performed in Sec. III A 5. It results with $\bar{y}^d(\mu,k) = \mathcal{Z}^{-1}\{\bar{Y}^d(\mu,z)\}$, with $\bar{y}^d(\mu,k)$ being the sum of the idealized signal $\bar{y}^{\text{ideal}}(\mu,k)$ and the superposed signal $\bar{y}^{\text{super}}(\mu,k)$, in

$$\bar{y}^{\text{ideal}}(\mu_T,k) = e^{-\widetilde{\beta}_{\mu_T}T}\bar{y}^d(\mu_T,k-1) + \bar{v}(\mu_T,k), \tag{28a}$$

$$f_y^{\text{ideal}}(l,k) = \sum_{\mu_T} \frac{1}{N_{\mu_T}} \bar{y}^{\text{ideal}}(\mu_T,k)K_2(\mu_T,l), \tag{28b}$$

$$\phi_l^{\text{super}}(l,k) = Ta^d(k) * f_y^{\text{ideal}}(l,k), \tag{28c}$$

$$\bar{y}^d(\mu_T,k) = \bar{y}^{\text{ideal}}(\mu_T,k) - T\phi_l^{\text{super}}(l,k)\widetilde{K}_1^*(\mu_T,l). \tag{28d}$$

Here, $a^d(k)$ is the discretized impulse response of $a(t)$. It can be included in the system in two ways:
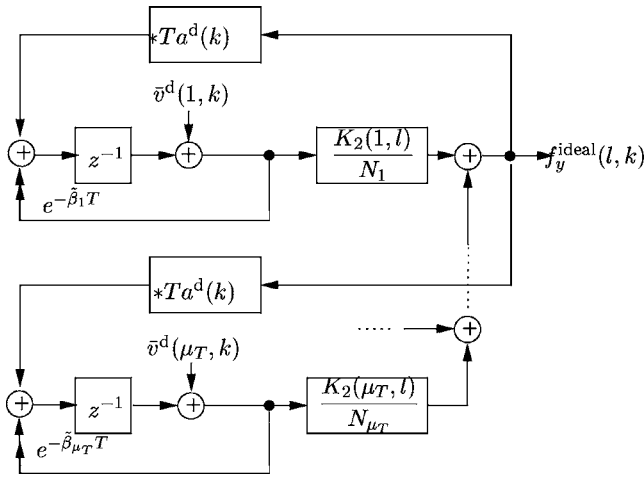
FIG. 7. Discrete structure for the simulation of the elastic restoring force $f_y^d(x_a,k)$ derived with the FTM in a transversal vibrating string with boundary conditions of third kind at the bridge position realized with the time-domain coupling approach.

(i)   By exciting several parallel recursive systems of second order representing the bridge modes similar to the eigenfrequencies of the string. The output of these bridge modes are then the input of the recursive systems for the string.

(ii) By direct convolution with a precalculated impulse response of the bridge.

The first realization is preferable for bridge models that contain only a few modes, as for example given in (4), and that have a long impulse response. The first realization has the additional advantage that it is a parametric representation such that online modifications of the bridge are possible. The realization of the TDBEA is shown in Fig. 7. For more complicated bridge models, consisting of several highly damped modes, the second realization is recommended since the impulse response is short in that case. The calculated force of the idealized model at the bridge position, $f_y^{ideal}(l,k)$, is convolved with $Ta^d(k)$ and then fed back into all first-order recursive systems to excite the superposed signal. This procedure can be realized for all, purely resistive terminations $Z_B \equiv$ const, arbitrary reactive networks, and interconnections of distributed systems with different parameters. For the in-

terconnection of different distributed systems each string is terminated at the interconnection position with the characteristic impedance of the interconnected string. The calculated superposed signal is then applied to both strings.

In this TDBEA the impact of the boundary conditions to the frequency domain cannot be directly seen. Therefore, a frequency-domain boundary excitation approach is derived in the Appendix .

The feedback path shown in Fig. 7 can be directly implemented as no delay-free loops occur. However, applying a filter-bank approach to this system as proposed in Sec. III A 5 causes simulation inaccuracies because the feedback paths, unlike the string eigenfrequencies, do not necessarily have a limited spectral spread. Thus, the feedback path impulse-responses cannot be subsampled without causing additional delays in the feedback path and transient errors in the output signal. Details of the usage of the filter-bank approach for FTM models containing feedback paths are discussed in Ref. 15.

a. Example. By simulating the vibrating string with boundary conditions of third kind with the simulation model derived in Sec. III B 2 the string velocity and the elastic restoring force are directly calculated. For this example the bridge admittances, as defined in the second example of Sec. III A 4, are used. The left plot of Fig. 8 shows the elastic restoring force of the ideally fixed string, $f_y^{ideal}(x,t)$, whereas the right plot shows the elastic restoring force of the superposed signal, $f_y^{super}(x,t)$. The superposed signal is zero as long as there is no force at the bridge position. When the wave of the idealized model reaches the bridge position at $t=1.2$ ms the bridge is excited to vibrate in its modes, resulting in a temporally distributed signal acting on the string by the superposed wave. Note that due to the extremely elastic bridge the effect is much larger here than for normal instrument bridges.

IV. COMPARISON OF THE FSA AND THE TDBEA

In this section, the FSA of Sec. III A and the TDBEA of Sec. III B 1 are compared. First, both methods are compared with respect to their computational complexities in Sec. IV A. The simulation accuracies in the frequency-domain as
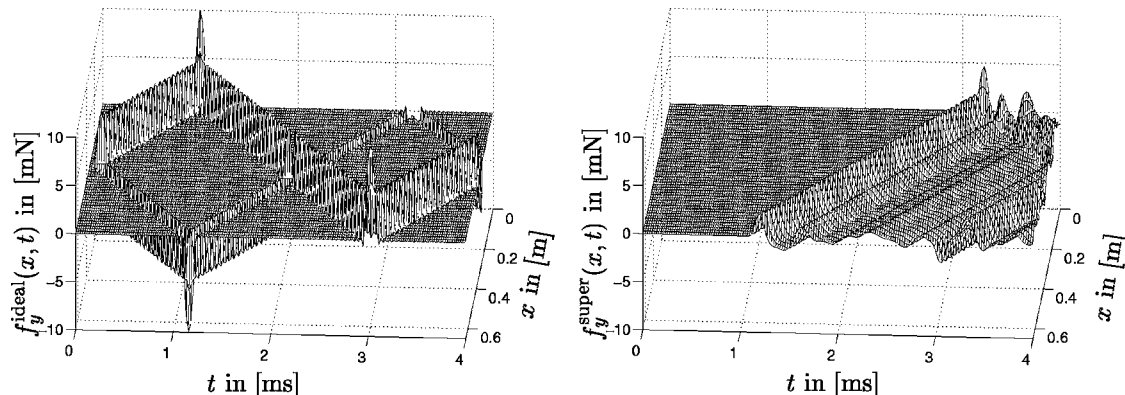
FIG. 8. Elastic restoring forces of a transversal vibrating string. Left plot: vibrations of the idealized string with boundary conditions of first kind. Right plot: superposed signal caused by the boundary conditions of third kind.

well as in the time domain are compared in Sec. IV B. In Sec. IV C the conclusions for the comparison are drawn.

A. Computational complexity

As discussed in Sec. III A 5, the calculation of the sound output at sample rate requires in the audio rate FSA about 231 MPOS for a typical guitar string with an average number of 70 simulated modes. Taking advantage of the filter-bank approach reduces the simulation computation of a six-string guitar to 305 MPOS,[14] that is, an average number of 51 MPOS per string is needed. In comparison, the TDBEA of Sec. III B 1 cannot deploy the efficiency of the filter-bank approach. Thus, the computational cost per string stays at 231 MPOS. On top of this, each additional output position needs on average 70 MPOS per string. If the elastic restoring force at the bridge position is not the desired output, this additional output has to be evaluated for the calculation of the boundary excitation function $\phi_l^{super}(l,k)$. Additionally, the bridge vibration in the time-domain approach is convolved with the impulse response $b^d(k)$ as given in (28c). Independent of the choice of the realization of this convolution, the parametric or the direct approach as described in Sec. III B 2, it introduces an additional computational cost. Thus, the computational complexity of the TDBEA is much higher than the frequency-shift approach with respect to sample-rate computations.

However, not only the sample-rate computations have to be taken into account, also the computational cost for the system setup has to be considered. Each time one of the system's parameters is changed, for example the length of the string, the coefficients of the recursive systems of second order have to be calculated from scratch. This does not have to be done at sample rate but only at a lower control-rate, e.g., each 10 ms.

The control-rate computations are quite efficient for the TDBEA since the eigenfrequencies of the idealized system are given analytically in (23). Also the calculation of the impulse response $a^d(k)$ or its corresponding parametric representation needs only a few calculations and has to be performed only when the bridge impedance or the characteristic string impedance are modified. When the string length is changed, which is assumed to be the most often parameter change, $a^d(k)$ stays unchanged. In the FSA the control-rate setup requires much more computational cost since the eigenfrequencies and their residuals have to be calculated iteratively with (11), (18), and (17). Assuming approximately 60 cycles on a digital signal processor for each iteration of (18), ten iterations, and 20 cycles for the calculation of the residual, each eigenfrequency needs 620 cycles for the update of the eigenfrequencies. With an update rate of 10 ms, a sampling frequency of 44.1 kHz, and 70 modes, the control-rate computational cost is about 100 MPOS which is no longer negligible. However, the parameters will not be updated every 10 ms, such that 100 MPOS must be seen as a peak computational cost.

In conclusion, the audio-rate computational cost is much lower in the FSA since a filter-bank approach can be used and no audio-rate convolution with an impulse response must be performed. However, the system setup is much more complicated and computationally costly for the FSA than for the TDBEA.

B. Simulation accuracy

Since the frequency-domain FSA should be compared with the TDBEA with respect to their simulation accuracies, the comparison has to be performed in the time domain as well as in the frequency domain. Thus, strengths and weaknesses of both methods can be identified. First, FSA and TDBEA are compared in the time domain in Sec. IV B 1, then the frequency-domain comparison is performed in Sec. IV B 2.

1. Time domain

First of all, one has to carefully design the boundary excitation functions for the TDBEA since in some cases the simulation can become instable. This can be explained best with the realization of the bridge modes by recursive systems of second order. Even though the eigenfrequencies of the string as well as the modes of the bridge are realized with stable recursive systems of second order, the feedback of the string's output through the bridge mode system to the string's input result in fourth-order recursive systems. The poles of these combined second-order systems might lie outside the unit circle which results in instable discrete systems. Nevertheless, stable physical systems will produce stable models, but one cannot instantaneously decide whether a specific boundary condition may produce instabilities, while this is possible with the FSA. Furthermore, the TDBEA shows inaccuracies due to the limitation of the simulated eigenfrequencies of the string. This results in the effect that each boundary excitation has simultaneous impact to all spatial string positions which is not observed in the real string. For details of this effect see Ref. 9. The inaccuracies accumulate over time since they occur with each reflection of the ideal wave and thus with each non zero boundary excitation function. Fortunately, this effect is small enough to be neglected for audio applications.

In the FSA, the simplification of the transfer function $\mathbf{H}(\mu,s,x)$ to $\mathbf{H}(\mu,s_{\infty\mu,i},x)$ introduces transient effects as stated in Sec. III A 3. Similar to the frequency limitation effect described above, also these transient effects are small enough to be neglected for audio applications.

Furthermore, the stability can be guaranteed at any time in the FSA since the damping coefficients $\sigma_\mu < 0$ are calculated explicitly. Inaccuracies might only be introduced when using this method by an insufficient number of iterations for the derivation of the eigenvalues β_μ. An example with the parameters of the second example of Sec. III A 4 is shown in Fig. 9.

The left plot shows the elastic restoring force simulated with the FSA and similarities to Fig. 8 can be easily recognized. The differences between the two signals is shown in the right plot of Fig. 9. Only very small differences between both simulations are present before the first bridge reflection occurs at $t=1.2$ ms. These differences disperse over the entire string and are caused by an inaccurate calculation of the

Trautmann *et al.*: Strings with third kind boundary conditions
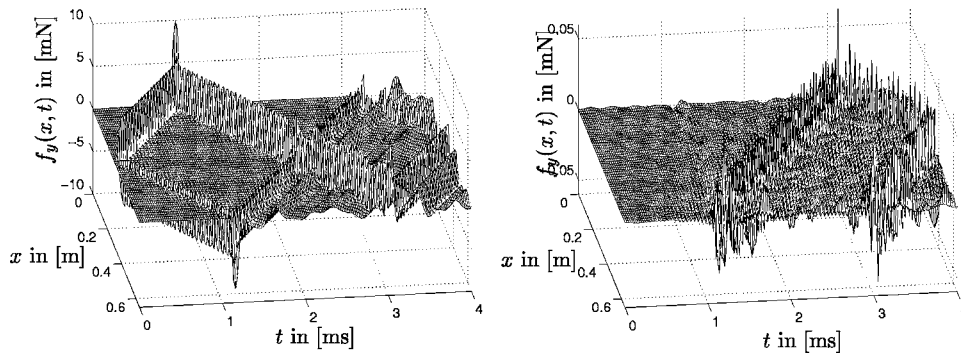
FIG. 9. Elastic restoring forces of a transversal vibrating string. Left plot: vibrations simulated with the FSA. Right plot: difference between frequency-shift simulations and time-domain boundary excitation simulations.

eigenvalues β_μ in the FSA as discussed above. At time instance $t = 1.2$ ms when the first reflection is taking place, an even higher signal difference on the entire string can be observed. This is caused by the TDBEA. This approach also results in inaccuracies of the reflected wave which can be seen for time instances later than 1.2 ms. However, the amplitude of the difference signal is less than 0.5% of the original signal. Thus, the simulation errors can be neglected for audio applications.

In conclusion, the FSA inherits more inaccuracies shortly after the excitation time, whereas the TDBEA has more inaccuracies after the first and later reflections of the ideal wave.

2. Frequency domain

In the frequency domain only small inaccuracies can be observed in both methods if only the onset of the output signal is taken into account. As mentioned above, the time-domain boundary excitation approach can become instable and the instability can be observed in the frequency domain. Even in stable systems the short-time spectra of the TDBEA vary over time since the inaccuracies of the impulse response used for the boundary excitation function accumulate over time.

In contrast, the eigenfrequencies of the FSA will not change over time. Hence, the short-time spectra only differ in the amplitude of the damped eigenfrequencies over time, not in their positions. Inaccuracies of the eigenfrequencies only occur if the iterative search of the eigenvalues β_μ, as described in Sec. III A 4, is not succeeding due to inaccurate starting positions or if only few iterations are performed for computational reasons. This effect is visible in Fig. 9.

C. Conclusions

It has been discussed that the audio-rate computational complexity of the FSA is much lower than that of the TDBEA, irrespective of the complexity of the bridge impedance. However, the computational cost for the system setup at control rate is much higher in the FSA compared to the TDBEA. This complexity even increases for physical systems that are more complicated than a system terminated with a fixed bridge impedance.

The simulation accuracy gives better results with the FSA, if eigenvalues β_μ are detected accurately. This task becomes problematic when considering more complicated physical systems.

V. SUMMARY AND CONCLUSIONS

In this article, the feasibility of the functional transformation method (FTM) for the simulation of string vibrations with boundary conditions of third kind is proven. Starting from a physical description in the form of a partial differential equation with initial and boundary conditions, two different approaches for the realization are given.

The frequency shift approach (FSA), on the one hand, introduces the boundary conditions of third kind in the eigenvalue problem. By means of an iterative algorithm, the deviation of the eigenfrequencies from the modes of a problem with boundary conditions of first kind are derived. With these numerically but accurately calculated modes, the remaining procedure of the FTM is identical to boundary conditions of first kind. The string is simulated efficiently by the summation of several second-order systems that can be realized with a filterbank approach.

Alternatively, the time-domain boundary excitation approach (TDBEA) introduces the boundary conditions of third kind at a later step. It uses the full analytical solution of the idealized model with boundary conditions of first kind for the simulation. The boundary conditions of third kind are then implemented in the algorithm by feedback of a superposed boundary-value-dependent signal that itself excites the ideal model.

The comparison of performance and accuracy shows the expected results. The FSA, as the more analytical approach, is more accurate due to its late discretization and it can be better optimized concerning the computational performance. The TDBEA realizes the boundary conditions of third kind only in the discrete implementation and therefore inherits more inaccuracies and is restricted to this particular discrete implementation. In conclusion the FSA is preferable for this specific application scenario. Nevertheless, the FSA is not always the method of choice for further developments, as it only shifts the complexity of the problem to the derivation of the algorithm. This approach will soon come to its limits when more complex problems, like coupled strings, are encountered. Combinations of both algorithms may be useful, to keep the effort to set up the algorithms below a certain level while maximizing performance and accuracy of the simulations.

APPENDIX: FREQUENCY-DOMAIN APPROACH OF THE BOUNDARY CONDITIONS SIMULATED WITH BOUNDARY EXCITATIONS

Compared to the frequency-shift approach (FSA) a disadvantage of the time-domain boundary excitation approach (TDBEA), as described in Sec. III B 1, is that its implementation is not as straightforward as the simple summation of harmonics for the FSA. Especially in the frequency domain the impact of including the output-dependent excitation function cannot be easily reconstructed. To overcome this obstacle, the coupling is now realized in the frequency domain. Inserting (28a)–(28c) into (28d) results in the total output:

$$\bar{y}^{\mathrm{d}}(\mu_T,k) = e^{-\tilde{\beta}_{\mu_T}T}\bar{y}^{\mathrm{d}}(\mu_T,k-1) + \widetilde{K}_1^*(\mu_T,l)Ta^{\mathrm{d}}(k) * f_y^{\mathrm{ideal}}(l,k)$$
$$+ \bar{v}(\mu_T,k), \tag{A1a}$$

$$= e^{-\tilde{\beta}_{\mu_T}T}\bar{y}^{\mathrm{d}}(\mu_T,k-1) + \widetilde{K}_1^*(\mu_T,l)Tb^{\mathrm{d}}(k) * f_y^{\mathrm{d}}(l,k)$$
$$+ \bar{v}(\mu_T,k) \tag{A1b}$$

$$= e^{-\tilde{\beta}_{\mu_T}T}\bar{y}^{\mathrm{d}}(\mu_T,k-1) + \widetilde{K}_1^*(\mu_T,l)Tb^{\mathrm{d}}(k)$$
$$* \sum_\nu \frac{1}{N_{\nu_T}}\bar{y}^{\mathrm{d}}(\nu_T,k)K_2(\nu_T,l) + \bar{v}(M_T,k), \tag{A1c}$$

with

$$B(s) = \frac{A(s)}{r(s)}, \quad b^{\mathrm{d}}(k) = \mathcal{Z}^{-1}\{B^{\mathrm{d}}(z)\}. \tag{A1d}$$

Since $f_y^{\mathrm{ideal}}(l,k)$ is not calculated explicitly it is replaced in (A1b) with the output $f_y^{\mathrm{d}}(l,k)$ where the idealized signal and the superposed signals are included. Both signals are related to each other by the reflection coefficient as defined in (11), such that $F_y^{\mathrm{d}}(l,s) = r(s)F_y^{\mathrm{ideal}}(l,s)$ in the Laplace domain. This relationship shows that the idealized vibration $f_y^{\mathrm{ideal}}(l,k)$ at the bridge position can be interpreted as the incoming signal. The reflected output signal is the sum of both, the idealized and the superposed signals, $f_y^{\mathrm{d}}(l,k)$.

Equation (A1c) is an implicit equation with respect to $\bar{y}^{\mathrm{d}}(\mu_T,k)$ since $\bar{y}^{\mathrm{d}}$ also appears in the sum on the right-hand side of the equation. However, it can be solved explicitly with respect to $\bar{y}^{\mathrm{d}}(\mu_T,k-1)$:

$$\bar{y}^{\mathrm{d}}(\mu_T,k-1) = e^{\tilde{\beta}_{\mu_T}T}\left(\bar{y}^{\mathrm{d}}(\mu,k) - \bar{y}^{\mathrm{d}}(\mu,k)\right.$$
$$\left. * \frac{T}{N_{\mu_T}}b^{\mathrm{d}}(k)K_2(\mu_T,l)\widetilde{K}_1^*(\mu_T,l)\right)$$
$$- e^{\tilde{\beta}_{\mu_T}T}\widetilde{K}_1^*(\mu_T,l)\sum_{\nu_T \neq \mu_T}\frac{T}{N_{\nu_T}}b^{\mathrm{d}}(k)$$
$$* \bar{y}^{\mathrm{d}}(\nu_T,k)K_2(\nu_T,l). \tag{A2}$$

The external excitation term $\bar{v}(\mu_T,k)$ is omitted here for concise notation. Equation (A2) can also be expressed in matrix notation $\mathbf{D}\bar{\mathbf{y}}_k = \bar{\mathbf{y}}_{k-1}$ with

$$\mathbf{D} = \begin{bmatrix} c_1 & d_{1,2} & \cdots & d_{1,N} \\ d_{2,1} & c_2 & \cdots & d_{2,N} \\ \vdots & & \ddots & \vdots \\ d_{N,1} & & \cdots & c_N \end{bmatrix},$$

$$\bar{\mathbf{y}}_k = \begin{bmatrix} \bar{y}^{\mathrm{d}}(1,k) \\ \bar{y}^{\mathrm{d}}(2,k) \\ \vdots \\ \bar{y}^{\mathrm{d}}(N,k) \end{bmatrix}, \quad \bar{\mathbf{y}}_{k-1} = \begin{bmatrix} \bar{y}^{\mathrm{d}}(1,k-1) \\ \bar{y}^{\mathrm{d}}(2,k-1) \\ \vdots \\ \bar{y}^{\mathrm{d}}(N,k-1) \end{bmatrix}. \tag{A3}$$

The constant functions c_μ and $d_{\mu,\nu}$ are given by

$$c_\mu = e^{\tilde{\beta}_{\mu_T}T}\left(1 - \frac{T}{N_{\mu_T}}b^{\mathrm{d}}(k)K_2(\mu_T,l)\widetilde{K}_1^*(\mu_T,l)\right), \tag{A4}$$

$$d_{\mu,\nu} = -e^{\tilde{\beta}_{\mu_T}T}\widetilde{K}_1^*(\mu_T,l)\frac{T}{N_{\nu_T}}b^{\mathrm{d}}(k)K_2(\nu_T,l). \tag{A5}$$

To solve for the output vector $\bar{\mathbf{y}}_k$, the matrix $\mathbf{D}$ must be inverted. The updated algorithm of the discrete realization in (A1c) is then replaced by

$$\bar{\mathbf{y}}_k = \mathbf{D}^{-1}\mathbf{y}_{k-1} + \bar{\mathbf{v}}_k. \tag{A6}$$

The external excitation is written likewise the other terms defined in (A3). Some properties can be extracted from the frequency-domain realization of the output-dependent excitation force (A5) and (A6):

(i) An interaction exists between all eigenfrequencies of the ideally terminated string as long as the coefficients $d_{\mu,\nu}$ are nonzero,

(ii) $d_{\mu,\nu}$ are only zero for $Z_B = \infty$, which corresponds to the ideally fixed end, and

(iii) the computational cost for the recursive systems is almost squared compared to the idealized case of the coupling matrix $\mathbf{D}^{-1}$.

The first item shows that the incorrectnesses of the modal shapes is fixed by additional excitations of all existing eigenfrequencies. This can be interpreted as a Fourier approximation of the desired eigenfrequency-basis with the basis of the idealized model. The procedure introduces inaccuracies of the resulting signal, particularly at the boundary, since the number of eigenfrequencies performing this correction is limited to the Nyquist frequency. This inaccuracy is obvious for the simulated signal at $x=l$, since all eigenfrequencies of the ideally fixed string are zero for $v_y(l,t)$, such that no combination of the eigenfrequencies can realize the desired nonzero values. The third item states that the computational cost of this frequency-domain solution is much higher than the TDBEA derived in Sec. III B 1. The number of MPOS is almost squared due to the coupling matrix $\mathbf{D}^{-1}$ and only the calculation of the additional output at the bridge position is saved compared to the previous method. Moreover, the inversion of the coupling matrix $\mathbf{D}$ might introduce numerical inaccuracies as long as finite precision arithmetic is used. Since c_μ and $d_{\mu,\nu}$ consists of impulse responses, in the case

of frequency-dependent bridge admittances the filter-bank realizations as discussed in Ref. 15 cannot be used here.

[1] L. Hiller and P. Ruiz, J. Audio Eng. Soc. **19**, 462–470 (1971).

[2] A. Chaigne and V. Doutaut, J. Acoust. Soc. Am. **101**, 539–557 (1997).

[3] C. Cadoz, A. Luciani, and J. Florens, Comput. Music J. **8**, 495–508 (1984).

[4] J. Adrien, in *Proc. AES 7th International Conf.* (Audio Engineering Society, Toronto, Canada, 1989), pp. 291–299.

[5] D. Jaffe and J. Smith, Comput. Music J. **7**, 56–69 (1983).

[6] M. Karjalainen, V. Välimäki, and T. Tolonen, Comput. Music J. **22**, 17–32 (1998).

[7] R. Rabenstein, in *Proc. of the EUROSIM CONGRESS*, Wien, 1995, pp. 177–182.

[8] M. Kahrs and K. Brandenburg, *Applications of Digital Signal Processing to Audio and Acoustics* (Kluwer Academic, Boston, 1998).

[9] L. Trautmann and R. Rabenstein, *Digital Sound Synthesis by Physical Modeling Using the Functional Transformation Method* (Kluwer Academic, New York, 2003).

[10] M. Karjalainen, in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'02)* (IEEE, New York, 2002), Vol. **2**, pp. 1869–1872.

[11] C. Erkut and M. Karjalainen, in *Proc. International Symposium on Musical Acoustics (ISMA 2002)*, Mexico City, Mexico, 2002.

[12] L. Trautmann and R. Rabenstein, in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, New York, 1999), pp. 83–86.

[13] R. Rabenstein and L. Trautmann, Signal Process. **83**, 1673–1688 (2003).

[14] L. Trautmann and V. Välimäki, in *Proc. Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (IEEE, New York, 2003), pp. 221–224.

[15] L. Trautmann and R. Rabenstein, J. Appl: Signal Process. **2004**, 949–963 (2004).

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

[17] P. Morse, *Vibration and Sound* (McGraw-Hill, New York, 1948).

[18] K. Legge and N. Fletcher, J. Acoust. Soc. Am. **76**, 5–12 (1984).

[19] T. Hill, B. Richardson, and S. Richardson, in *Proc. International Symposium on Musical Acoustics (ISMA)*, Perugia, Italy, 2001, pp. 425–429.

[20] C. Lambourg and A. Chaigne, in *Proc. Stockholm Music Acoustics Conf. (SMAC)* (Royal Swedish Academy of Music, Stockholm, Sweden, 1993), pp. 448–453.

[21] N. Giordano, J. Acoust. Soc. Am. **103**, 2128 (1998).

[22] R. Rabenstein and L. Trautmann, IEEE Trans. Circuits Syst., I: Fundam. Theory Appl. **49**, 852–861 (2002).

[23] C. Lanczos, *Linear Differential Operators* (Van Nostrand, London, 1961).

[24] B. Widrow and S. Stearns, *Adaptive Signal Processing* (Prentice–Hall, Englewood Cliffs, NJ, 1985).

[25] C. Gough, Acustica **49**, 124–141 (1981).

On waveguide modeling of stiff piano strings

Éric Ducasse[a]

Laboratoire de Mécanique Physique, UMR CNRS 5469, Université Bordeaux 1, 351 cours de la Libération, 33405 Talence Cedex, France

Bensa *et al.* [J. Acoust. Soc. Am. **114**, 1095–1107 (2003), Sec. IV] recently proposed a waveguide model for the transverse displacement of a stiff piano string. The study described here is an attempt to cast a complementary light on this topic, based on a common *wave approach* instead of a *modal approach*. A pair of weakly attenuated traveling waves and a pair of fast-decaying waves both satisfy the one-dimensional wave equation developed by Bensa *et al.* These solutions have to be carefully considered, however, for portions of string interacting with the hammer felt, the bridge, or the *capo d'astro* bar. © 2005 Acoustical Society of America. [DOI: 10.1121/1.1993127]

Pages: 1776–1781

I. INTRODUCTION

The first attempt for synthesizing musical sounds using physical models was made more than 30 years ago by Hiller and Ruiz.[1] Starting from the *one-dimensional wave equation*[2] of the transverse displacement of a string, three main approaches are generally used for building a computational model: either the use of a finite difference scheme (e.g., Refs. 3 and 4), or a modal synthesis (e.g., Ref. 5), or the building of a *digital waveguide* model (e.g., Refs. 6, 7, and 3). A *waveguide* is considered here as a "*filterlike structure modeling one-dimensional wave propagation as purely lossless throughout the length of the string, with loss and dispersion lumped in terminating filters*."[3] The hammer action and the reflection conditions at both ends of the string are also modeled as *lumped* filters. Portions of string are *distributed* elements[6] represented by two-port networks. In this paper we focus on this latter approach.

Bensa *et al.* propose an improved *one-dimensional wave equation* [Ref. 3, Eq. (6)] corresponding to a *well-posed* model of stiff piano strings. In Sec. IV B they use a *modal approach* to extract the waveguide parameters from the partial differential equation (PDE). This approach requires the knowledge of both the length of the string and the boundary conditions [Ref. 3, Eq. (10)] for the calculation of the vibration modes associated with standing waves. The waveguide parameters are then deduced from the characteristics of the first mode including the fundamental frequency of the ideal string. The present paper is an attempt to show that a *wave approach* using complex wave-numbers appears as more convenient because each dispersive attenuated traveling wave is directly obtained from the one-dimensional wave equation, apart from the reflections at the ends of the string. Assuming that the one-dimensional wave equation is linear and time and space invariant, a two-dimensional Fourier-Laplace transformation gives an algebraic equation relating the transverse displacement of the string (output) to the space distribution of external forces (input) in the *wave–number/frequency* domain. The solution of this equation can be seen as the superposition of traveling waves which occur in a waveguide model.

Some general points about waveguides, traveling waves, and complex wave numbers are outlined in Sec. II to help the reader to understand the following sections. In Sec. III, after the one-dimensional wave equation advanced by Bensa *et al.*[3] is recalled, this equation is solved to obtain the response of an infinite string to a point impulse in the *wave-number/frequency* domain, showing that two fast-decaying traveling waves are omitted by the modal approach in Bensa *et al.*[3] These fast-decaying waves can be neglected only for portions of string without sources which are long enough. They should, however, be taken into account at the neighborhood of the excitation point, as detailed in Sec. IV.

II. WAVEGUIDE MODELING AND COMPLEX WAVE NUMBERS

The question is this: how do we pass from the wave equation of a one-dimensional system to a *waveguide* model?

A. From the one-dimensional wave equation to the transfer function

1. A general one-dimensional wave equation

The vibration state of a one-dimensional system (e.g., the stiff piano string) is assumed to be defined by a function y (e.g., the transverse displacement of the string) of time t and position x along the system.

A general one-dimensional wave equation, assuming that it is linear and time and space invariant, can be written as

$$\sum_{m=0}^{d_x} \sum_{n=0}^{d_t} \alpha_{m,n} \frac{\partial^{m+n} y}{\partial x^m \, \partial t^n}(x,t) = f(x,t), \tag{1}$$

where f is the space distribution of external forces (e.g., exerted by the hammer felt, the bridge, and the *capo d'astro* bar).

2. Time/frequency and space/wave-number conversions

In the *space/frequency* domain, each signal $s(x,t)$ is converted into $\hat{s}(x,\omega)$ by Fourier transformation:

$$\hat{s}(x,\omega) = \int_{-\infty}^{+\infty} s(x,t)e^{-i\omega t}\, dt, \tag{2}$$

where ω is the angular frequency and $i^2 = -1$.

In the *wave-number/frequency* domain, each signal $s(x,t)$ becomes $S(k,\omega)$ by Fourier-Laplace transformation:

$$S(k,\omega) = \int_{-\infty}^{+\infty} \hat{s}(x,\omega)e^{-ikx}\, dx$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} s(x,t)e^{-i(\omega t + kx)}\, dt\, dx, \tag{3}$$

where k is the complex wave-number.

3. An ordinary differential equation in the space/frequency domain

The Fourier transformation of the wave equation (1) leads to an ordinary linear constant-coefficient differential equation with respect to x:

$$\sum_{m=0}^{d_x}\left[\sum_{n=0}^{d_t}\alpha_{m,n}(i\omega)^n\right]\frac{\partial^m \hat{y}}{\partial x^m}(x,\omega) = \hat{f}(x,\omega). \tag{4}$$

4. The transfer function

In the *wave-number/frequency* domain, we obtain

$$\left[\sum_{m=0}^{d_x}\sum_{n=0}^{d_t}\alpha_{m,n}i^{m+n}\omega^n k^m\right]Y(k,\omega)$$

$$= F(k,\omega) \Leftrightarrow Y(k,\omega) = H(k,\omega)F(k,\omega), \tag{5}$$

where $H(k,\omega)$ is the transfer function of the system. The denominator of $H(k,\omega)$ is a polynomial of order d_x in k and order d_t in ω. The poles of $H(k,\omega)$ [where $H(k,\omega)$ is considered as a function of k] are the roots of a polynomial equation commonly called the *dispersion equation*.

B. From the transfer function to the block-diagram model

1. The impulse response as the sum of pairs of symmetrical traveling waves

Assuming both that the propagation model corresponds to a well-posed physical problem (cf., e.g., Ref. 3 for more details) and that the problem is symmetrical (i.e., unchanged by substituting $-x$ for x), the coefficients $\alpha_{2k+1,n}$ are zero, which implies that $d_x = 2J$ is an even number, and the transfer function $H(k,\omega)$ is rewritten as follows:

$$H(k,\omega) = \frac{a(\omega)}{\displaystyle\prod_{j=1}^{J}[k^2 - k_j(\omega)^2]}$$

$$= \sum_{j=1}^{J}\frac{a_j(\omega)}{k^2 - k_j(\omega)^2}$$

$$= \sum_{j=1}^{J} g_j(\omega)\left[\frac{1}{k - k_j(\omega)} - \frac{1}{k + k_j(\omega)}\right], \tag{6}$$

where $a_j(\omega) = a(\omega)/\prod_{m=1,m\neq j}^{J}[k_m(\omega)^2 - k_j(\omega)^2]$ and $g_j(\omega) = a_j(\omega)/[2k_j(\omega)]$, assuming $k_j(\omega) \neq 0$ and $k_j(\omega) \neq k_m(\omega)$ if $j \neq m$

For an infinite system, the assumed boundary (no sources at infinity) and initial conditions (rest initial condition) are

$$y|_{x\to+\infty} = y|_{x\to-\infty} = 0; \quad y|_{t<0} = 0;$$
$$\tag{7}$$
$$\forall n, \quad 1 \leqslant n \leqslant d_t - 1, \quad \left.\frac{\partial^n y}{\partial t^n}\right|_{t<0} = 0.$$

Coming back to the *space/frequency* domain, Eqs. (6) and (7) lead to the convolution product with respect to the spatial variable:

$$\hat{y}(x,\omega) = \int_{-\infty}^{+\infty} \hat{f}(x_0,\omega)\hat{h}(x - x_0,\omega)\, dx_0, \tag{8}$$

$\hat{h}(x,\omega)$ satisfying

$$\hat{h}(x,\omega) = \sum_{j=1}^{J} g_j(\omega)[e^{-ik_j(\omega)x}u(x) + e^{+ik_j(\omega)x}u(-x)], \tag{9}$$

where u is the Heaviside unit step function:

$$\forall x < 0, \quad u(x) = 0 \quad \text{and} \quad \forall x \geqslant 0, \quad u(x) = 1.$$

The function $\hat{h}(x,\omega)$ corresponds to the Fourier transform of the impulse response $h(x,t)$ (also named Green's function) to $f(x,t) = \delta(x)\delta(t)$ where δ is the Dirac impulse. This response $h(x,t)$ is the effect of a time impulse (at $t=0$) exerted on a single point (at $x=0$) of the system. The unit of h is the unit of y per Newton and per second. An inverse Fourier transformation gives

$$h(x,t) = \frac{1}{2\pi}\sum_{j=1}^{J}\int_{-\infty}^{+\infty} g_j(\omega)\{e^{i[\omega t - k_j(\omega)x]}u(x)$$
$$+ e^{i[\omega t + k_j(\omega)x]}u(-x)\}\, d\omega. \tag{10}$$

Note that $h(x,t) = h(-x,t)$: the symmetry of the problem is satisfied.

Each $k_j(\omega)$ is a complex wave-number such that

$$k_j(\omega) = \omega\tau_j(\omega) - i\alpha_j(\omega) = \frac{\omega}{c_j(\omega)} - i\alpha_j(\omega), \quad (11)$$

where $\tau_j(\omega) \geq 0$ is the propagation delay per unit length (or slowness), $c_j(\omega)$ is the phase velocity, and $\alpha_j(\omega) \geq 0$ is the attenuation per unit length.

The impulse response can be rewritten as

$$h(x,t) = \frac{1}{2\pi}\sum_{j=1}^{J}\int_{-\infty}^{+\infty} g_j(\omega)\{e^{-\alpha_j(\omega)x}e^{i\omega[t-x/c_j(\omega)]}u(x)$$
$$+ e^{+\alpha_j(\omega)x}e^{i\omega[t+x/c_j(\omega)]}u(-x)\}\,d\omega. \quad (12)$$

This impulse response is consequently the superposition of J pairs of symmetrical decaying traveling waves starting from the excitation point and propagating in opposite directions.

Thus, the general solution y is given by the two-dimensional convolution:

$$y(x,t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x_0,t_0)h(x-x_0,t-t_0)\,dx_0\,dt_0. \quad (13)$$

2. Waveguide modeling of a region without sources

Under the assumption that no sources exist in a region between $x=x_{min}$ and $x=x_{max}$, one can demonstrate that the vibration state is the sum of J decaying traveling waves y_j^+ in the increasing x direction and of J decaying traveling waves y_j^- in the decreasing x direction:

$$\forall x, \quad x_{min} \leq x \leq x_{max}, \quad y(x,t) = \sum_{j=1}^{J} y_j^+(x,t) + \sum_{j=1}^{J} y_j^-(x,t),$$

where $\quad \widehat{y_j^+}(x,\omega) = e^{-ik_j(\omega)(x-x_{min})}\widehat{y_j^+}(x_{min},\omega) \quad (14)$

and $\quad \widehat{y_j^-}(x,\omega) = e^{+ik_j(\omega)(x-x_{max})}\widehat{y_j^-}(x_{max},\omega).$

This region can be modeled as J waveguides in parallel which become J digital waveguides[6,7] in discrete time. Each transfer function $e^{-ik_j(\omega)\Delta x}$ is generally designed as a delay line in series with a digital causal filter. After suitable digital filters have been found, efficient time-domain simulations can be made.

In the case of piano strings, the sources are located in the segment of hammer felt/string contact and at both ends. A block-diagram model is drawn in Fig. 1, including an observation point. This figure is a generalization of the model involved in Bensa et al. (Ref. 3, Fig. 2). Note that the boundary conditions characterizing the string terminations can be changed by only modifying one lumped filter at each end, apart from the waveguides modeling the portions of string without sources.

III. THE FOUR WAVES IN A STIFF PIANO STRING

In this section and the next one, we apply the principles given in the previous section to the one-dimensional wave
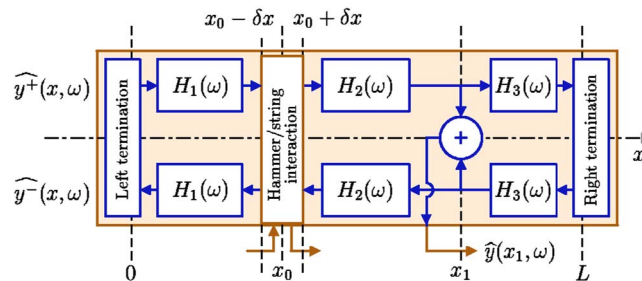
FIG. 1. Block diagram of the waveguide model of a piano string. The string segment interacting with the hammer felt is assumed to be $[x_0-\delta x;x_0+\delta x]$. x_1 is the position of observation. $x=0$ is the *capo d'astro* bar position and $x=L$ is the bridge position. Each $H_n(\omega)$ is equivalent to J delay line/filter blocks in parallel: $H_1(\omega) = \sum_{j=1}^{J} e^{-ik_j(\omega)(x_0-\delta x)}$, $H_2(\omega) = \sum_{j=1}^{J} e^{-ik_j(\omega)(x_1-x_0-\delta x)}$, and $H_3(\omega) = \sum_{j=1}^{J} e^{-ik_j(\omega)(L-x_1)}$.

equation introduced by Bensa et al. [Ref. 3, Eq. (6)]. The dispersion equation provides two pairs of decaying traveling waves which need a detailed analysis.

A. The transfer function of a stiff piano string

The transverse displacement y is assumed to satisfy [Ref. 3, Eq. (6)]

$$\frac{\partial^2 y}{\partial t^2} + 2b_1\frac{\partial y}{\partial t} - 2b_2\frac{\partial^3 y}{\partial x^2\,\partial t} - c^2\frac{\partial^2 y}{\partial x^2} + \kappa^2\frac{\partial^4 y}{\partial x^4} = \frac{1}{\mu}f, \quad (15)$$

where b_1 is the first coefficient of damping (due to viscosity of the air), b_2 is the second damping coefficient (inner losses), $c = \sqrt{T/\mu}$ is the transverse wave velocity of string,[4] T is the string tension, μ is the linear mass density of string, $\kappa = c\sqrt{EI/T}$ is a stiffness coefficient, E is the Young modulus, $I = Sr^2$ is the moment of inertia of the cross-section, S is the cross-section, and r is the radius of gyration.

In the *wave-number/frequency* domain, Eqs. (5) and (15) lead to the transfer function:

$$H(k,\omega) = \frac{Y(k,\omega)}{F(k,\omega)}$$

$$= \frac{1}{\mu[(-\omega^2 + 2i\omega b_1) + (c^2 + 2i\omega b_2)k^2 + \kappa^2 k^4]}. \quad (16)$$

B. Complex wave numbers

The two ($J=2$) wave numbers k_s and k_d satisfy

$$k_s^2(\omega) = \frac{c^2}{2\kappa^2}\left[-1 - i\frac{2\omega}{c^2}b_2\right.$$
$$\left. + \sqrt{1 + \frac{4(\kappa^2 - b_2^2)\omega^2}{c^4} + 4i\omega\left(\frac{1}{c^2}b_2 - 2\frac{\kappa^2}{c^4}b_1\right)}\right], \quad (17)$$

$$k_d^2(\omega) = \frac{c^2}{2\kappa^2}\left[-1 - i\frac{2\omega}{c^2}b_2\right.$$
$$\left. - \sqrt{1 + \frac{4(\kappa^2 - b_2^2)\omega^2}{c^4} + 4i\omega\left(\frac{1}{c^2}b_2 - 2\frac{\kappa^2}{c^4}b_1\right)}\right].$$

$$(18)$$

These equations are similar to Eq. (11) in Ref. 3. The approaches, however, are different. In the modal approach[3] β_+ and β_- are real-valued functions of the complex frequency s. In the present wave approach k_s and k_d are the roots of the *dispersion equation* and are complex-valued functions of the common angular frequency ω (real number).

The assumptions [Ref. 3, Eqs. (24) and (25)]

$$b_1 b_2 \ll c^2; \quad b_2^2 \ll \kappa^2 \text{ and } b_1^2 \ll \omega^2 \qquad (19)$$

lead to

$$k_s(\omega) \simeq \frac{c}{\kappa}\sqrt{\frac{\xi}{2}} - i\frac{\kappa\omega}{c^3(1+\xi)}\sqrt{\frac{2}{\xi}}\left(b_1 + \frac{c^2\xi}{2\kappa^2}b_2\right) \qquad (20)$$

and

$$k_d(\omega) \simeq \frac{1}{c^3\sqrt{2\xi}(1+\xi)}(2\omega^2 b_2 - c^2\xi b_1) - i\sqrt{\frac{2}{\xi}}\frac{\omega}{c}, \qquad (21)$$

with

$$\xi = -1 + \sqrt{1 + \frac{4\kappa^2}{c^4}\omega^2}. \qquad (22)$$

C. Waveguide modeling of portions of string without sources

The first wave number k_s characterizes a weakly attenuated dispersive traveling wave similar to the one described by Bensa *et al.* [Ref. 3, Eq. (34)]. At low frequencies [$b_1^2 \ll \omega^2 \ll c^4/(4\kappa^2)$] the phase velocity is close to c. Its numerical values are between 160 and 420 m s^{-1} for piano tones C2, C4, and C7 (Ref. 3, Table I). The attenuation per unit length $\alpha_s \simeq b_1/c$ is small: its numerical values [$10\alpha_s/\log(10)$] are less than 0.1dB m^{-1}.

The second wave number k_d represents a fast-decaying wave with a very high phase velocity. At low frequencies the phase velocity is approximately $c\kappa/(b_2 - \kappa^2 b_1/c^2)$. Its numerical values are in the range 2.6×10^5 to 1.62×10^6 m s^{-1}. The attenuation per unit length is $\alpha_d \simeq c/\kappa$. The numerical values [$10\alpha_d/\log(10)$] of α_d are greater than 1.1×10^3dB m^{-1}. For any length Δx greater than 1.8 cm, $e^{-\alpha_d\Delta x}$ (the modulus of $e^{-ik_d\Delta x}$) is less than 10^{-2}. This second wave is *quasi-evanescent* i.e., its phase velocity is almost infinite and its attenuation per unit length is high. It could represent the fact that a part of the energy of a hammer strike would instantly propagate along the stiff string around the contact region such as to avoid the formation of a sharp corner, as shown below and in accordance with Cremer.[8]

Consequently, the waveguide model of a portion of string without sources will contain only a single delay line/filter (wave number k_s) for each direction of propagation, provided that the portion of string is longer than a few cen-

timeters. But the other wave characterized by k_d cannot be neglected at the neighborhood of the excitation region, as shown in the following section.

IV. THE IMPULSE RESPONSE OF THE STRING

A. Space/frequency domain

In the space/frequency domain Eqs. (16)–(18), (6), and (9) imply that the Fourier transform of the impulse response is

$$\hat{h}(x,\omega) = \begin{cases} \widehat{h_s^+}(0,\omega)e^{-ik_s(\omega)x} + \widehat{h_d^+}(0,\omega)e^{-ik_d(\omega)x}, & x > 0, \\ \widehat{h_s^+}(0,\omega) + \widehat{h_s^-}(0,\omega) + \widehat{h_d^+}(0,\omega) + \widehat{h_d^-}(0,\omega), & x = 0, \\ \widehat{h_s^-}(0,\omega)e^{+ik_s(\omega)x} + \widehat{h_d^-}(0,\omega)e^{+ik_d(\omega)x}, & x < 0, \end{cases}$$

$$(23)$$

where

$$\widehat{h_s^+}(0,\omega) = \widehat{h_s^-}(0,\omega) = g_s(\omega) = \frac{-i}{2\mu\kappa^2[k_s(\omega)^2 - k_d(\omega)^2]k_s(\omega)},$$

$$(24)$$

$$\widehat{h_d^+}(0,\omega) = \widehat{h_d^-}(0,\omega) = g_d(\omega) = \frac{i}{2\mu\kappa^2[k_s(\omega)^2 - k_d(\omega)^2]k_s(\omega)}.$$

The complex compliances $g_s(\omega)$ and $g_d(\omega)$ (unit: m N^{-1}) characterize the conversion of force to motion.

B. Example of a C2 piano string

The response y of a C2 piano string (Ref. 3, Table I) to a downward Gaussian time impulse at a single point $x=0$ is given by Fig. 2. This response is very close to the negative of the impulse response h in the bandwidth 0–20 kHz. The transverse displacement y is the sum of two weakly attenuated traveling waves $y_s^\pm$ (wave numbers $\pm k_s$) and of two fast-decaying waves $y_d^\pm$ (wave numbers $\pm k_d$) going away symmetrically from the excitation point. In agreement with Refs. 8 and 2 (Sec. 2.18), "*the bend is rounded appreciably by the stiffness of the string*." The fast-decaying waves inhibit the sharp corner which could be generated by the weakly attenuated traveling waves, if the former were not considered. As shown in the previous section, only the weakly attenuated traveling waves exist far from the excitation point (more than 2 cm), as well as in the neighborhood of the excitation point after 0.1 ms.

C. Noncausal components of the causal impulse response

An important additional point is that $y_s^\pm$ and $y_d^\pm$ are not causal responses whereas their sum is causal. A mathematical explanation can be found for this by focusing on the impulse response at the excitation point [Eq. (24)]. The signals $h_s^\pm(0,t)$ and $h_d^\pm(0,t)$ and the modulus of their Fourier transforms $g_s(\omega) = \widehat{h_s^\pm}(0,\omega)$ and $g_d(\omega) = \widehat{h_d^\pm}(0,\omega)$ are plotted in Fig. 3 (in the case of a C2 piano string). The latter signal $h_d^\pm(0,t)$ is nearly an even function (with respect to t). The former signal $h_s^\pm(0,t)$ is nearly the sum of a constant [$1/(4\mu c)$] and of an odd function.
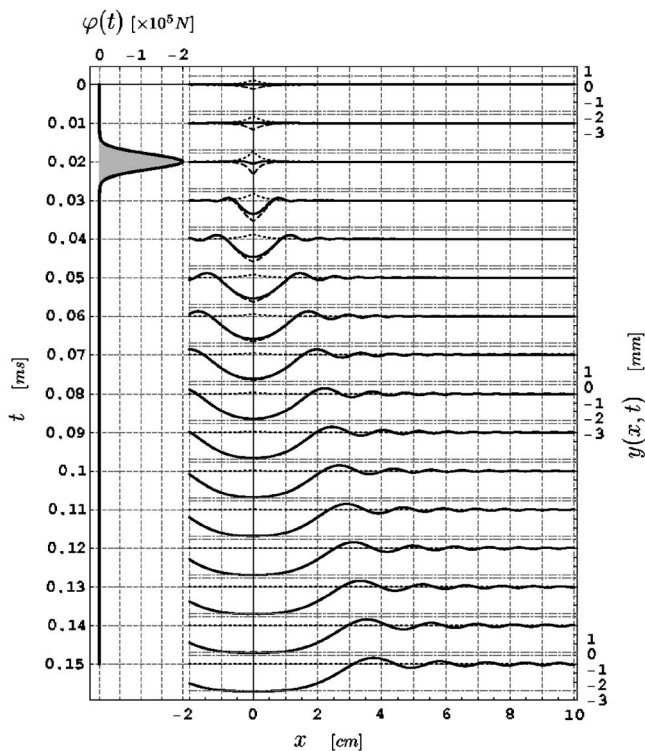
FIG. 2. Response of a C2 piano string to a downward Gaussian impulse f. $f(x,t)=f_0\delta(x)\varphi(t)$ where $\varphi(t)=-\lambda e^{-\pi\lambda^2(t-\tau)^2}$. $f_0=1N$, $\lambda=2\times10^5$ s^{-1}, and $\tau=0.02$ ms. The parameters of the C2 piano string are (Ref. 3, Table I) $c=160.9$ m s^{-1}, $\kappa=0.58$ m^2 s^{-1}, $b_1=0.25$ s^{-1}, and $b_2=7.5\times10^{-5}$ m^2 s^{-1}. The response y (plain line) is the sum of a weakly attenuated wave y_s (dashed line) and of a quasi-evanescent wave y_d (dotted line).

(a)

(b)
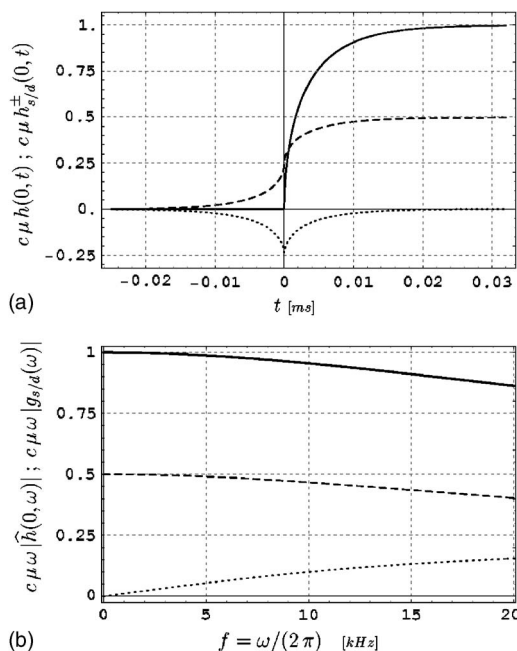
FIG. 3. (a) Impulse responses of a C2 piano string and (b) the modulus of their Fourier transforms times the angular frequency. The compliance $g_s(\omega)=\widehat{h_s^{\pm}}(0,\omega)$ is the Fourier transform of the signal $h_s^{\pm}(0,t)$ (weakly attenuated waves, dashed line). $g_d(\omega)=\widehat{h_d^{\pm}}(0,\omega)$ is the Fourier transform of $h_d^{\pm}(0,t)$ (quasi-evanescent waves, dotted line). The impulse response $h(0,t)$ (plain line) is two times the sum of $h_s^{\pm}(0,t)$ and $h_d^{\pm}(0,t)$. Its Fourier transform is $\hat{h}(0,\omega)=2[g_s(\omega)+g_d(\omega)]$.
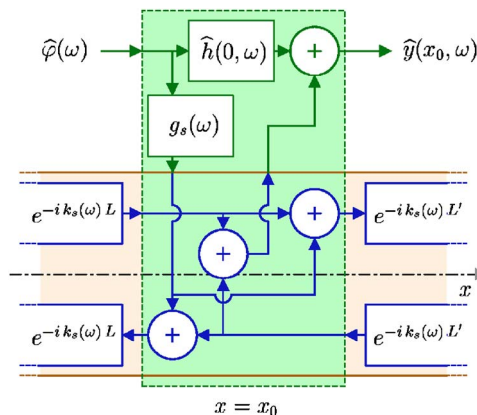
FIG. 4. Block diagram of a portion of string excited by a point force. This force $\varphi(t)$ is exerted by the hammer felt at the position $x=x_0$. The force distribution is assumed to be $f(x,t)=\varphi(t)\delta(x-x_0)$ for x in the range x_0-L to x_0+L' (L and L' are both greater than a few centimeters).

This is natural because, on assuming that b_1 and b_2 are zero, and ω is greater than zero [see Eqs. (17), (18), (22), and (24)], the compliances become

$$g_s(\omega)\simeq\frac{ik}{\mu c^3\sqrt{2\xi}(1+\xi)} \text{ and } g_d(\omega)\simeq\frac{-\sqrt{\xi}}{2\mu\sqrt{2}c(1+\xi)\omega}.$$
(25)

.

Consequently, the compliance g_s is an imaginary-valued function and g_d is a real-valued function. Because the signals $h_s^{\pm}(0,t)$ and $h_d^{\pm}(0,t)$ are real-valued, the former should be an odd function and the latter an even function. But the Fourier transform of $h_s^{\pm}(0,t)$ has a singularity at $\omega=0$ and the limit of $h_s^{\pm}(0,t)$, when t tends to $-\infty$, has to be zero. This implies that a constant has to be added to $h_s^{\pm}(0,t)$, giving an additional term $\pi\delta(\omega)/(2\mu c)$ in its Fourier transform.

Applying the final value theorem to the signal $h_s^{\pm}(0,t)$, the limit of $h_s^{\pm}(0,t)$ with t tending to infinity is $1/(2\mu c)$ (see Fig. 3). The final value of the impulse response $h(0,t)$ is similarly $1/(\mu c)$. This last property is false if $b_1>0$: all the signals tend slowly to zero when t tends to infinity. Indeed, $h_d^{\pm}(0,t)$ is close to zero for great values of t and $h(0,t)\simeq h_s^{+}(0,t)+h_s^{-}(0,t)$.

In summary, the two waves $h_s^{\pm}$ and $h_d^{\pm}$ have to be considered together at the position $x=0$ to avoid nonphysical effects with respect to time. The conversion of force to motion is characterized by a single compliance $\hat{h}(0,\omega)=2[g_s(\omega)+g_d(\omega)]$, which is plotted in Fig. 3.

This implies that if the force distribution $f(x,t)=\delta(x-x_0)\varphi(t)$ is assumed to represent a point force exerted by the hammer felt at the position $x=x_0$, then the response $y(x_0,t)$ at the excitation point satisfies $\hat{y}(x_0,\omega)=2[g_s(\omega)+g_d(\omega)]\hat{\varphi}(\omega)$. This case is represented in Fig. 4 as a block diagram. Similar block diagrams are used for bowed strings (cf., e.g., the review article by Smith,[9] Fig. 14).

V. CONCLUSION AND PROSPECTS

The conclusion of this paper is that a complete mathematical treatment of the one-dimensional wave equation

(15) carefully takes into account all the waves provided by the dispersion equation. Fast-decaying waves can be neglected in the waveguide modeling of portions of string without sources if they are long enough. However, they cannot accurately be neglected at the neighborhood of the excitation region (hammer) and of the fixation points (bridge, *capo d'astro* bar).

In particular, if a point force is exerted by the hammer felt on the string, the waveguide model proposed by Bensa *et al.* (Ref. 3, Fig. 2) needs to be complemented at the excitation point, as shown in Fig. 4 which summarizes the results of the last section. Note that a realistic excitation by the hammer is distributed over a certain width.[4] This extension could be made by a spatial convolution. A future study may determine whether or not this correction would noticeably improve the quality of synthesized stiff-string sounds.

This paper is an attempt to extract as much information as possible from the physical model defined by a one-dimensional wave equation, in a *waveguide modeling* context and from a mathematical point of view. It is hoped that this study will help to improve future waveguide models of strings and to enhance the accuracy of efficient computational models using *digital waveguides*,[6,7,3] after a substantial effort to design suitable digital filters.

[1] L. Hiller and P. Ruiz: "Synthesizing musical sounds by solving the wave equation for vibrating objects," J. Audio Eng. Soc. **19**(6), 462–470 (Part I) (1971); **19**(7), 542–551 (Part II) (1971).

[2] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, 1999).

[3] J. Bensa, S. Bilbao, R. Kronland-Martinet, and J. O. Smith III, "The simulation of piano string vibration: from physical model to finite difference schemes and digital waveguides," J. Acoust. Soc. Am. **114**, 1095–1107 (2003).

[4] A. Chaigne and A. Askenfelt, "Numerical simulations of struck strings. I. A physical model for string using finite difference methods," J. Acoust. Soc. Am. **95** 1112–1118 (1994).

[5] J.-M. Adrien, "The Missing Link: Modal Synthesis," in *Representations of Musical Signals* (MIT, Cambridge, 1991), pp. 269–297.

[6] J. O. Smith III, "Physical modeling using digital waveguides," Comput. Music J. **16**(4), 74–91 (1992).

[7] J. O. Smith III, "Physical modeling synthesis update," Comput. Music J. **20**(2), 44–56 (1996).

[8] L. Cremer, *The Physics of Violin*, translation by J. S. Allen (MIT, Cambridge, MA 1984).

[9] J. O. Smith III, "Virtual acoustic musical instruments: Review and update," J. New Music Res. **33**(3), 283–304 (2004), available online at http://ccrma.stanford.edu/~jos/jnmr/.

Simulation of ultrasound propagation through bovine cancellous bone using elastic and Biot's finite-difference time-domain methods

A. Hosokawa

Department of Electrical and Computer Engineering, Akashi National College of Technology,
679-3 Nishioka, Uozumi, Akashi, 674-8501 Hyogo, Japan

The propagation of ultrasonic pulse waves in bovine cancellous bone has been numerically analyzed in two dimensions by using two finite-difference time-domain (FDTD) methods: the commonly used elastic FDTD method and an FDTD method extended with Biot's theory for a porous elastic solid saturated with viscous fluid. Both FDTD results were compared with the results of previous experiments by Hosokawa and Otani [J. Acoust. Soc. Am. **101**, 558–562 (1997)], in which the Biot's fast and slow longitudinal waves were clearly identified. It was difficult to analyze both the fast and slow waves with the elastic FDTD method because of the inadequate 2D model of cancellous bone. On the other hand, in Biot's FDTD results that consider the pore fluid motion relative to the solid frame, both waves could be clearly observed. The experimental and simulated values of the speeds of these waves were in good agreement. Using the modified Biot's FDTD equations containing the possible attenuations for the fast wave other than the viscous loss due to the pore fluid motion, the amplitude ratio of the slow wave to the fast wave largely increased with the porosity, which also agrees with the experimental results. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.2000767]

I. INTRODUCTION

Cancellous bone is comprised of a skeletal frame formed by numerous rod-like and plate-like bone elements (namely trabeculae), with soft tissue (bone marrow) in the pore spaces. The structural arrangement of the trabeculae changes with the porosity.[1,2] At low porosity, a network of plate-like trabeculae with few rods, i.e., plate–plate-type trabeculae, are formed. As the porosity increases, the plate–plate-type trabeculae transform into plate–rod-type trabeculae, forming an open network. This structure is formed of rod-like trabeculae between parallel plates and has a strong structural anisotropy. At higher porosity, a more open rod–rod-type structure is formed primarily of rod-like trabeculae. Thus, the structure in cancellous bone is extremely complex. This complexity has made it difficult to theoretically analyze the propagation phenomena of ultrasounds in cancellous bone.

In previous research,[3–5] ultrasound propagation in bovine cancellous bone was examined experimentally. In experiments, both the fast and slow longitudinal waves could be clearly observed in the direction parallel to the trabecular alignment. It was then shown that these waves corresponded to "waves of the first and second kinds" as predicted in Biot's theory.[6,7] In the present study, two-dimensional analysis of fast and slow wave propagation in cancellous bone has been performed numerically using two finite-difference time-domain (FDTD) methods. One is the commonly used elastic FDTD method for elastic wave propagation in an elastic medium, and the other is an FDTD method extended with Biot's theory for propagation in a two-component medium comprised of a porous solid frame and pore fluid. In the elastic FDTD method, cancellous bone was modeled as numerous

trabecular rods and bone marrow in these spaces. The model used in Biot's FDTD method, on the other hand, was based on Biot's theory for an isotropic medium, in which the viscous loss caused by the pore fluid motion relative to the solid frame[6,7] was taken into account. The fast and slow waveforms were simulated by both FDTD methods, and the wave speeds and the amplitude ratio of the slow wave to the fast wave were compared with previous experimental results.

II. FDTD METHODS

A. Acoustic FDTD method

The propagation of acoustic waves in a 2D fluid is governed by the following three equations:

$$-\rho\frac{\partial \dot{u}_x}{\partial t} = \frac{\partial p}{\partial x}, \tag{1a}$$

$$-\rho\frac{\partial \dot{u}_y}{\partial t} = \frac{\partial p}{\partial y}, \tag{1b}$$

$$-\frac{1}{K}\frac{\partial p}{\partial t} = \frac{\partial \dot{u}_x}{\partial x} + \frac{\partial \dot{u}_y}{\partial y}, \tag{1c}$$

where u_l $(l=x,y)$ are the particle displacement components of the fluid in the l direction (an overdot denotes a time derivative), p is the sound pressure, ρ is the fluid density, and K is the bulk modulus.

As functions of discrete space and time, the FDTD equations are derived by discretizing $\dot{u}_l$ and p, and by approximating partial differential equations (1a)–(1c) using a central differencing scheme with staggered grids. A leap-frog

system[8] is then constituted for $\dot{u}_l$ and p. The values of $\dot{u}_l$ and p for the entire set of spatial points at a discrete time step can be solved by substituting these values at the previous step for the FDTD equations with distinct values of the medium parameters (ρ and K) at each point. Given a sound source excitation, this system can provide the pressure field at any time.

B. Elastic FDTD method

The propagation of elastic waves in a 2D elastic medium is governed by the following five equations:

$$\rho \frac{\partial \dot{u}_x}{\partial t} = \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y}, \tag{2a}$$

$$\rho \frac{\partial \dot{u}_y}{\partial t} = \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y}, \tag{2b}$$

$$\frac{\partial \tau_{xx}}{\partial t} = (\lambda + 2\mu) \frac{\partial \dot{u}_x}{\partial x} + \lambda \frac{\partial \dot{u}_y}{\partial y}, \tag{2c}$$

$$\frac{\partial \tau_{yy}}{\partial t} = \lambda \frac{\partial \dot{u}_x}{\partial x} + (\lambda + 2\mu) \frac{\partial \dot{u}_y}{\partial y}, \tag{2d}$$

$$\frac{\partial \tau_{xy}}{\partial t} = \mu \left(\frac{\partial \dot{u}_x}{\partial y} + \frac{\partial \dot{u}_y}{\partial x} \right), \tag{2e}$$

where u_l are the particle displacement components of the medium in the l direction, τ_{ll} are the normal stress components in the l direction, τ_{xy} is the shear stress component on the $x-y$ plane, ρ is the medium density, and λ and μ are, respectively, the first and second Lamé coefficients. Central differencing for Eqs. (2a)–(2e) is used to constitute a leap-frog system for $\dot{u}_l$ and τ_{ll} and τ_{xy}.

C. Biot's FDTD method

In Biot's theory,[6,7] the elastic wave propagation in a fluid-saturated porous medium is described by considering the dynamic coupling between the solid skeletal frame and pore fluid. This theory predicts two distinct longitudinal waves propagating through such a medium, which are defined by fast and slow longitudinal waves (denoted as waves of the first and second kinds by Biot). The fast wave corresponds to in-phase motion of the solid frame and pore fluid, and the slow wave corresponds to their out-of-phase motion. Wave propagation in a 2D porous medium is governed by the following eight equations:

$$\frac{\partial}{\partial t} (\rho \dot{u}_x + \rho_f \dot{w}_x) = \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y}, \tag{3a}$$

$$\frac{\partial}{\partial t} (\rho \dot{u}_y + \rho_f \dot{w}_y) = \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y}, \tag{3b}$$

$$\frac{\partial}{\partial t} (\rho_f \dot{u}_x + m \dot{w}_x) = -\frac{\partial p_f}{\partial x} - \frac{F_r \eta}{k} \dot{w}_x, \tag{3c}$$

$$\frac{\partial}{\partial t} (\rho_f \dot{u}_y + m \dot{w}_y) = -\frac{\partial p_f}{\partial y} - \frac{F_r \eta}{k} \dot{w}_y, \tag{3d}$$

$$\frac{\partial \tau_{xx}}{\partial t} = H \frac{\partial \dot{u}_x}{\partial x} + (H - 2\mu) \frac{\partial \dot{u}_y}{\partial y} + C \left(\frac{\partial \dot{w}_x}{\partial x} + \frac{\partial \dot{w}_y}{\partial y} \right), \tag{3e}$$

$$\frac{\partial \tau_{yy}}{\partial t} = (H - 2\mu) \frac{\partial \dot{u}_x}{\partial x} + H \frac{\partial \dot{u}_y}{\partial y} + C \left(\frac{\partial \dot{w}_x}{\partial x} + \frac{\partial \dot{w}_y}{\partial y} \right), \tag{3f}$$

$$\frac{\partial \tau_{xy}}{\partial t} = \mu \left(\frac{\partial \dot{u}_x}{\partial y} + \frac{\partial \dot{u}_y}{\partial x} \right), \tag{3g}$$

$$\frac{\partial p_f}{\partial t} = -C \left(\frac{\partial \dot{u}_x}{\partial x} + \frac{\partial \dot{u}_y}{\partial y} \right) - M \left(\frac{\partial \dot{w}_x}{\partial x} + \frac{\partial \dot{w}_y}{\partial y} \right), \tag{3h}$$

where u_l are the particle displacement components of the solid frame in the l direction, w_l are the displacement components of the pore fluid relative to the solid frame, τ_{ll} are the normal stress components of the solid frame in the l direction, τ_{xy} is the shear stress component on the $x-y$ plane, p_f is the pressure in the pore fluid, ρ is the total density of the fluid-saturated medium, and ρ_f is the fluid density. In addition, m is the density parameter, which is determined by the relation

$$m = \frac{\alpha \rho_f}{\beta}, \tag{4}$$

where α, which is denoted as the structural factor, is dependent on the pore geometry, and β is the porosity. The last terms $(F_r \eta / k) \dot{w}_l$ on the right-hand side in Eqs. (3c) and (3d) represent the viscous resistance to the fluid motion, where η is the fluid viscosity, k is the permeability, and F_r is a correlation factor to consider the frequency dependence of the viscous resistance. Finally, H, C, and M are generalized elastic coefficients deduced by Biot, and can be expressed in terms of the bulk moduli of the solid and fluid, the bulk and shear moduli μ of the solid frame, and the porosity. A leap-frog system is constituted for $\dot{u}_l$, $\dot{w}_l$ and τ_{ll}, τ_{xy}, and p_f.

III. NUMERICAL IMPLEMENTATION

A. FDTD analysis model

Ultrasound propagation through bovine femoral cancellous bone in the direction parallel to the trabecular alignment, which had been experimentally investigated by Hosokawa and Otani,[3] was numerically analyzed using both the elastic and Biot's FDTD methods. In order to reduce the memory requirement, the numerical analysis was performed in 2D rather than 3D. The geometry of the FDTD analysis model is shown in Fig. 1. As shown in this figure, the spatial region is 100×50 mm on the $x-y$ plane. A transmitter and a receiver (10 and 8 mm in size, respectively) separated by a distance of 75 mm, and the cancellous bone specimen (9 $\times 40$ mm) were set in this region. The transmitting and receiving lines were flat (not focused), and the backing materials were assumed to be perfectly rigid. As the excitation condition, pulsed particle displacements in the x direction
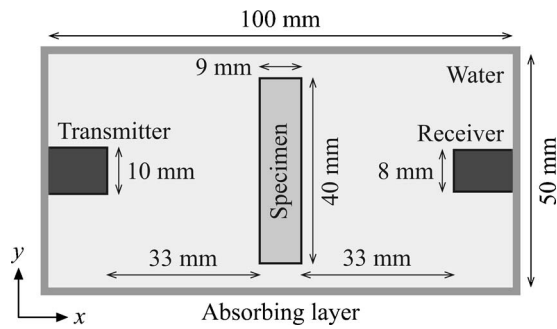
FIG. 1. Schematic drawing of the FDTD analysis model for ultrasound propagation in the cancellous bone specimen.

were given at the corresponding points on the transmitting line. The time function is a single sinusoid multiplied by a Hamming window

$$u_x(t) = U_0 \sin(2\pi f_0 t)[0.54 - 0.46 \cos(2\pi f_0 t)]$$

$$\text{at } 0 \leqslant t \leqslant 1/f_0, \tag{5}$$

where U_0 is a constant and f_0 is a frequency. The output was calculated as the summation of the sound pressures at the receiving points. In order to eliminate the reflections from the boundaries surrounding the analysis region, Higdon's second-order absorbing boundary conditions[9] were adopted at these boundaries. The spatial intervals Δx and Δy in the x- and y directions were the same at 25 μm (the computational domain consisted of 4000×2000 grids), and the time interval Δt was 2.5 ns. For fewer unknowns, the acoustic FDTD method was used for the calculation in the water region and the elastic or Biot's FDTD method in the specimen region. The bulk modulus $K_w = 2.0$ GPa and density $\rho_w = 1000$ kg/m^3 of water were given in the acoustic FDTD method.

To verify the FDTD analysis model, except for the cancellous bone model, the pulsed waveform propagating in water without any specimen was simulated and compared with the previous experimental result reported in Ref. 3. In the experiments, the transmitter was driven by a single sinusoidal voltage at 1 MHz. On the other hand, in the simulations, Eq. (6) with $f_0 = 0.75$ MHz was given by considering the transmitting and receiving characteristics in the experimental settings. The experimental and simulated results are shown in Figs. 2(a) and 2(b), respectively. These waveforms differ slightly in the rise-up and rise-down parts, but generally are in agreement.

B. Elastic cancellous bone model

A schematic drawing of the cancellous bone model used in the elastic FDTD method is shown in Fig. 3. The skeletal frame of cancellous bone was made of numerous trabecular rods. The porosity β was adjusted by changing the trabecular thickness a_s and pore space between trabeculae a_f, and the trabecular arrangement was adjusted by changing the trabecular length l_s and angle θ. In order to form an inhomogeneous trabecular structure, these values were uniformly distributed in the range of $\pm 25\%$ in each of a_s, a_f, and l_s, and $\pm 10°$ in θ. It was assumed that the pore spaces were perfectly
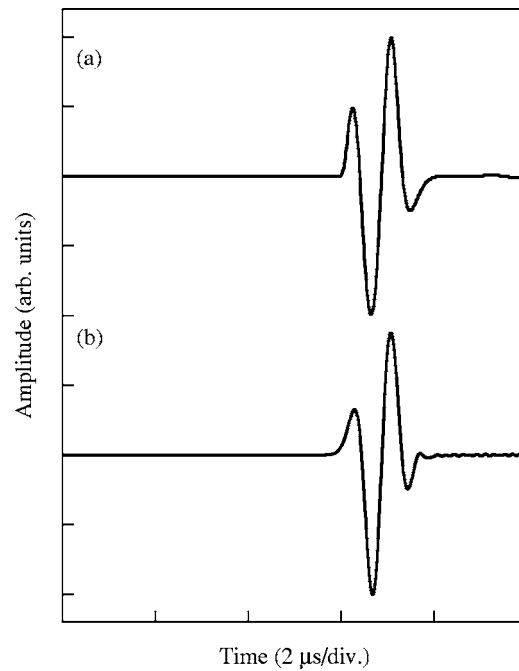
FIG. 2. Pulsed waveforms propagating in water without any cancellous bone specimen: (a) experimental result (Ref. 3); (b) simulated result obtained using the acoustic FDTD method.

filled with bone marrow. The corresponding values of the medium parameters (ρ, λ, and μ) were given at each spatial point in this model.

Assuming bovine femoral cancellous bone with trabeculae parallel to the thickness direction (the plate–rod-type trabeculae described in Sec. I), which was used in previous experiments, the mean values of a_s, a_f, l_s, and θ were determined. The values for the low- and high-porosity models are listed in Table I, and geometrical drawings of these models are shown in Fig. 4. The porosities β for the low- and high-porosity models are 0.75 and 0.83, respectively, which are identical to the experimental values. The physical parameters of bovine cancellous bone are listed in Table II. The parameters for the trabeculae were assumed to be those of cortical bone. The first and second Lamé coefficients $\lambda_s = 14.8$ GPa and $\mu_s = 8.3$ Gpa, respectively, were calculated from the Young's modulus $E_s = 22.0$ GPa and Poisson's ratio $\nu_s = 0.32$ by Williams and Johnson.[10] The density $\rho_s = 1960$ kg/m^3 was
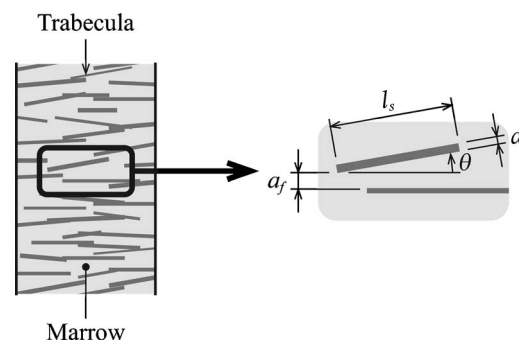
FIG. 3. Schematic drawing of the cancellous bone model used in the elastic FDTD method.

TABLE I. Trabecular sizes of bovine cancellous bone models with low and high porosities used in the elastic FDTD method. These values were uniformly distributed in the range of ±25% in each of a_s, a_f, and l_s, and ±10° in θ.

	Low	High
Trabecular thickness a_s	0.15 mm	0.15 mm
Pore space between trabeculae a_f	0.80 mm	1.35 mm
Trabecular length l_s	4.5 mm	4.5 mm
Trabecular angle θ	0°	0°
Porosity β	0.75	0.83

TABLE II. Physical parameters of bovine cancellous bone.

	Trabecula	Bone marrow
First Lamé coefficient λ	14.8 GPa	2.0 GPa
Second Lamé coefficient μ	8.3 GPa	0 GPa
Density ρ	1960 kg/m³	930 kg/m³

given by Lang.[11] The values λ_f=2.0 GPa, μ_f=0 GPa, and ρ_f=930 kg/m³ for bone marrow were measured by Hosokawa and Otani.[3]

C. Biot's cancellous bone model

The structure in cancellous bone is regarded as anisotropic and inhomogeneous. This structure was considered in the elastic FDTD method, whereas the cancellous bone
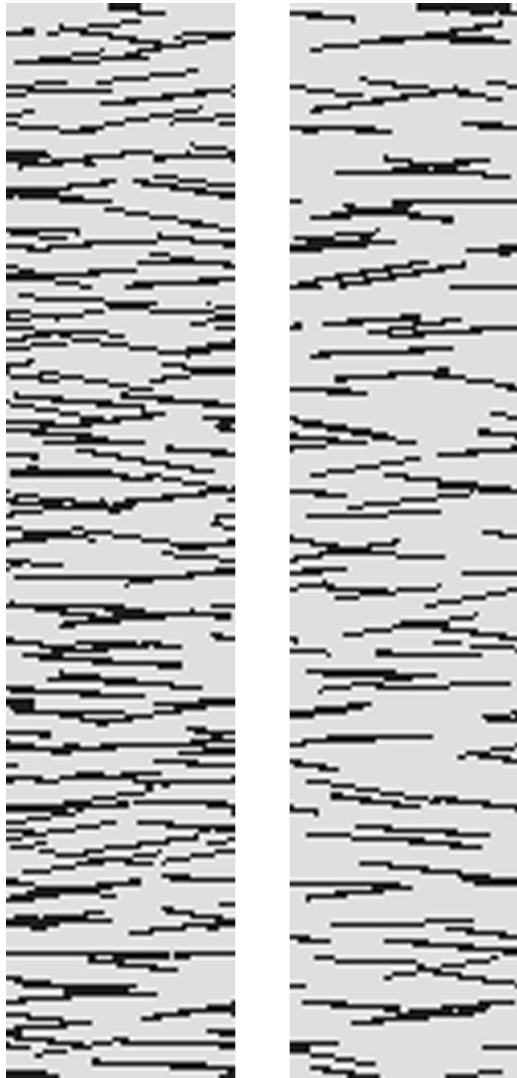
FIG. 4. Geometric drawings of the bovine cancellous bone models used in the elastic FDTD method (a) at low porosity (β=0.75) and (b) at high porosity (β=0.83).

model in Biot's FDTD method was assumed to be isotropic and homogeneous. The medium parameter values (ρ, ρ_f, m, F_r, η, k, H, C, M, and μ) were treated identically for all points in the model because determining such a large number of values at each point was exceedingly difficult. In a previous study,[3] the propagation speeds of both the fast and slow longitudinal waves in bovine cancellous bone could be estimated by the basic Biot's theory for a macroscopically isotropic and homogeneous medium. Numerical analysis using Biot's FDTD method was performed by similar method. The Young's modulus E_b of the trabecular frame in cancellous bone was calculated as

$$\frac{E_b}{E_s} = (1 - \beta)^n, \tag{6}$$

where E_s is the Young's modulus of the solid bone and n is a parameter that depends on the trabecular structure.[1] The structural factor α in Eq. (4) can be calculated by Berryman's relation[12]

$$\alpha = 1 - r(1 - \beta^{-1}), \tag{7}$$

where r is a variable related to the microscopic structure of the solid frame. This relation was originally applied to bone by Williams.[10] In Biot's theory,[6,7] the correction factor $F(\kappa)$ for the viscous resistance is defined as a function of frequency f

$$F(\kappa) = \frac{1}{4}\frac{\kappa T(\kappa)}{1 - 2T(\kappa)/j\kappa}, \tag{8a}$$

$$T(\kappa) = \frac{\text{ber}'(\kappa) + j\text{bei}'(\kappa)}{\text{ber}(\kappa) + j\text{bei}(\kappa)}, \tag{8b}$$

$$\kappa = a\sqrt{\frac{\omega\rho_f}{\eta}}, \tag{8c}$$

where the functions of $\text{ber}(\kappa)$ and $\text{bei}(\kappa)$ are the real and imaginary parts of the Kelvin function, a is the pore size parameter, which depends on both the size and shape of the pore, and $\omega(=2\pi f)$ is the angular frequency. The real part of $F(\kappa)$, with the center frequency of the transmitted wave substituted for f and the average pore size a_0 set as a, was used as F_r in Eqs. (3c) and (3d).

The parameters applied to Biot's model of bovine cancellous bone are listed in Table III. The Young's modulus E_s=22.0 GPa and Poisson's ratio ν_s=0.32 of solid bone (as trabeculae),[10] the density of the solid bone ρ_s=1960 kg/m³,[11] the bulk modulus K_f=2.2 GPa and the density ρ_f=930 kg/m³ of bone marrow,[3] the Poisson's ratio of the trabecular frame ν_b=0.32,[10] r=0.25 by Williams,[13] the viscosity of bone marrow η=1.5 P s by McKelvie and

TABLE III. Biot's parameters of bovine cancellous bone (as reported in Ref. 3).

Young's modulus of solid bone E_s	22.0 GPa
Poisson's ratio of solid bone ν_s	0.32
Bulk modulus of bone marrow K_f	2.0 GPa
Density of solid bone ρ_s	1960 kg/m^3
Density of bone marrow ρ_f	930 kg/m^3
Poisson's ratio of trabecular frame ν_b	0.32
Variable r	0.25
Viscosity of bone marrow η	1.5 P s
Parameter n	1.46

Palmer,[14] and $n=1.46$ for propagation in the direction of the trabecular alignment,[3] were given. The average pore size a_0 and permeability k for low and high porosities ($\beta=0.75$ and 0.83, respectively), as reported in Ref. 3, are listed in Table IV. All parameter values in Tables III and IV are identical to those determined in previous wave speed estimations.

IV. SIMULATED RESULTS

A. Previous experimental results (Ref. 3)

Previous experimental results for observations of the pulsed waveforms propagating through bovine cancellous bone specimens (9 mm in thickness) are shown in Fig. 5, together with the simulated results. The pulsed wave in Fig. 2(b) was applied to the specimens in the direction of the trabecular alignment. Figure 5(a-1) shows the waveform for a low porosity of $\beta=0.75$ (described as bone volume fraction $V_f=0.25$ in a previous paper), and Fig. 5(b-1) shows that for a high porosity of $\beta=0.83$ ($V_f=0.17$). In both figures, the fast and slow longitudinal waves corresponding to the first and second kind of waves could be clearly observed. As the porosity β increased, i.e., as the density or bone volume fraction decreased, the amplitude of the fast wave became smaller. On the other hand, the amplitude of the slow wave became much greater. Moreover, the fast wave speed decreased slightly with β, and the slow wave speed was almost constant. The fast and slow wave speeds at 1 MHz and the amplitude ratio of the slow wave to the fast wave are summarized in Table V, together with the simulated values.

In both the experiments and simulations, the wave speeds were measured using pulse spectrum analysis,[3,15,16] and the amplitude ratio was the peak–peak amplitude of the slow wave divided by that of the fast wave. Here, it is noted that both the experimental and simulated speeds could be regarded as nondispersive in the measured frequency range of 0.5–1.5 MHz.

TABLE IV. Pore sizes and permeabilities of bovine cancellous bone with low and high porosities (as reported in Ref. 3).

	Low	High
Pore size a_0	0.80 mm	1.35 mm
Permeability k	7×10^{-9}	3×10^{-8}
Porosity β	0.75	0.83

B. Simulated results obtained using the elastic FDTD method

The simulated results obtained using the elastic FDTD method are shown in Figs. 5(a-2) and (b-2), which are the waveforms for the low- and high-porosity cancellous bone models [Figs. 4(a) and 4(b)], respectively. Two types of waves, which seem to correspond to the fast and slow waves in the experiments, can be observed in these figures. Compared to the experimental results in Figs. 5(a-1) and (b-1), the analyzed fast waves are largely deformed since individual waves of distinct speeds overlap. The elastic FDTD and the experimental results of the wave speeds and amplitude ratios listed in Table V also differ. In the slow wave, however, the shapes are analogous and the speeds agree.

Several differences in the fast wave can be observed between the simulated and experimental results. The reason for these differences appears to be the inadequacy of the elastic cancellous bone models. In general, all trabeculae in cancellous bone are mutually connected to form a elastic skeletal frame. As shown in Fig. 4, the trabeculae in the 2D elastic model do not form a sufficient frame. It is difficult to analyze the fast wave propagating in cancellous bone using a model that forms an insufficiently elastic frame because the fast wave is related to the motion of the frame. The "fast-speed wave" analyzed using the elastic FDTD method is therefore regarded as the overlap of a number of waves propagating in individual trabecular rods, rather than the "fast wave" observed in the experiments.

A model with a elastic frame can easily be realized by incorporating cross struts that join the trabeculae. In the 2D model, however, a discontinuity consisting of a group of disconnected pore spaces surrounded by the trabeculae on every side inevitably exists in the pore fluid (bone marrow). For the propagation of the slow wave, which is related to the pore fluid motion, it is necessary for the pore spaces to be continuous in the propagation direction.[17] Therefore, the 2D elastic frame model is expected to make it difficult to analyze the propagation of the slow wave. (Some agreements in the slow wave, which were described above, appear to have resulted from satisfying this continuous condition in the models of Fig. 4.) Accordingly, it can be concluded that, in the cancellous bone model, both the connected trabecular frame and continuous pore spaces are required for analyzing the propagations of both the fast wave and the slow wave. This model, although not possible in two dimensions, can be easily realized in three dimensions.

If the computational grids N_x and N_y in the x- and y directions, respectively, are kept the same, and the grids N_z in the z direction are extended, then the memory and CPU requirements in the 3D acoustic FDTD method grow at least $4N_z/3$ times as large as that in two dimensions, because the number of unknowns increases from three ($\dot{u}_x$, $\dot{u}_y$, and p) to four ($\dot{u}_x$, $\dot{u}_y$, $\dot{u}_z$, and p). The 3D elastic FDTD method requires greater memory and CPU performance.

Luo et al.[18] carried out 2D simulations of ultrasound propagation through human calcaneal cancellous bone using the finite-difference (FD) method for an elastic wave equation. Pulsed waveforms propagating in the parallel and perpendicular directions to the trabecular alignment were satis-
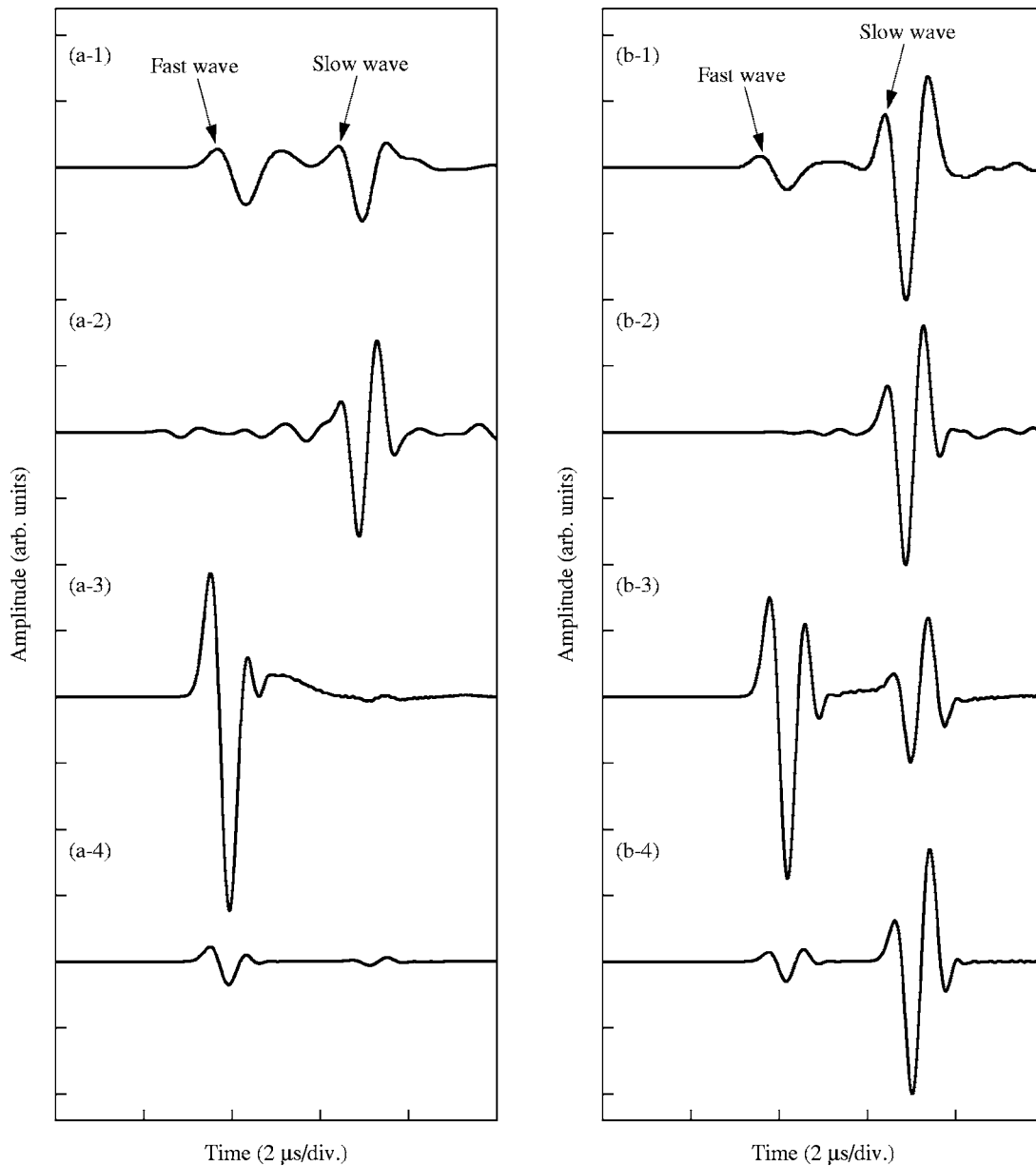
FIG. 5. Experimental [(a-1), (b-1)], elastic FDTD [(a-2), (b-2)], Biot's FDTD [(a-3), (b-3)], and modified Biot's FDTD [(a-4), (b-4)] results for observations of pulsed waveforms propagating through bovine cancellous bone in the direction of the trabecular alignment (a) at low porosity ($\beta=0.75$) and (b) at high porosity ($\beta=0.83$). The experimental waveforms are given in Ref. 3.

factorily analyzed. Moreover, it was shown that the wave speed in each direction was linearly related to low bone volume fraction (or high porosity). Hosokawa and Otani[4] experimentally showed that, due to weak dynamic coupling between the trabecular frame and the pore fluid, the wave speed ν in bovine femoral cancellous bone with rod–rod trabecular structure (Sec. I) followed Wyllie's equation[19,20]

$$\frac{1}{\nu} = \frac{1-\beta}{\nu_s} + \frac{\beta}{\nu_f},\tag{9}$$

where ν_s and ν_f are the speeds in the trabeculae and the pore fluid, respectively. This equation is dependent only on the porosity, not on the elasticity of the solid frame. The results of Luo can be considered to follow this simple relation, because the wave speed ν in Eq. (9) changes linearly in the high-porosity range. The 2D elastic model is probably appli-

cable to cancellous bone having a rod–rod trabecular structure, in which the wave propagation follows a simple relation, such as Wyllie's equation, and the trabecular structure of the image of calcaneal bone presented in Luo's paper appears to be a rod–rod-type structure.

C. Simulated results obtained using Biot's FDTD method

The simulated results obtained using Biot's FDTD method are shown in Figs. 5(a-3) and (b-3) for porosities β =0.75 and 0.83, respectively. Both the fast and slow waves could be clearly analyzed because both the solid skeletal frame and continuous pore fluid are considered in the original Biot's theory,[6,7] although the slow wave amplitude at β =0.75 is too small. As shown in Table V, these wave speeds

TABLE V. Experimental, elastic FDTD, Biot's FDTD, and modified Biot's FDTD results for the fast and slow wave speeds and amplitude ratios of the slow to fast waves at low and high porosities ($\beta=0.75$ and 0.83).

		Experimental results	Elastic FDTD results	Biot's FDTD results	Modified Biot's FDTD results
Fast wave speed	$\beta=0.75$	2670 m/s	4080 m/s	2730 m/s	2730 m/s
	$\beta=0.83$	2500 m/s	2010 m/s	2530 m/s	2530 m/s
Slow wave speed	$\beta=0.75$	1440 m/s	1440 m/s	1380 m/s	1380 m/s
	$\beta=0.83$	1460 m/s	1440 m/s	1440 m/s	1440 m/s
Amplitude ratio of	$\beta=0.75$	1.39	19.9	0.06	0.21
slow to fast waves	$\beta=0.83$	6.74	35.2	0.51	7.70

agree well with the experimental values, whereas the amplitude ratios are much smaller at both porosities. As shown in Figs. 5(a-3) and (b-3), the amplitudes of the analyzed fast wave are too large. In the basic Biot's equations [Eqs. (3a)–(3h)], only the loss due to the viscous resistance caused by the relative motion of the pore fluid is considered. The fast wave in cancellous bone, on the other hand, can be attenuated not only by the viscous resistance but also by other mechanisms, e.g., absorption within the trabecular frame and scattering by the trabeculae. In fact, the attenuation coefficients of the fast wave in bovine cancellous bone have been reported to be much larger than those of the slow wave,[3] although this is contrary to Biot's theory.[6,7] In order to consider the other possible attenuation for the fast wave, in addition to the shear wave, Eqs. (3e)–(3g) were modified as follows:

$$\frac{\partial \tau_{xx}}{\partial t} + \gamma_{xx}\tau_{xx} = H\frac{\partial \dot{u}_x}{\partial x} + (H - 2\mu)\frac{\partial \dot{u}_y}{\partial y} + C\left(\frac{\partial \dot{w}_x}{\partial x} + \frac{\partial \dot{w}_y}{\partial y}\right), \tag{3e'}$$

$$\frac{\partial \tau_{yy}}{\partial t} + \gamma_{yy}\tau_{yy} = (H - 2\mu)\frac{\partial \dot{u}_x}{\partial x} + H\frac{\partial \dot{u}_y}{\partial y} + C\left(\frac{\partial \dot{w}_x}{\partial x} + \frac{\partial \dot{w}_y}{\partial y}\right), \tag{3f'}$$

$$\frac{\partial \tau_{xy}}{\partial t} + \gamma_{xy}\tau_{xy} = \mu\left(\frac{\partial \dot{u}_x}{\partial y} + \frac{\partial \dot{u}_y}{\partial x}\right). \tag{3g'}$$

The fast wave attenuation in the l direction can be characterized by the last terms $\gamma_{ll}\tau_{ll}$ on the left-hand side in Eqs. (3e') and (3f'), and the shear wave attenuation by the term $\gamma_{xy}\tau_{xy}$ in Eq. (3g').

Previous experiments[3] showed that the attenuation coefficients of the fast wave in cancellous bone were independent of the porosity. Therefore, the fast wave attenuation can be considered to be slightly affected by the viscous loss caused by the pore fluid motion. Without the pore fluid motion, the modified Biot's equations (3a)–(3d), (3e'), (3f'), (3g'), and (3h) are reduced as follows:

$$\frac{\partial}{\partial t}(\rho \dot{u}_x) = \frac{\partial \tau_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y}, \tag{10a}$$

$$\frac{\partial}{\partial t}(\rho \dot{u}_y) = \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \tau_{yy}}{\partial y}, \tag{10b}$$

$$\frac{\partial \tau_{xx}}{\partial t} + \gamma_{xx}\tau_{xx} = H\frac{\partial \dot{u}_x}{\partial x} + (H - 2\mu)\frac{\partial \dot{u}_y}{\partial y}, \tag{10c}$$

$$\frac{\partial \tau_{yy}}{\partial t} + \gamma_{yy}\tau_{yy} = (H - 2\mu)\frac{\partial \dot{u}_x}{\partial x} + H\frac{\partial \dot{u}_y}{\partial y}, \tag{10d}$$

$$\frac{\partial \tau_{xy}}{\partial t} + \gamma_{xy}\tau_{xy} = \mu\left(\frac{\partial \dot{u}_x}{\partial y} + \frac{\partial \dot{u}_y}{\partial x}\right). \tag{10e}$$

The resistance coefficients γ_{ll} and γ_{xy} are then related to

$$\gamma_{xx} = \gamma_{yy} = \frac{2k_{\text{fast}}\alpha_{\text{fast}}\omega}{k_{\text{fast}}^2 - \alpha_{\text{fast}}^2}, \tag{11a}$$

$$\gamma_{xy} = \frac{2k_{\text{shear}}\alpha_{\text{shear}}\omega}{k_{\text{shear}}^2 - \alpha_{\text{shear}}^2}, \tag{11b}$$

where k_{fast} and k_{shear}, and α_{fast} and α_{shear} are the wave numbers and attenuation coefficients at a frequency of $\omega/2\pi$, for the fast and shear waves, respectively. Using Eqs. (11a) and (11b), the values of $\gamma_{xx} = \gamma_{yy} = 1.0 \times 10^6$ were calculated in reference to the attenuation coefficient of approximately $\alpha_{\text{fast}} = 0.2$ Np/mm at 1 MHz for the fast wave, which was measured previously,[3] and $\gamma_{xy} = 10 \times 10^6$ in the estimation of α_{shear} as $10 \times \alpha_{\text{fast}}$.

The results simulated using the modified Biot's FDTD equations are shown in Figs. 5(a-4) and (b-4), and the recalculated amplitude ratios are listed in Table V. At the porosity of $\beta=0.83$, the analyzed waveform is in good agreement with the experimental waveform. The amplitude ratio (i.e., the slow wave amplitude) largely increased with the porosity β, which also agrees with the experimental results. The large increment of the slow wave amplitude can be deduced to be caused by the viscous loss considered in Biot's FDTD method because this increment is not observed in the simulated results [Figs. 5(a-3) and (b-3)] obtained using the elastic FDTD method. In conclusion, Biot's FDTD method is more adequate for the analysis of fast and slow wave propagations in cancellous bone.

The values of the amplitude ratios are not in good agreement, particularly at $\beta=0.75$. The primary reason for this disagreement is the assumption that cancellous bone is isotropic and inhomogeneous. Moreover, the pulse widths of the analyzed waves are obviously narrower than those in the experiments because the frequency dependences of the attenuations were not exactly incorporated in the FDTD equations.

V. CONCLUSIONS

Two 2D FDTD models were developed to numerically analyze the Biot's fast and slow longitudinal waves propagating through bovine cancellous bone. Using the elastic FDTD model comprised of numerous trabecular rods and bone marrow in these spaces, neither the fast nor slow waves could be analyzed because of the insufficient trabecular frame and because the viscous loss caused by the pore fluid motion was not considered. A 3D model that considers viscous loss appear to be necessary for analysis by the elastic FDTD method. On the other hand, in the simulated results obtained using Biot's FDTD method based on the basic Biot's equations for an isotropic and inhomogeneous medium, both waves could be clearly observed. The wave speeds were found to be in good agreement with previously obtained experimental values. Moreover, the modified Biot's FDTD equations with the fast wave attenuation were applied to exactly analyze the change in the amplitude ratio of the fast to slow waves with the porosity. The results indicated that the analyzed waveforms are sufficiently analogous to the experimental waveforms.

The results obtained in the present study indicate that it is possible to numerically analyze the fast and slow waves propagating in cancellous bone using Biot's FDTD method. Extension of this method to waves propagating in an anisotropic and frequency-dependent medium is needed for more adequate application and is a subject for future investigation.

ACKNOWLEDGMENT

[1] L. J. Gibson, "The mechanical behavior of cancellous bone," J. Biomech. **18**, 317–328 (1985).

[2] J. L. Williams and J. L. Lewis, "Properties and an anisotropic model of cancellous bone from the proximal tibial epiphysis," J. Biomech. Eng. **104**, 50–56 (1982).

[3] A. Hosokawa and T. Otani, "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562 (1997).

[4] A. Hosokawa, T. Otani, T. Suzaki, Y. Kubo, and S. Takai, "Influence of trabecular structure on ultrasonic wave propagation in bovine cancellous bone," Jpn. J. Appl. Phys., Part 1 **36**, 3233–3237 (1997).

[5] A. Hosokawa and T. Otani, "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722 (1998).

[6] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).

[7] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).

[8] J. Virieux, "P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method," Geophysics **51**, 889–901 (1986).

[9] R. L. Higdon, "Absorbing boundary conditions for difference approximations to the multi-dimensional wave equation," Math. Comput. **47**, 437–459 (1986).

[10] J. L. Williams and W. J. H. Johnson, "Elastic constants of composites formed from PMMA bone cement and anisotropic bovine tibial cancellous bone," J. Biomech. **22**, 673–682 (1989).

[11] S. B. Lang, "Ultrasonic method for measuring elastic coefficients of bone and results on fresh and dried bovine bones," IEEE Trans. Biomed. Eng. **BME-17**, 101–105 (1970).

[12] J. G. Berryman, "Confirmation of Biot's theory," Appl. Phys. Lett. **37**, 382–384 (1980).

[13] J. L. Williams, "Ultrasonic wave propagation in cancellous and cortical bone: Prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112 (1992).

[14] M. L. McKelvie and S. B. Palmer, "The interaction of ultrasound with cancellous bone," Phys. Med. Biol. **36**, 1331–1340 (1991).

[15] P. H. F. Nicholson, G. Lowet, C. M. Langton, J. Dequeker, and G. Van der Perre, "A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435 (1996).

[16] W. Sachse and Y. H. Pao, "On the determination of phase and group velocities of dispersion waves in solid," J. Appl. Phys. **49**, 4320–4327 (1978).

[17] T. Bourbié, O. Coussy, and B. Zinszner, *Acoustics of Porous Media* (Technip, Paris, 1987), pp. 85–87.

[18] G. Luo, J. J. Kaufman, A. Chiabrera, B. Bianco, J. H. Kinney, D. Haupt, J. T. Ryaby, and R. S. Siffert, "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830 (1999).

[19] M. R. J. Wyllie, A. R. Gregory, and L. W. Gardner, "Elastic wave velocities in heterogeneous and porous media," Geophysics **21**, 41–70 (1956).

[20] M. R. J. Wyllie, A. R. Gregory, and G. H. F. Gardner, "An experimental investigation of factors affecting elastic wave velocities in porous media," Geophysics **23**, 459–493 (1958).

Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences

John L. Spiesberger

Department of Earth and Environmental Science, 240 South 33rd Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6316

A new nonlinear sequential Monte Carlo technique is used to estimate *posterior* probability distributions for the location of a calling animal, the locations of acoustic receivers, sound speeds, winds, and the differences in sonic travel time between pairs of receivers from measurements of those differences, while adopting realistic *prior* distributions of the variables. Other algorithms in the literature appear to be too inefficient to yield distributions for this large number of variables (up to 41) without recourse to a linear approximation. The new technique overcomes the computational inefficiency of other algorithms because it does not sequentially propagate the joint probability distribution of the variables between adjacent data. Instead, the lower and upper bounds of the distributions are propagated. The technique is applied to commonly encountered problems that were previously intractable such as estimating how accurately sound speed and poorly known initial locations of receivers can be estimated from the differences in sonic travel time from calling animals, while explicitly modeling distributions of *all* the variables in the problem. In both cases, the new technique yields one or two orders of magnitude improvements compared with initial uncertainties. The technique is suitable for accurately estimating receiver locations from animal calls. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1992708]

I. INTRODUCTION

Calling animals can be located by measuring the differences in sonic arrival times at pairs of receivers.[1–8] When the speed of sound is spatially homogeneous, the difference in arrival time multiplied by the speed gives the difference in distance of the animal from a pair of receivers. The locus of points in space for which this difference is constant is a hyperboloid. Given sufficient numbers of receivers, one intersects the hyperboloid from each pair to yield location. This so-called hyperbolic location technique has the property that the location of the animal is nonlinearly related to some of the relevant variables in the problem and there are typically many such variables such as the uncertain speed of propagation and the locations of the animal and the receivers. The nonlinearity and, particularly, the large number of variables has apparently made it computationally difficult for *any* approach to yield probability distributions for all the variables without making a linear approximation that is often invalid.[9] Since probability distributions are complete estimates of variables, it is desirable to have a method that can produce them in a computationally feasible manner. It would be desirable to be able to update distributions for the locations of the receivers and the speed of propagation as each new animal call is processed. The receivers could be stationary or mobile, and the speed of propagation could change with time. It would be desirable to use any realistic *prior* probability distributions for locations and speeds. A practical method for accomplishing all these goals is given here without using a linear approximation.

First consider the difficulties associated with nonlinearity. For each pair of receivers at coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$, one wants the possible locations of the animal, $\mathbf{s}$, satisfying

$$\frac{|\mathbf{s} - \mathbf{r}_i|}{c} - \frac{|\mathbf{s} - \mathbf{r}_j|}{c} = \tau_{ij}, \tag{1}$$

where the signal speed is c and the measured lag,

$$\tau_{ij} = t_i - t_j, \tag{2}$$

is the difference in the times for the call to reach receivers i and j, respectively. Each Cartesian coordinate for the source, (s_x, s_y, s_z), and receiver, $(r_i(x), r_i(y), r_i(z))$, is nonlinearly related to the measured lag, τ_{ij}, because

$$|\mathbf{s} - \mathbf{r}_i| = \sqrt{(s_x - r_i(x))^2 + (s_y - r_i(y))^2 + (s_z - r_i(z))^2}.$$

There are three problems stemming from this nonlinearity. First, the probability distribution for the animal's location has no known analytical solution given probability distributions for the lag, receiver locations, and the speed of sound unless some of these distributions are set to an unwavering value. Second, even when one linearizes the relationship between the coordinate of the animal and the other variables, the distribution for its location can be in error by one or two orders of magnitude.[9] Third, a linear estimation scheme can converge to an incorrect solution corresponding to a nonglobal minimum.[10] Thus linear estimation techniques such as least squares, Wiener, and Kalman filters[11] can provide unreliable and inaccurate results. We are thus faced with the problem of how to obtain reliable and robust estimates of probability distributions.

Consider next the difficulties with the large number of pertinent variables. It is assumed that estimated locations of the receivers and the speed of propagation have errors, so their values are to be improved in the light of data. For $\mathcal{R}$ receivers, there are $3\mathcal{R}-6$ associated Cartesian coordinates to estimate. The reduction by 6 merely means we are uninterested in the absolute location and rotation of the coordinate system. Additionally, there are three unknown variables for the animal's location, and one unknown variable for the speed of sound. Summing these there are

$$V = 3\mathcal{R} - 2 \qquad (3)$$

variables of interest in hyperbolic location problems. For example, there are 16 variables with six receivers. When the effective speed of the acoustic signal between the animal and each receiver is modified due to winds, refraction of the acoustic path,[12] or propagation through two different media such as the air and water,[13] the hyperbolic location problem can be generalized to obtain locations, but one needs to introduce a different effective speed, c_i, along the acoustic path between the animal and receiver i. This increases the number of variables to be estimated by $\mathcal{R}-1$. The number of relevant variables increases further if the probability distributions of the lags are estimated from the data.

Bayes theorem[14] provides optimal probability distributions but is impractical. Suppose one wants the joint probability distribution, $f(\mathbf{s},\mathbf{r},\mathbf{c}|\tau_{ij})$, of source location, receiver locations, $\mathbf{r}$, and speeds of propagation between the source and each receiver, $\mathbf{c}$, given a lag, τ_{ij}. Bayes theorem supplies the desired result,

$$f(\mathbf{s},\mathbf{r},\mathbf{c}|\tau_{ij}) = \frac{\mathbf{f}(\tau_{ij}|\mathbf{s},\mathbf{r},\mathbf{c})\pi(\mathbf{s},\mathbf{r},\mathbf{c})}{\int \mathbf{f}(\tau_{ij}|\mathbf{s},\mathbf{r},\mathbf{c})\pi(\mathbf{s},\mathbf{r},\mathbf{c})\mathbf{ds}\,\mathbf{dr}\,\mathbf{dc}}, \qquad (4)$$

in terms of the conditional joint distribution of the data on the source, receivers, and speeds of propagation, $f(\tau_{ij}|\mathbf{s},\mathbf{r},\mathbf{c})$, and the prior joint distribution of the source location, receiver locations, and speeds, $\pi(\mathbf{s},\mathbf{r},\mathbf{c})$. If the distributions on the right side could be evaluated analytically, then evaluating Eq. (4) would be the end of the problem. One could introduce new data and keep finding better estimates of the distribution on the left given updated distributions on the right. Brute force evaluation of the distributions on the right appears to be computationally difficult because there are many variables. For hyperbolic location and six receivers, one needs to estimate the joint distributions of 16 variables for each introduced datum. Suppose each variable is divided into ten bins. Accurate estimation of the joint distribution requires a reliable probability of occurrence in 10^{16} bins. Instead of binning, one can estimate distributions using Gibbs sampling or Markov Chain Monte Carlo approaches,[14,15] but they appear to be computationally impractical for a large number of variables such as 16. Despite the fact that this problem is commonly encountered in the fields of acoustic and electromagnetic tracking, the author is unaware of any publication where the distributions of all the variables (about 16 or more) are estimated without making a linearizing approximation. Instead, the literature appears to treat other problems with elegant approaches where some of the variables are known without error (such as some or all of the receiver coordinates) and/or some linear approximation is adopted.[12,15–20]

The approach taken here has its root in a method for estimating the distribution of an animal's location given realistic *a priori* estimates for the distributions of sound speed, receiver locations, and measurement error.[9] We explain how that approach can be generalized to estimate the distributions of all the variables and how to include dynamical models for the evolution of all variables between the receptions of different animal calls. The approach is more convenient to discuss when given the name "sequential bound estimation" for reasons that are apparent after explaining how it works.

II. SEQUENTIAL BOUND ESTIMATION

Sequential bound estimation could be applied to many problems that are amenable to solution using Bayes theorem. The technique is demonstrated in the context of locating calling animals in situations where one has more receivers than needed to obtain a mathematical solution. The sequential nature is evident because data are sequentially processed from different subsets of the acoustic receivers. The idea of treating these data in subsets and transitioning estimated variables between one datum and the next is analogous to other problems where data are sequentially processed. Section IV explains how sequential bound estimation could be applied in a different location problem.

A. Receiver constellation

We discuss how to treat situations where one has just the right number of receivers to estimate location without mathematical ambiguity. In general, location problems in two spatial dimensions require three or four receivers[21] and three-dimensional problems require four or five receivers.[2] The location of the animal with respect to the receivers dictates the required number.[2,21] A "receiver constellation" is the minimum number of receivers required to obtain an unambiguous location. A systematic way of determining the necessary number of receivers is to use an analytical solution of location based on the lesser number, and use the datum from an additional receiver if needed.

B. Analytical solution for location from each constellation

When the speed of the signal varies between the animal and each receiver, the general solution for the animal's location in three spatial dimensions from a group of four receivers is[9]

$$\vec{s} = \mathbf{R}^{-1}\frac{\vec{b}}{2} - \mathbf{R}^{-1}\vec{f}t_1 - \mathbf{R}^{-1}\vec{g}t_1^2, \qquad (5)$$

where

$$\mathbf{R} \equiv \begin{pmatrix} r_2(x) & r_2(y) & r_2(z) \\ r_3(x) & r_3(y) & r_3(z) \\ r_4(x) & r_4(y) & r_4(z) \end{pmatrix}, \quad \vec{b} \equiv \begin{pmatrix} \|\vec{r}_2\|^2 - c_2^2\tau_{21}^2 \\ \|\vec{r}_3\|^2 - c_3^2\tau_{31}^2 \\ \|\vec{r}_4\|^2 - c_4^2\tau_{41}^2 \end{pmatrix},$$

$$\vec{f} \equiv \begin{pmatrix} c_2^2\tau_{21} \\ c_3^2\tau_{31} \\ c_4^2\tau_{41} \end{pmatrix}, \qquad (6)$$

and

$$\vec{g} \equiv \frac{1}{2}\begin{pmatrix} c_2^2 - c_1^2 \\ c_3^2 - c_1^2 \\ c_4^2 - c_1^2 \end{pmatrix}, \tag{7}$$

and t_1 is the solution from the quartic equation,

$$a_6 t_1^4 + 2a_5 t_1^3 + (a_3 - a_4 - c_1^2)t_1^2 - a_2 t_1 + \frac{a_1}{4} = 0, \tag{8}$$

where

$$a_1 \equiv (\mathbf{R}^{-1}\vec{b})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{b}), \quad a_2 \equiv (\mathbf{R}^{-1}\vec{b})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{f}),$$

$$a_3 \equiv (\mathbf{R}^{-1}\vec{f})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{f}), \quad a_4 \equiv (\mathbf{R}^{-1}\vec{b})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{g}), \tag{9}$$

$$a_5 \equiv (\mathbf{R}^{-1}\vec{f})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{g}), \quad a_6 \equiv (\mathbf{R}^{-1}\vec{g})^{\mathrm{T}}(\mathbf{R}^{-1}\vec{g}),$$

and $\mathbf{R}^{-1}$ is the inverse of $\mathbf{R}$. Quartic equations have analytical solutions. When the speed is spatially homogeneous, i.e., $c_i = c_1 \forall i$, the cubic and quartic terms in Eq. (8) vanish and the resulting quadratic equation is that found before for hyperbolic location.[3] The analytical solution for location in two spatial dimensions has the same form as above except one assumes the constellation has three receivers so one removes the last rows in the expressions for $\mathbf{R}$, $\mathbf{b}$, $\mathbf{f}$, and $\mathbf{g}$.

The signal speed, c_i, is more general than the speed appearing in the scalar wave equation for pressure perturbations. Here, c_i denotes the effective speed of sound between the source and receiver, which is the distance divided by the travel time. Effects of winds, currents, and nonstraight propagation paths due to refraction and diffraction all affect the time for sound to reach a receiver.

Ambiguous solutions for location occur when Eq. (8) yields more than one nonnegative real solution. Such ambiguities occur in practice. For each ambiguity, one generates a model for τ_{51}, which one can do because one knows where receiver 5 is, and chooses the root of t_1 that is closest to that measured.

C. Probability distributions for all variables

We show how one generalizes the idea for obtaining the probability distribution for animal location[9] to obtain probability distributions for all variables including the lags. The lags, τ_{ij}, are treated as random variables because they contain errors, just like all the other variables in the problem such as the locations of the receivers. We also consider situations where there are more receivers than needed to obtain a mathematically unambiguous solution for location.

Since $\mathbf{r}_i$, τ_{ij}, and c_i are random variables, then t_1 and $\mathbf{s}$ are random variables because of Eqs. (5) and (8). One assigns realistic *prior* probability distributions to each of these variables except t_1, whose distribution depends deterministically on the other distributions. For example, in a case of most ignorance, one can assign *prior* uniform distributions such as intervals $[-3000, 3000]$, $[-3000, 3000]$, and $[0, -20]$ m for the (x, y, z) Cartesian coordinates of a snapping shrimp where it is known that one cannot hear its sound at a

distance of 3000 m and it must be above a bottom depth of 20 m.

For three-dimensional problems, the number of ways of choosing a four-receiver constellation among $\mathcal{R}$ receivers is

$$N = \frac{\mathcal{R}!}{(\mathcal{R} - 4)!4!}. \tag{10}$$

A computer is used to draw a single *configuration* of random variables for the first constellation. Each configuration consists of the set $\{r_i(x), r_i(y), r_i(z), \tau_{ij}, c_i\}$ where i and j are taken from the set of four of the $\mathcal{R}$ receivers for the chosen constellation, and where $i > j$ to avoid using redundant data (e.g., $\tau_{23} = -\tau_{32}$). For each configuration, an animal location is computed from Eq. (5) if that equation yields a unique location. If a configuration yields more than one location, a location is chosen to be that yielding the closest difference in travel time to a randomly chosen fifth receiver and one of the receivers in the constellation.

A *valid configuration* must pass two criteria. First, the animal's location must lie within its *a priori* spatial limits. Second, there must be a real-valued solution for location from Eq. (5) since configurations can yield only complex-valued roots. Configurations not passing these criteria are discarded because they cannot occur in reality.

Valid configurations from a receiver constellation define a cloud of animal locations, receiver locations, estimates for the effective speeds of sound for each path, and the values of each of the lags. Accurate probability distributions for each require a sufficient number of valid configurations.

Some constellations yield better estimates of a variable than others. For example, an animal surrounded by four receivers would be more accurately located than one where four receivers were clumped together along a line at great distance from an animal.

Since each receiver constellation individually yields limits for the upper and lower values of each variable, one can enforce these limits to constrain random selections for that variable when seeking the valid configurations from the next constellation. That is why this method is called sequential bound estimation.

For example, suppose the *prior* distribution for variable k has some shape within the interval $[\check{v}(k)_0, \hat{v}(k)_0]$ with $\check{v}(k)_0 < \hat{v}(k)_0$ and the subscript denotes the bound after using constellation p with 0 denoting *a priori* values. After using data from constellation $p=1$, its *a posteriori* distribution must be contained in the interval, $[\max\{\check{v}(k)_0, \check{v}(k)_1\}, \min\{\hat{v}(k)_0, \hat{v}(k)_1\}]$. In general, the interval following use of constellation p is

$$[\max\{\check{v}(k)_{p-1}, \check{v}(k)_p\}, \min\{\hat{v}(k)_{p-1}, \hat{v}(k)_p\}]. \tag{11}$$

Sequential bound estimation *assigns* a distribution for the variable within these bounds. If one chooses a uniform distribution, then one is not overconstraining its distribution and valid configurations for the next constellation are obtained without prejudice. (In the parlance of information theory, the maximum entropy principle is used to prove that the uniform distribution contains the least information given only a variable's bounds.[22]) If one could implement Bayes theorem, the

variable would be contained within the bounds given by Eq. (11), but the distribution would likely not be uniform. It is the author's opinion that it is better to underconstrain the distribution of a variable than to overconstrain it in an incorrect manner, as might occur if one was to assign the variable a narrow Gaussian distribution. Sequential bound estimation could then assign a uniform distribution to each variable after assimilating data from all but the last constellation. After the last constellation, one does not reassign a distribution to the variable, but rather simply computes its probability distribution from its Monte Carlo samples. In fact, one can estimate the joint distribution of the variables from the output of the last constellation. Any desired statistic can be estimated from the final joint distribution such as a percent confidence limit, the mean, and a maximum likelihood value.

A consistent estimator, such as Bayes theorem, tends toward the true distribution as the number of samples goes to infinity. Sequential bound estimation is not a consistent estimator but it can provide distributions that bound the correct distributions. This may not be such a great drawback for two reasons. First, many, if not most, problems of interest are ones where the *prior* distributions of the variables are unknown and need to be guessed. So even if one could implement Bayes theorem, the resulting joint distribution would be in error. When *prior* distributions are in error, methods that work hard to maintain the mathematical rigidity of their assumptions may yield results that are less accurate than sequential bound estimation where the bounds of the variables are enforced with minimum (i.e., uniform) constraints on their distributions at intermediate steps. Second, it is better to have an algorithm that provides distributions of pertinent variables than to have no means to compute them at all.

D. Transitioning variables from one animal call to the next

Like a Kalman filter,[11] sequential bound estimation can incorporate a model to transition variables from one animal call to the next. A few examples are given.

Suppose the speed of sound varies with time. Bounds for the speed between times t_1 and t_2 relative to a reference speed, $c_r(t_1)$, can be modeled as

$$\delta\hat{c}_r(t_2) = \min\left\{+|\delta c_{max}|,\ \hat{c}_r(t_1) - c_r(t_1) + |t_2 - t_1|\left|\frac{d\delta c_r}{dt}\right|\right\},$$

$$\delta\check{c}_r(t_2) = \max\left\{-|\delta c_{max}|,\ \check{c}_r(t_1) - c_r(t_1) - |t_2 - t_1|\left|\frac{d\delta c_r}{dt}\right|\right\},$$

(12)

where the path-averaged deviation of speed between the source and receiver r is $\delta c_r(t)$, the maximum deviation is $|\delta c_{max}|$, and the maximum allowed rate of change of the fluctuation is $|d\delta c_r/dt|$. Equation (12) expresses a method to relax prior constraints on bounds to maximum limits at a specified maximum rate of change. In the absence of further data, the bounds expand to *a priori* values. Bounds for the winds could obey similar relationships.

Similarly, we can transition probability distributions for a mobile animal. Let the prior probability distribution for the animal's location be uniformly distributed in the (x, y, z) Cartesian
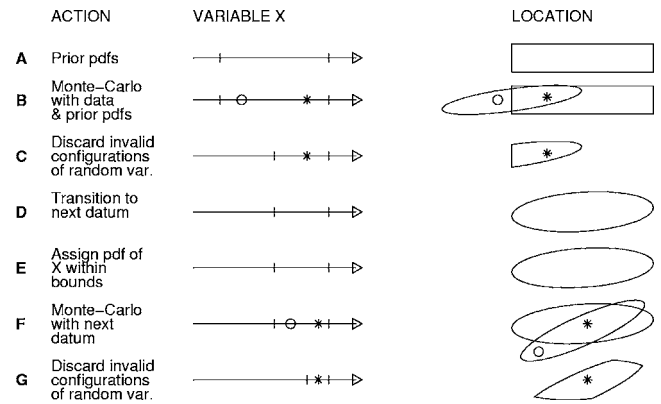
FIG. 1. Flow diagram for sequential bound estimation where location of some object is a function of a variable, x. Probability distribution function is pdf.

tesian intervals $[\check{s}_x(t_i^-), \hat{s}_x(t_i^-)]$, $[\check{s}_y(t_i^-), \hat{s}_y(t_i^-)]$, and $[\check{s}_z(t_i^-), \hat{s}_z(t_i^-)]$ for a call received at time t_i. The $-$ superscript denotes the time just prior to assimilation of the data from the call at time t_i. After assimilating the data from time t_i, we have new lower and upper bounds for the animal at time t_i^+ given by $[\check{s}_x(t_i^+), \hat{s}_x(t_i^+)]$, $[\check{s}_y(t_i^+), \hat{s}_y(t_i^+)]$, and $[\check{s}_z(t_i^+), \hat{s}_z(t_i^+)]$ where the $+$ denotes time just after the call. Assign lower and upper bounds, denoted respectively by superscripts $\check{}$ and $\hat{}$, for each component of the Cartesian velocity, (U, V, W), of the animal between calls. Then we know that

$$\check{s}_x(t_{i+1}^-) = \check{s}_x(t_i^-) + (t_{i+1} - t_i)\check{U},$$

$$\check{s}_y(t_{i+1}^-) = \check{s}_y(t_i^-) + (t_{i+1} - t_i)\check{V},$$

(13)

$$\check{s}_z(t_{i+1}^-) = \check{s}_z(t_i^-) + (t_{i+1} - t_i)\check{W},$$

with analogous equations for the transition to the upper bounds. There are numerous ways to transition such bounds.

E. Flow diagram

Figure 1 summarizes sequential bound estimation for location problems. Step A assigns *prior* distributions and bounds to each variable, e.g., x. Its *prior* bounds are indicated by tic marks on its axis. For hyperbolic locations of calling animals, x could be any Cartesian coordinate of a receiver, an effective speed of propagation, etc. Also, one assigns *prior* bounds (rectangle) for location of the desired object.

In step B, one computes configurations of variables using the first set of data, where the variables are drawn from their *prior* distributions. Each random configuration of variables associates a set of variables with a location. For example, values of x indicated by the circle and asterisks lead to locations given by the circle and asterisks, respectively. All configurations define a cloud of locations in the ellipse.

Step C discards invalid configurations of random variables that are identified by locations outside *prior* bounds (rectangle). Valid configurations have bounds indicated by the new tic marks on the x axis and by the intersection of the

VARIABLE X

A Posterior Bounds

B Bounds for new datum

C Intersect Bounds and Assign pdf
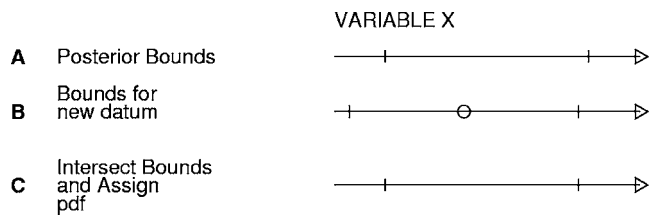
FIG. 2. How sequential bound estimation derives new bounds and probability distribution (pdf) when a variable X is a datum.

TABLE I. Cartesian (x,y,z) coordinates for the five receivers shown in Fig. 3 of Ref. 24. Nonzero locations are measured within ±0.05 m.

Receiver	Cartesian coordinate (m)		
	x	y	z
R1	0	0	0
R2	19.76	0	0
R3	17.96	−18.80	0
R4	2.34	−29.92	−0.02
R5	−12.41	−14.35	−0.43

ellipse and the rectangle. The asterisk is contained in a set with a valid configuration.

Bounds for the variables are transitioned to the next datum in step D. The bounds for x do not change, but they could in general. For example, if x was a coordinate of a stationary (mobile) receiver, its bounds would not (would) change in the transition.

Step E assigns new distributions to each variable. A uniform distribution is suggested to be the most appropriate to use within the transitioned bounds.

Step F uses the next datum to generate configurations of variables by drawing from the current distributions of variables. Two configurations are indicated by the circle and asterisks. The cloud of locations for the new random configuration is a new ellipse.

Step G discards invalid configurations as in step C, leaving only valid configurations. Not only does one have better bounds for location, but one also has better bounds for each variable. One can estimate distributions of variables from valid configurations.

When variable X is a datum, bounds and probability distributions are determined in a slightly different way. Suppose the posterior bounds for X from a previous step are transitioned to the time just before the next datum, X, is assimilated [Fig. 2(a)]. The next datum value for X [circle, Fig. 2(b)] has some *a priori* probability distribution bounded by the tic marks [Fig. 2(b)]. Sequential bound estimation assigns a probability distribution that is bounded by the intersection of the bounds in panels (a) and (b), i.e., panel (c). Then the distribution in (c) is drawn from to compute configurations of random variables. If all assumptions and data are handled properly, the bounds in (a) and (b) will overlap. When they do not overlap, the algorithm can notify the user that something is wrong. For example, one could have an instrumental error or an incorrect assumption concerning a probability distribution.

III. EXAMPLES

We provide simple examples to show how sequential bound estimation may be used to solve problems of interest that come up when locating calling animals from widely separated receivers. Each example treats more variables than dealt with in the literature. Convergence of distributions is obtained in each example by increasing the number of valid configurations until reaching 4000, at which point no significant change is found.

A. Verifying sequential bound estimation with data

We show that probability distributions for location are consistent with data. Previously, an undeveloped form of sequential bound estimation was used to locate a whale.[23] The only variable whose bounds were sequentially modified was the whale's location. Here, all variables are sequentially modified except the distributions of the lags.

On 4 June 1995, five omni-directional microphones recorded sounds from a Red-winged Blackbird in Port Matilda, PA (bird B1 in Fig. 3 of Ref. 24). Isodiachronic location is done using published differences in travel time (Table 2, Ref. 24).

A priori distributions of the variables are uniform. Lags have means as measured with intervals of ±0.000 067 s on either side due to effects of noise and interference between multipath.[24] The speed of sound has a mean of 344 m/s and an interval of ±2 m/s on either side. Cartesian components of the wind have zero means with intervals of ±2, ±2, and ±0.5 m/s on each side for the two horizontal and vertical components, respectively. Cartesian coordinates of the receivers have means given by measured values. The interval on each side of the mean is ±0.05 m except as follows. Receiver 1 is defined to be at the origin of the coordinate system. Receiver 2's location is defined such that the y axis passes through it and its z coordinate is zero. Receiver 3 is defined such that its z coordinate is zero. These definitions define the absolute location and rotation of the coordinate system (Table I). There are five receiver constellations [Eq. (10)].

It is assumed that the location of the Red-winged Blackbird is initially described as a uniform random variable in the Cartesian $x-y-z$ intervals $(-30, 30)$ m, $(-30, 30)$ m, and $(-5, 10)$ m, respectively, where receiver one is about a meter above the ground.

Following sequential bound estimation, Cartesian coordinates of the Red-winged Blackbird have 95% and 100% confidence limits of $(9.7 \pm 0.1, 7.7 \pm 1.1, 0.4 \pm 5.2)$ m and $(9.7 \pm 0.2, 7.9 \pm 1.6, 0.9 \pm 5.9)$ m, respectively. These are consistent with the optical/visual survey for its location which has a 95% confidence limit of $(9.8 \pm 0.5, 6.8 \pm 0.5, 2.3 \pm 1)$ m (Table IV, Ref. 24).

Twenty probability distributions are estimated in this example (nine Cartesian receiver coordinates, three Cartesian source coordinates, three Cartesian wind components, and five variables for the speed of sound between the animal and each receiver).

John L. Spiesberger: Locating calling animals and estimating everything
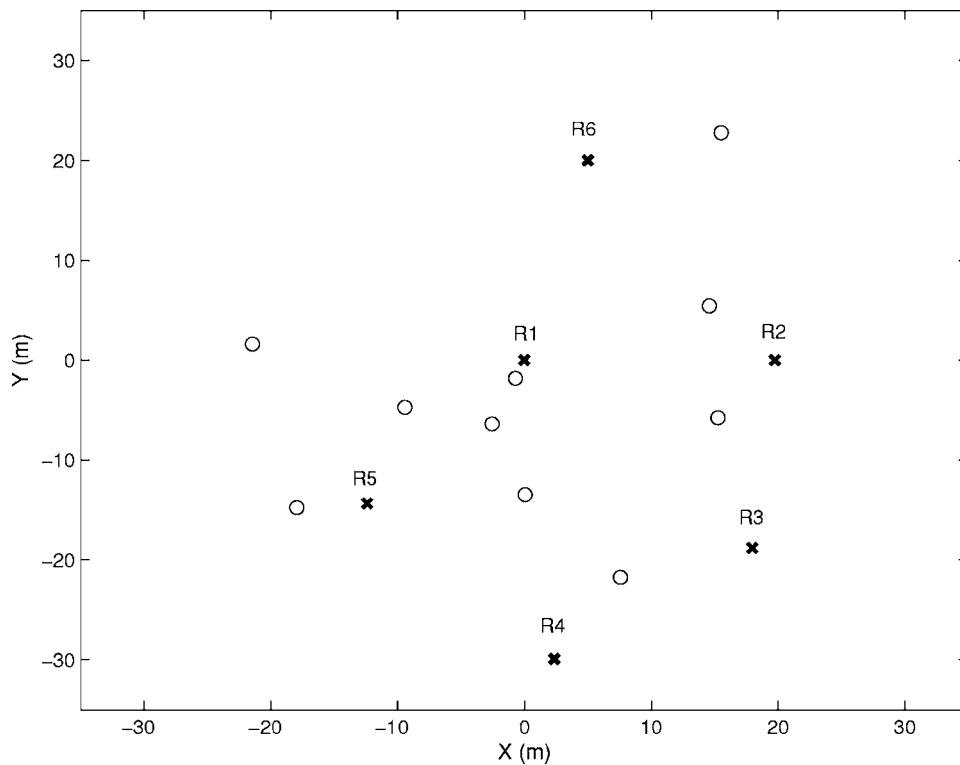
FIG. 3. Horizontal locations of six receivers and ten sources (circles) used to estimate the spatially homogeneous component of sound speed. The receivers have elevations of 2, 2, 2, 1.98, 1.57, and 7.0 m, respectively.

B. Estimating speed of sound: Simulation

We find a probability distribution for the spatially homogeneous component of the sound speed field when the locations of receivers are accurately but imperfectly measured. Six receivers are distributed in air over a region of about 30 m by 30 m (Fig. 3). Their elevations are about 2 m, except for receiver 6 which is at an elevation of 7 m. The initially unknown locations for ten sources are chosen at random (Fig. 3). The sources are assumed to produce sound at 10-s intervals at elevations between between 3 and 6 m.

All random variables have a uniform distribution in the intervals quoted below. Lags are distributed within $\pm 0.000\,05$ s about their noiseless values.

The initial guesses for the Cartesian coordinates of the receivers are drawn from a uniform distribution within ± 0.05 m of their true values. Their initial distributions are taken to have means about these values within an interval of ± 0.1 m. An error of zero is assigned to the location of receiver 1 (defined to be the origin), to the y and z coordinates of receiver 2, and the z coordinate of receiver 3 because these define the origin and orientation of the coordinate system.

Mild winds are simulated to be spatially homogeneous with a temporal scale of 5 s. The model assumes their initial x, y, and z components are specified to fluctuate within ± 1, ± 1, and 0.5 m s^{-1} about a mean of 0. The maximum rate of change of any component is 1 m s^{-2}.

The actual speed of sound is composed of spatially homogeneous and inhomogeneous components. The homogeneous component is about 330 m s^{-1}. It is allowed to vary at a maximum rate of 0.01 m s^{-2}, so in 10 s it can vary by 0.1 m s^{-1}. The limits for the homogeneous component are constrained to be within 330 ± 0.5 m s^{-1}. The inhomogeneous component may vary from section to section within ± 0.1 m s^{-1} at a maximum rate of 0.01 m s^{-2}.

In order to test the ability of sequential bound estimation to estimate the speed of sound, the algorithm is given a much wider range of speeds than actually occur. The spatially homogeneous component of sound speed is assumed to have a *prior* distribution that is uniformly distributed within 330 ± 50 m s^{-1}. After sequential bound estimation assimilates simulated data from source 1, the lower and upper bounds for the homogeneous component of sound speed are between $330 - 6.3$ and $330 + 3.8$ m s^{-1}, respectively. Following the use of source 7, these bounds are $330 - 2.1$ and $330 + 1.3$ m s^{-1}, which is equivalent to a variation of ± 2.8 °C if due to temperature. The bounds do not change significantly following the use of sources 8–10. The correct value of the speed of sound, about 330 m s^{-1}, falls within the bounds provided by sequential bound estimation. Sequential bound estimation reduces the initial uncertainty of ± 50 by a factor of about 29. The 95% confidence limits for the speed of sound along each section are about ± 1.3 m s^{-1}, which is equivalent to ± 2.1 °C.

Consider adding four additional receivers at Cartesian coordinates (20, 0, 3), (−20, 22, 4), (25, 21, 2), and (−5, 3, 3) m. Ten new randomly chosen source locations are different than above, but are similarly situated and not chosen to optimize any measure of performance. When the same statistics are assumed as above, it is found that the homogeneous component of the sound speed field has 95% and 100% confidence limits of ± 0.6 and ± 0.7 m/s, respectively. Sequential bound estimation reduces the initial uncertainty of ± 50 m s^{-1} by a factor of about 71. If due to temperature, the limits correspond to ± 1.0 and ± 1.1 °C,

respectively. This is an accurate measurement of average air temperature, and it is about twice as good as the case with six receivers above.

The number of distributions estimated for the six- and ten-receiver simulations are 25 and 41, respectively for each animal. The number of Cartesian receiver coordinates are 12 and 24, respectively, for the six- and ten-receiver simulations. Both simulations use three Cartesian source coordinates, three Cartesian wind components, one homogeneous component of speed, and six and ten variables for each inhomogeneous component of speed corresponding to six and ten receivers, respectively.

C. Estimating wind field: Simulation

Suppose an animal is located at Cartesian coordinate (20,100,7) m and its signals are monitored at five microphones at (0,0,0), (25,0,3), (50,3,5), (30,40,9), and (1,30,4) m, respectively. Each coordinate is initially assumed to be distributed uniformly within ±0.04 m about the true values. This accuracy is typical of that obtained from optical surveys. *A priori* errors are zero for receiver 1, the y and z coordinates of receiver 2, and the z coordinate for receiver 3. These coordinates define the origin and orientation of the coordinate system.

For definiteness, assume the animal's call has a rms bandwidth of 1000 Hz and, following the cross correlation of the signal between each pair of microphones, the peak signal-to-noise ratio is 20 dB. The standard deviation of the peak lag in the cross-correlation function has a standard deviation of 16 μs (Ref. 25), an accuracy that is achieved in practice. Initial errors of the lags are taken to be uniformly distributed within ±32 μs of their noiseless values.

Simulated lags are computed for a speed of sound of 330 m/s and for a horizontal wind blowing at 10 m/s toward the positive y Cartesian axis. The speed of sound has zero variation about a mean of 330 m/s, and the y component of the wind is initially assumed to be uniformly distributed in the interval 0 to +20 m/s. Winds along the x and z axes are set to zero.

After the first constellation, the distributions of the variables (except the lags) are assumed to be uniformly distributed about the most recent values of their sample means from the valid configurations. The limits of the uniform distribution are determined using sequential bound estimation.

With isodiachronic location, 95% confidence limits for the animal are x: 16.6–20.4 m, y: 98.8–101.8 m, and z: 3.3–40.3 m. These are statistically consistent with the correct location at $(20, 100, 7)$ m. The large variation in z stems from the fact that the animal and receivers are nearly coplanar.

Valid configurations of the y component of the wind have 95% confidence limits of 8.9 and 19 m/s. This is consistent with the true speed of 10 m/s.

The number of distributions estimated in this simulation is 13 (nine Cartesian receiver coordinates, three Cartesian animal coordinates, and one Cartesian wind component).

D. Surveying locations of receivers

1. Simulation

Suppose the calls of 100 animals are used to survey the locations of poorly positioned receivers. For simplicity, suppose the calls occur at 10-s intervals at elevations between 0 and 3 m. These calls could be due to a person walking around and blowing a whistle or due to animals who naturally call within 3 m of the ground. The true locations of the receivers are shown in Fig. 3. Horizontal locations of the calls are chosen to be uniformly distributed within this plotted domain (not shown). *A priori* distributions of all variables are uniformly distributed in the intervals given below.

The x, y, and z components of the wind are distributed within $[-1, +1]$, $[-1, +1]$, and $[-0.5, +0.5]$ m s^{-1}, respectively. Without fluctuations, the mean speed of sound is 330 m s^{-1}. The spatially homogeneous and inhomogeneous fluctuations of sound speed are distributed in $[-0.5, +0.5]$ and $[-0.2, +0.2]$ m s^{-1}, respectively. The initial estimates of the receiver locations are incorrect. Their initial locations are chosen by drawing from a uniform distribution within ±2 m of each of their true Cartesian coordinates. Their error distributions are assumed to be within ±3.94 m of each of their incorrect Cartesian coordinates except as follows. As before, no error is assigned to the location of receiver 1 (defined to be the origin), to the y and z coordinates of receiver 2, and the z coordinate of receiver 3 because these define the origin and orientation of the coordinate system.

The spatially homogeneous and inhomogeneous fluctuations of the sound speed are modeled as in Eq. (12). For homogeneous fluctuations, we take δc_{max} and $d\delta c_r/dt$ equal to 0.5 m s^{-1} and 0.01 m s^{-2}, respectively. The time series for this component has a temporal scale of 500 s. For inhomogeneous fluctuations, we take δc_{max} and $d\delta c_r/dt$ equal to 0.2 m s^{-1} and 0.01 m s^{-2}, respectively. The time series of this component has a temporal scale of 10 s.

The Cartesian components of the wind have temporal variations modeled in the same way as the speed of sound. Maximum values of the x, y, and z components are within the intervals stated above. The maximum rate of change of each component is 1 m s^{-2}. The time series of winds have a temporal scale of 5 s.

Unlike previous examples, the probability distributions of the lags are also updated using sequential bound estimation. The lags are distributed in intervals of ±0.0001 s on either side of their noiseless values.

The 100% confidence limits for the Cartesian x, y, and z coordinates of each receiver are obtained from the lags (e.g., Figs. 4 and 5). Locations improve significantly with the number of calls processed. *A priori* errors of ±3.94 m decrease to values between 0.10 and 3.3 m (Table II). Surveying accuracy is better for receivers 2 and 3. This may occur because *a priori* errors are zero for receiver 2's y and z components and for receiver 3's z component. Receiver 6 is less accurately navigated, perhaps because it is higher, by at least 4 m, than any animals, and the other receivers are within 2 m of elevation of the calls. The animal calls contain some information to locate the vertical coordinates (Table II), but because the geometry is mostly horizontal, vertical coordi-
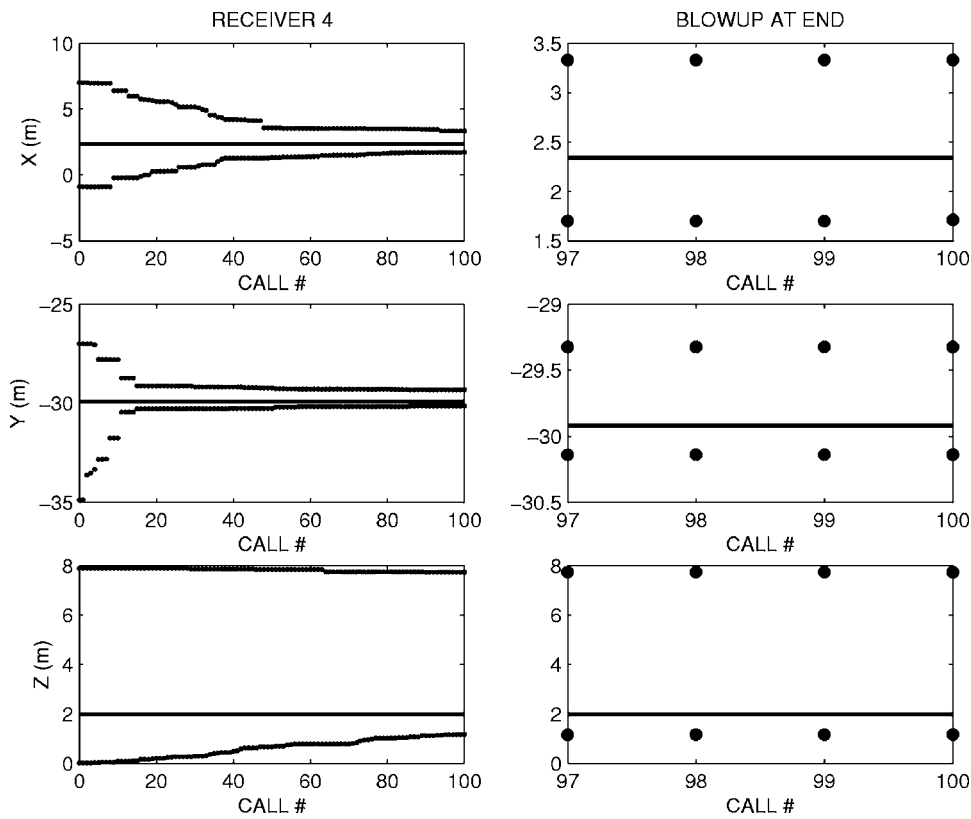
FIG. 4. Example of surveying receiver locations from animal calls at unknown locations near the elevations of the receivers. 100% confidence limits (astericks) for the horizontal (x, y) and vertical z limits for receiver 4 in Fig. 3 are given as a function of the number of calls used for surveying. Each Cartesian coordinate of a receiver is initially known within ±3.94 m. The correct Cartesian coordinates (solid line) are bounded by the 100% limits. Results from the last four calls are shown in the blow-up in the right-column.

nates are not surveyed as accurately. The ratio of *a priori* to final surveying accuracy varies from 39 to 5 for x and y components (Table II).

The number of distributions estimated from each call is 40 (12 Cartesian receiver coordinates, three Cartesian animal coordinates, one homogeneous component of speed, six inhomogeneous components of sound speed, three Cartesian wind components, and 15 lags).
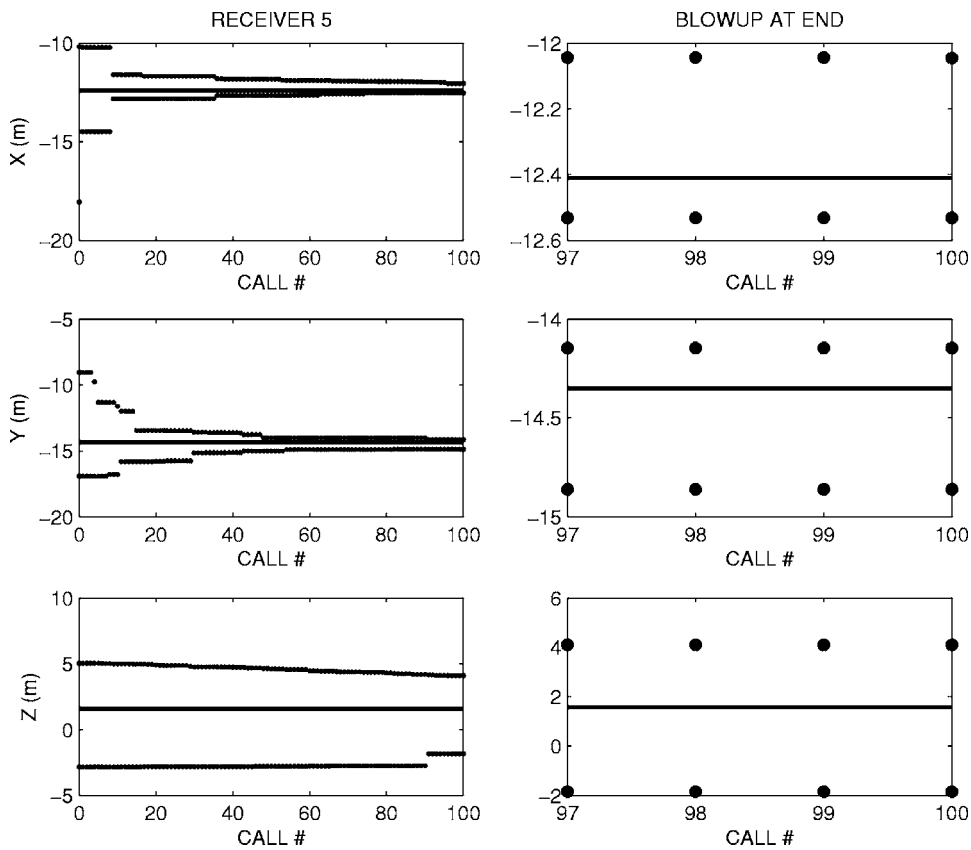
FIG. 5. Same as Fig. 4 except this is receiver 5.

TABLE II. 100% confidence intervals for x, y, and z Cartesian coordinates of receivers 2–6 after using sequential bound estimation on the lags from 100 animal calls at unknown locations (e.g., Figs. 4 and 5). For example, the 100% limits for the x component of receiver 2 are ±0.10 m about a mean value. The mean value is not shown but the mean value plus and minus the indicated limit encompasses the correct coordinate of the receiver in all cases. The *a priori* errors of each Cartesian coordinate are ±3.94 m. By definition, the y coordinate of receiver 2 and the z components of receivers 2 and 3 are 0 (Sec. III D).

Receiver no.	±100% Confidence limits		
	x (m)	y (m)	z (m)
2	0.10	0.00	0.0
3	0.20	0.13	0.0
4	0.81	0.41	3.3
5	0.24	0.36	3.0
6	1.2	0.50	2.9

2. Experiment

Sequential bound estimation is used to provide information about the location of the x coordinate of receiver 2 from Sec. III A (Table I). The information is provided only from a single call of a Red-winged Blackbird recorded on five microphones.

A priori statistics are the same as Sec. III A for isodiachronic location except as follows. The reference location of receiver 2 is changed from (19.76,0,0) to (21,0,0). It's a *a priori* probability distribution is uniform within ±1.5 m of its mean of 21 m. It is important that the algorithm does not know that the x coordinate of receiver 2 is really 19.76 ± 0.05 m.

We find that the 95% confidence limits for the x coordinate of receiver 2 are 19.53 and 21.57 m. This is consistent with the correct answer of 19.76 ± 0.05 m. Its probability distribution is skewed toward the correct answer of 19.76 m (Fig. 6). Sequential bound estimation can estimate joint distributions. For example, the joint distribution of the x coordinates of receiver 2 and the bird are highly correlated (Fig. 7).

IV. ANOTHER KIND OF LOCATION PROBLEM

To illustrate sequential bound estimation in another context, we show a different way to locate a calling whale or other source (Fig. 8). An array is towed along course θ_p at time p when a calling whale is detected at bearing β_p at sound pressure level Z_p from a beamformer. With array center at $\mathbf{r}_p$, the location of the whale is

$$\mathbf{s}_p = \mathbf{r}_p + \rho_p \sin[\theta_p + \beta_p]\hat{\mathbf{i}} + \rho_p \cos[\theta_p + \beta_p]\hat{\mathbf{j}}, \qquad (14)$$

where the distance from the array center to the whale is ρ_p, and unit vectors pointing north and east are $\hat{\mathbf{j}}$ and $\hat{\mathbf{i}}$, respectively. At each time p, we take measurements of array location, tow direction, beam angle, and source level. Source level is converted to distance using some algorithm. For $p = 1$, we assign upper and lower bounds and *prior* distributions for each measurement and for distance ρ_1 given by $f(\mathbf{r}_p)$, $f(\theta_p)$, $f(\beta_p)$, and $f(\rho_p)$, respectively. For example, the *prior* distribution for distance could be estimated from source level and be uniformly distributed between 10 and 40 000 m. A *prior* region is assigned to the location of the whale, e.g., within a radius of 500 km from the array. Configurations of variables are computed by drawing from their *prior* distributions. Locations for each configuration are obtained from Eq. (14). Invalid configurations are discarded (those with the whale outside its *a priori* bounds), leaving valid configurations. We transition the *posterior* bounds for all variables and spatial bounds for the whale to the time of the next datum, at $p = 2$. Since each variable corresponds to a datum, *prior* bounds are assigned to each datum at $p = 2$, and these are intersected with those transitioned from the data that were transitioned from $p = 1$ (Fig. 2). Then, uniform distributions are assigned to all variables within their new intersected bounds, and configurations are drawn from these distributions. Probability distributions for any variable can be made for each time p. Sequential bound estimation may be as accurate or more accurate than any other Bayes-type estimation scheme since the *prior* distributions of the variables are usually not known very well in this scenario.

V. OTHER CONSIDERATIONS

Bounds taken from the set of valid configurations almost always define a smaller interval than the *prior* bounds for each variable. We would not want to accept a smaller bound if it is likely to have occurred by chance because we draw from a *finite* number of samples. This "bound-creep" can be mitigated using standard statistical techniques. Bound-creep mitigation is used in some of the examples in Sec. III.

Experience with many different kinds of problems is needed to appreciate convergence criteria for distributions. In analogy to many Markov chain Monte Carlo approaches, convergence for sequential bound estimation is checked by increasing the number of valid configurations until no significant change is found. Convergence could be difficult when one seeks distributions for which a pertinent event occurs with small probability. On the other hand, there is no advantage in adopting a Gaussian distribution for a *prior* for
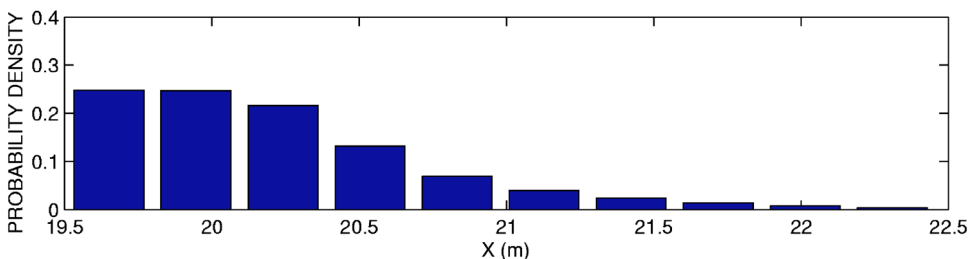
FIG. 6. Probability density function for the Cartesian x coordinate of receiver 2 (Table I) as determined from sequential bound estimation. The correct answer for the x coordinate is 19.76 ± 0.05 m.
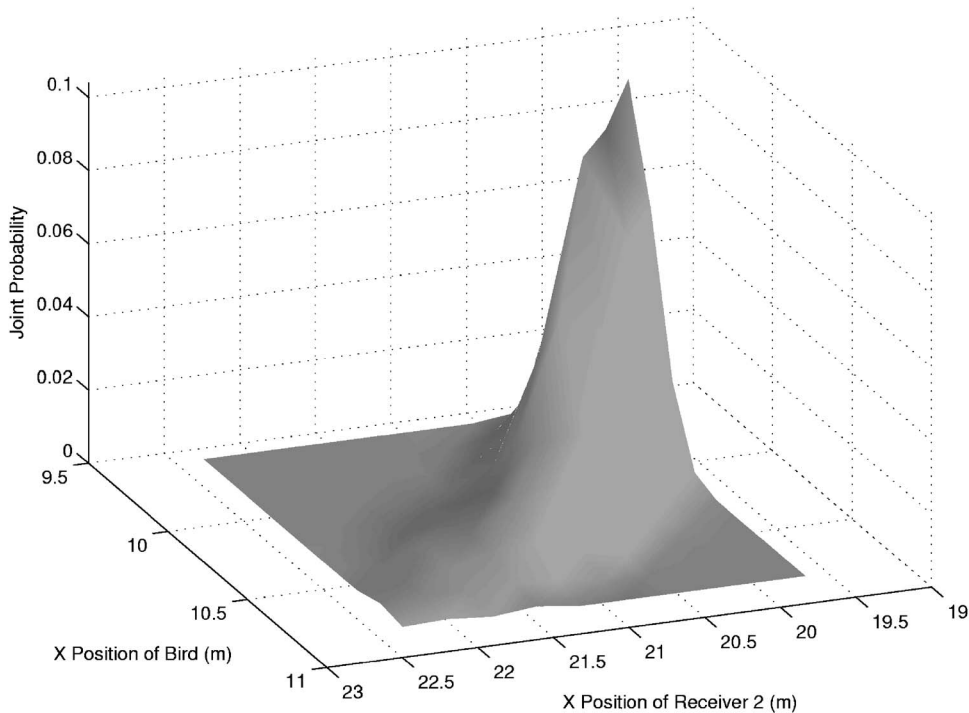
FIG. 7. Joint probability distribution, $f(r_2(x), s(x))$, from valid configurations of random variables following use of data from the fifth receiver constellation. The x coordinates of receiver 2 and the bird location are $r_2(x)$ and $s(x)$, respectively. The marginal distribution for $r_2(x)$ is shown in Fig. 6. The joint distribution shows the probability of joint occurrence for bin intervals of 0.3 and 0.14 m for $r_2(x)$ and $s(x)$, respectively.

the sake of analytical beauty when one distribution is usually as easy to draw from as another, such as a truncated Gaussian. Because data are finite, measurements cannot strictly obey a Gaussian distribution.

We define *efficiency* as the number of valid configurations divided by the number of configurations,

$$E \equiv \mathcal{V}/\mathcal{C}. \tag{15}$$

An efficiency of 1 means that each configuration is a valid configuration. Efficiency increases (decreases) when variables approach statistical independence (dependence). In Sec. III, the efficiency varies roughly between $O(1)$ and $O(10^{-3})$ and depends on the example and receiver constellation in question. Example III D 1 (receiver surveying)

FIG. 8. An array located between the tic marks is towed along course θ_p degrees True at time p. At this time, the center of the array is located at $\mathbf{r}_p$ and the bearing angle to the whale call is β_p.

has an efficiency of about 10^{-3} and takes about 12 days to run on an Advanced Micro Devices Athlon 1800+ CPU. Example III D 2 (bird location) has an efficiency of about 0.3 and takes 13 min to run. Little effort has been made to optimize run times for the software written in the MATLAB programming language.

VI. CONCLUSION

A new sequential Monte Carlo algorithm called "sequential bound estimation" was used to estimate the probability distributions of all pertinent random variables, numbering 13 to 41, in situations where calls from animals were recorded on widely separated receivers. The algorithm was efficient enough to provide distributions for all variables because it did not attempt to propagate joint probability distributions between the use of subsequent data as done using Bayes theorem,[14] Gibbs samplers,[14] and particle filters.[15] These other techniques may be too inefficient to estimate the distributions of so many variables and that is evidently why the literature does not appear to report solutions for the problems dealt with in this paper. Sequential bound estimation was able to estimate the distributions of the receiver locations, speeds of sound, winds, lags, and animal locations by sequentially processing lags between pairs of receivers. The algorithm is robust and leads to distributions that bound the distributions that would be obtained from Bayes theorem if that theorem was practical to implement. Sequential bound estimation need not make any linear approximation between the variables and the data, and it is able to use any realistic *prior* distributions for each of the variables in the problem. Like a Kalman filter, the algorithm is flexible enough to incorporate any model for the transition of all the variables between sequential use of the data.
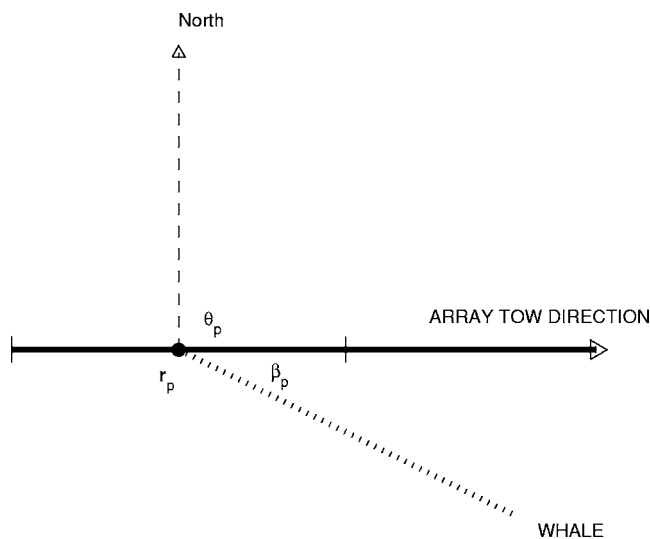
The covariance intersection method[26] is used for Kalman filters when the correlation between data is unknown, but is boundable.[26] Sequential bound estimation is somewhat analogous to that method in that bounds of distributions are sequentially propagated between data without propagating joint distributions. The latter approach appears to be useful because *prior* distributions are often unknown and it is too inefficient to propagate joint distributions of many variables using other sequential Monte Carlo methods.[15]

According to the new algorithm, the lags of travel time between widely separated receivers contain significant information about the location of a calling animal, the speed of sound in air, and the locations of receivers whose initial locations are poorly known. The algorithm indicates that large *prior* uncertainties for the speed of sound or receiver locations can be improved by factors of one or two orders of magnitude by using the information from the lags. One practical use of the new algorithm may be to use animal calls to accurately survey the locations of receivers that may be too expensive or inconvenient to navigate by other means.

ACKNOWLEDGMENTS

[1] R. A. Walker, "Some intense, low frequency, underwater sounds of wide geographic distribution, apparently of biological origin," J. Acoust. Soc. Am. **35**, 1816–1824 (1963).

[2] W. A. Tyrrell, "Design of acoustic systems," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford, 1964), pp. 65–86.

[3] W. A. Watkins and W. E. Schevill, "Four hydrophone array for acoustic three-dimensional location," Woods Hole Oceanographic Technical Report (1971), pp. 71–60.

[4] W. A. Watkins and W. E. Schevill, "Listening to Hawaiian Spinner Porpoises, *STENELLA CV. LONGIROSTRIS*. with a three-dimensional hydrophone array," J. Mammal. **55**, 319–328 (1974).

[5] W. C. Cummings and D. V. Holliday, "Passive acoustic location of bowhead whales in a population census off Point Barrow, Alaska," J. Acoust. Soc. Am. **78**, 1163–1169 (1985).

[6] C. Clark, W. T. Ellison, and K. Beeman, "Acoustic tracking of migrating bowhead whales," in *Oceans 86* (IEEE Oceanic Eng. Soc., New York, 1986), pp. 341–346.

[7] B. Mohl, M. Wahlberg, P. T. Madsen, L. A. Miller, and A. Surlykke, "Sperm whale clicks: Directionality and source level revisited," J. Acoust. Soc. Am. **107**, 638–648 (2000).

[8] C. O. Tiemann and M. B. Porter, "A comparison of model-based and hyperbolic localization techniques as applied to marine mammal calls," J. Acoust. Soc. Am. **114**, 2406 (2003).

[9] J. L. Spiesberger and M. Wahlberg, "Probability density functions for hyperbolic and isodiachronic location," J. Acoust. Soc. Am. **112**, 3046–3052 (2002).

[10] M. Wahlberg, "Estimating source position accuracy in acoustic localization of marine animals," Master thesis, Department of Naval Architecture and Ocean Engineering, Chalmers University of Technology, Sweden, 1999, pp. 29–30.

[11] A. Gelb, *Applied Optimal Estimation* (MIT, Cambridge, MA, 1974).

[12] H. T. Vincent II, "Models, algorithms, and measurements for underwater acoustic positioning," Ph.D. dissertation, University of Rhode Island, 2001.

[13] M. Wahlberg, S. G. Lunneryd, and H. Westerberg, "The source level of harbour seal flipper slaps," Aquat. Mammals **28**(1), 90–92 (2002).

[14] B. P. Carlin and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis* (Chapman and Hall/CRC, New York, 2000).

[15] *Sequential Monte-Carlo Methods in Practice*, edited by A. Doucet, J. F. G. deFreitas, and N. J. Gordon (Springer-Verlag, New York, 2001).

[16] Z. H. Michalopoulou, X. Ma, M. Picarelli, and U. Ghosh-Dastidar, "Fast matching methods for inversion with underwater sound," Proc. Oceans 2000, **1**, 647–651 (2000).

[17] X. Ma and Z.-H. Michalopoulou, "Inversion in an uncertain environment using linearization and ray path arrivals," Center for Applied Mathematics and Statistics Research Report 0001-33, New Jersey Institute of Technology, Newark, N. J., 07102 (2001).

[18] S. E. Dosso and N. E. B. Collison, "Acoustic tracking of a freely drifting sonobouy field," J. Acoust. Soc. Am. **111**, 2166–2176 (2002).

[19] P. Elisseeff, H. Schmidt, M. Johnson, D. Herold, N. R. Chapman, and M. M. McDonald, "Acoustic tomography of a coastal front in Haro Strait, British Columbia," J. Acoust. Soc. Am. **106**, 169–184 (1999).

[20] Z. H. Michalopoulou and U. Ghosh-Dastidar, "Tabu for matched-field source localization and geoacoustic inversion," J. Acoust. Soc. Am. **115**, 135–145 (2004).

[21] R. O. Schmidt, "A new approach to geometry of range difference location," IEEE Trans. Aerosp. Electron. Syst. **AES-8**, 821–835 (1972).

[22] S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).

[23] M. Wahlberg, "Comparing a linear with a non-linear method for acoustic localization," Can. Acoust. **32**, 125–131 (2004).

[24] J. L. Spiesberger, "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**, 837–846 (1999).

[25] C. W. Helstrom, *Statistical Theory of Signal Detection* (Pergamon, New York, 1975), pp. 470 and 276.

[26] L. Chen, P. O. Arambel, and R. K. Mehra, "Estimation under unknown correlation: Covariance intersection revisited," IEEE Trans. Autom. Control **47**, 1879–1882 (2002).

The influence of cortical end-plate on broadband ultrasound attenuation measurements at the human calcaneus using scanning confocal ultrasound

Yi Xia, Wei Lin, and Yi-Xian Qin[a]
Department of Biomedical Engineering, Stony Brook University, Stony Brook, New York, 11794

Quantitative ultrasound (QUS) assessment, including broadband ultrasound attenuation (BUA), is an efficient technique for assessing bone quality in various statuses, e.g., osteoporosis. While assessing trabecular bone loss is essential to bone quality, the existence of cortical bone can substantially reduce the accuracy of BUA measurement. In this study, we developed an approach to quantify the influence of the cortical end-plate in the QUS on 18 cadaver calcanei using both analytical and experimental analyses. A simplified cortical-trabecular-cortical sandwich model has been developed for simulation of wave propagations. Results show that the cortical end-plate has a significant effect on BUA (yielding 8.5 ± 3.6 dB/MHz in cortical bone alone), approximately 15% of the BUA value over the whole bone BUA measurement (54.1 ± 20.1 dB/MHz). The phenomenon has been predicted by the developed analytical model with a high correlation ($r^2 = 0.63, p < 0.0001$). The data have suggested that the mechanism of the BUA attributed to the cortical end-plate is primarily due to the ultrasonic wave transmission and reflection within the cortical layers. Therefore, the influence of the cortical end-plate in BUA can be quantified and incorporated into the QUS assessment for bone quality, which may provide insight into BUA measurement for accurate diagnosis of bone diseases. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1979428]

I. INTRODUCTION

Osteoporosis is a widespread public health problem (Lin and Lane, 2004; Melton *et al.*, 1998). Without symptoms except for the painless progressive bone loss, osteopenia and osteoporosis are often called the "silent disease." In the U.S. alone, 10 million individuals already have the disease and an additional 34 million have low bone mass, accounting for 55% of the population 50 years of age and older (National Osteoporosis Foundation Fact Sheet, 2005). Currently, there is no cure for this disease. Hence, early diagnose of osteoporosis that can monitor bone quality and predict fracture risk is extremely important for both the treatment and the prevention of osteoporosis.

Recent development in quantitative ultrasound (QUS) with imaging capability provides a new physical modality to characterize bone quality, which is noninvasive, radiation free, repeatable, safe, relatively accurate, and suitable for multiple skeleton sites (Barkmann *et al.*, 2000), Laugier *et al.*, 1996). Research has shown that QUS measured broadband ultrasound attenuation (BUA) can measure structural information in addition to density information (Laugier *et al.*, 1997; Grimm and Willams, 1997; Gluer *et al.*, 1994), which can improve the estimation of bone strength and fracture risk (Njeh *et al.*, 2001). This has been verified by both theoretical analysis (Lin *et al.*, 2001) and *in vitro* studies of

trabecular bone samples (Chaffai *et al.*, 2002; Nicholson *et al.*, 2001; Cortet *et al.*, 2003). *In vivo* studies suggested that BUA measured at human calcaneus is capable to predict fracture risk independent of bone density in both men and women (Bouxsein and Radloff, 1997; Bauer *et al.*, 1997; Huopio *et al.*, 2004; Hadji *et al.*, 2000; Khaw *et al.*, 2004; Welch *et al.*, 2004) and may indicate the fracture type (Drozdzowska and Pluskiewicz, 2002).

Human calcaneus is the most commonly used site in QUS measurement. Compared to the cortical bone, the quantity and quality of trabecular bone inside is a more vital indicator of osteoporosis and fracture risk because of its early response to bone loss induced by diseases such as osteopenia (Njeh *et al.*, 1999). The diagnostic accuracy depends primarily on the accurate measurement of the BUA value, particularly from the trabecular region that has the most incidence of osteoporotic fracture. The physical mechanism of BUA from human calcaneus is frequency-dependent attenuation primarily due to the highly attenuating trabecular bone structure (Chaffai *et al.*, 2000). The existence of the cortical shell often reduced the accuracy of the measurement due to the interference of ultrasound wave propagation thought the cortical end-plate. Langton *et al.* (1996; Langton and Subhen, 2001) have demonstrated that a significant artifact in the calcaneus BUA in the trabecular region is related to the presence of the calcaneal cortical end-plate. The identification of the effect of the cortical end-plate on the measurement of BUA would greatly enhance the capability of QUS in noninvasive assessment for bone diseases, particularly in the trabecular bone region.

---
[a]Author to whom correspondence should be addressed; Department of Biomedical Engineering Psychology-A Bldg., 3rd Floor, Stony Brook University, Stony Brook, NY 11794-2580.
Electronic mail: yi-xian.qin@sunysb.edu
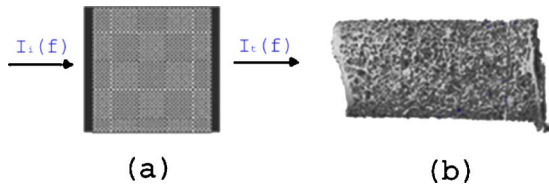
FIG. 1. (Color online) A cortical-trabecular-cortical sandwich structure (a) is proposed in the analytical acoustic model to mimic the intact calcaneus sample structure (b) extracted from the center of calcaneus. The central axis of the structure is along the medial-lateral direction of the calcaneus. Ultrasound waves enter the structure at the left with intensity $I_i(f)$ and exit at the right with intensity $I_t(f)$.

In this study, we hypothesized that the transmission coefficient of an ultrasound wave through the cortical layer of human calcaneus is frequency dependent and is quasilinear in the frequency range where BUA is measured. Thus it may alter the true BUA value in the trabecular region of human calcaneus. The primary objectives of this study were then to quantify the BUA artifact induced by the cortical end-plate experimentally, and to explain and verify the influence of cortical bone in an analytical acoustic layer model.

II. METHODS

A. Numerical model

A cortical-trabecular-cortical sandwich structure is proposed to simulate the ultrasound transmission coefficient in the medial-lateral direction of the intact calcaneus sample (Fig. 1). The structure begins with a thin solid layer of cortical plate at the left, a trabecular bone region in the middle, and ends with another paralleled thin layer of cortical bone at the right. The entire structure is immersed in water and an ultrasound wave enters the structure at the left cortical layer and exits the structure at the right cortical layer. Only the normal incident wave is considered in the model, and therefore it is assumed that no shear wave is generated in the cortical plates. A homogenous and isotropic material property is assumed for the cortical plate. The setup of this model mimics the *in vivo* measurement condition of diagnostic ultrasound, in which the medial-lateral axis of the sample is aligned with the ultrasound pathway and the ultrasound wave

passes through the first cortical end-plate, the fluid saturated trabecular region, and the second cortical plate in sequence.

BUA in this model is determined by the ultrasound attenuation coefficient $A(f)$, which is a function of frequency f, and can be mathematically described as follows:

$$A(f) = 10 \log_{10}\left(\frac{I_i(f)}{I_t(f)}\right). \tag{1}$$

$I_i(f)$ is the incident wave intensity and $I_t(f)$ is the transmitted wave intensity. The attenuation coefficient is the combination that resulted from both the cortical end-plates and the trabecular bone. Each section can be regarded as a wave intensity modulator with different transfer functions:

$$I_t(f) = T_{icl}(f)T_{it}(f)T_{icr}(f)I_i(f), \tag{2}$$

where $T_{icl}(f)$, $T_{it}(f)$, and $T_{icr}(f)$ are the intensity transmission coefficients of the left cortical plate, trabecular region, and right cortical plate respectively. Therefore, the $A(f)$ can be represented in terms of transfer functions as indicated in Eq. (3):

$$A(f) = 10 \log_{10}\left(\frac{1}{T_{icl}(f)}\right) + 10 \log_{10}\left(\frac{1}{T_{it}(f)}\right)$$
$$+ 10 \log_{10}\left(\frac{1}{T_{icr}(f)}\right). \tag{3}$$

Trabecular bone has the characteristic property that its attenuation coefficient is approximately linearly proportional to the frequency. Therefore, the influence of cortical plates $A_{cor}(f)$ on the attenuation coefficient can be separated from trabecular bone in the equation as shown below:

$$A_{cor}(f) = \left[10 \log_{10}\left(\frac{1}{T_{icl}(f)}\right) + 10 \log_{10}\left(\frac{1}{T_{icr}(f)}\right)\right]. \tag{4}$$

The ultrasound wave is reflected and transmitted at the water-cortical and cortical-trabecular boundaries due to the acoustic impedance mismatch, i.e., the speed of sound and density combination in the materials. The intensity transmission coefficients of cortical plates can be derived from the classic acoustic transmission coefficient through a solid layer (Kinsler *et al.*, 2000):

$$T_{icl}(f) = \frac{4}{2 + (r_w/r_t + r_t/r_w)\cos^2(kL_l) + (r_w r_t/r_c^2 + r_c^2/r_w r_t)\sin^2(kL_l)}, \tag{5}$$

$$T_{icr}(f) = \frac{4}{2 + (r_t/r_w + r_w/r_t)\cos^2(kL_r) + (r_t r_w/r_c^2 + r_c^2/r_t r_w)\sin^2(kL_r)}, \tag{6}$$

$$k = \frac{2\pi}{\lambda}, \quad r_t = \rho_t c_t, \quad r_w = \rho_w c_w, \quad r_c = \rho_c c_c. \tag{7}$$

In the above equations, L_l and L_r are the thicknesses of the left and right cortical plates. r_w, r_c, and r_t are the sound

impedance in water, cortical bone, and the trabecular region, respectively, which is the product of medium density ρ and ultrasonic wave velocity c within the medium. k is the wave number derived from wavelength λ. Those parameters are constant except that the wavelength λ is the function of fre-

Xia, Lin, and Qin: Cortical shell influence in acoustic assessment

quency. The relationship between cortical plate intensity transmission coefficient and frequency in this model can be established if the above-mentioned parameters can be reasonably quantified. Since the density in the trabecular region is contributed by both solid trabecular bone material and the water filled in the trabecular void, the ultimate density used in the model is the summation of trabecular material density and the water density with the ratio determined by the trabecular bone volume fraction (BVF) from micro-CT scanning:

$$\rho_t = \text{BVF} \cdot \rho_{tm} + (1 - \text{BVF})\rho_w. \qquad (8)$$

ρ_{tm} is the trabecular material density, essentially similar to the density of cortical material ($\rho_{tm} = \rho_c$) (Turner *et al.*, 1999). The thickness of the cortical plates and medium density in the trabecular region can be measured individually for each sample, and therefore the results of numeric simulation can be verified experimentally.

B. Experimental setup

1. Sample preparation

Eighteen human cadaver calcanei were obtained at the Stony Brook Medical School and were kept frozen at −40 °C. Prior to the ultrasound measurement, the sample was thawed to room temperature and soft tissue was removed. The center of the calcaneus was marked and a cylindrical sample in the medial-lateral direction was extracted using a coring drill with the cortical plates intact. The cylindrical samples had relatively flat and parallel surfaces with diameters of around 14 mm. Samples were first measured by micro-CT and then subject to two consecutive ultrasound measurements before and after cortical end-plates were removed using diamond wheel saw.

2. Micro-CT measurement

Micro-CT measurements were performed with 20-μm resolution on the calcaneus cylinders with intact cortical end-plates (Scanco, Micro-CT 40). A series of 2D bone slice images was obtained and a 3D image [Fig. 1(b)] was reconstructed for each bone sample by stacking the 2D slices together. From the reconstructed 3D image, cortical end-plates can be clearly differentiated and the thicknesses of cortical end-plates were measured using the scale bar from the same 3D image and recorded. The thickness of the cortical shell was then determined at both end-plates and averaged to two numbers at both cortical sides. In addition, the ratio of bone volume to total volume (BVF), trabecular number (Tb.N), thickness (Tb.Th), and spacing (Tb.Sp) at the trabecular bone region of each sample were recorded, which were automatically determined by the embedded Micro-CT software.

3. BUA measurement

Experimental measurement of BUA was performed under room temperature using the newly developed scanning confocal acoustic diagnostic (SCAD) system (Qin *et al.*, 2001) in our lab. It consisted of a computer-controlled 2-D scanner unit with focused transmitter and receiver transducers (1 MHz, Panametrics V303). Transducers are coaxially aligned to each other. The transmitter was driven by pulse signals and the signals passed through bone were received by the receiver amplifier unit (Panamertrics, Model 5072PR) and digitized at 25 MHz using a high-speed digitizer (Gagescope, CS1250) embedded in the computer (Dell Dimension). The control software was written in C++ language.

In the SCAD measurement, signals transmitted through bone were recorded as a sample wave over a maximum 40 × 40-mm^2 image array (0.5-mm pixel size). All these waves were processed to calculate the ultrasound attenuation (ATT; dB), the logarithm of the ratio of the reference wave energy to sample wave energy, and the BUA (dB/MHz), and the slope of the frequency-dependent attenuation between 0.3 and 0.7 MHz (Langton *et al.*, 1984), and then to generate ATT and BUA images. Although 0.3 to 0.7 MHz may not be the optimal frequency range for the 1-MHz broadband ultrasound transducer, it can still generate an ultrasound signal with good strength and signal-to-noise ratio in this range. The received signal was on the order of 300 mV (peak-peak) after proper amplification (20 dB). The cross section of the bone sample, which was perpendicular to the ultrasound pathway, can be easily identified in the ATT image with a contour delineating the shape of the bone sample. A suitable sized region of interest (ROI) can be determined from the ATT image and applied to the corresponding BUA image to calculate the average BUA (Qin *et al.*, 2001).

To experimentally measure the BUA artifact introduced by the cortical plates, each calcaneus sample took two consecutive scans before and after cortical plates were removed. For each scan, a 28 × 28 mm^2 ATT image was obtained. The location of the bone sample can be easily recognized in the image with the size around 15 × 15 mm^2, and an 8 × 8 mm^2 ROI located at the center of the sample was determined. This ROI was then applied to the corresponding BUA image to determine the average BUA in each scan. The BUA difference in the two consecutive scans was recorded for each sample.

C. Statistical analysis

Linear regression and correlation analysis were performed on the BUA data with Microcal Origin 6.0 to examine the capability with which the model predicts the actual BUA artifact. The normality of experimental data was determined by a one-sample Kolmogorov-Smirnov test using SPSS software. The experimental data were normally distributed and a paired two-tail *t* test was used to evaluate the difference between experimental and model calculated data using the same SPSS software, in which a *p* value less than 0.05 was considered significant.

III. RESULTS

Table I listed the summary of the data (mean±standard deviation) obtained from the experiment. The cortical endplate thickness for the bone samples was 0.6±0.3 mm for the medial side and 0.3±0.1 mm for the lateral side, and the BVF in the trabecular region of these samples ranged from 5% to 30%. The BUA values measured from samples with intact cortical end-plates were larger than the BUA values measured after cortical plates were removed. The difference

TABLE I. Summary of experiment data; $N = 18$.

	Original length (mm)	Trabecular length (mm)	Cortical plates thickness (mm)			BUA (dB/MHz)	
			Med	Lat	BVF	w plate	w/o plate
Mean±SD	24.2±3.2	20.2±2	0.6±0.3	0.3±0.1	18±6%	54.1±20.1	45.6±19.9

was 8.5±3.6 dB/MHz, approximately 15% of the BUA measured from samples with intact cortical plates (54.1±20.1 dB/MHz).

Numerical simulations were performed by setting the model parameters as follows. The density of water is 1000 kg/m³ and the sound propagation speed in the water is measured around 1480 m/s. The density of the cortical material is 1900 kg/m³ (Cameron *et al.*, 1999) and the acoustic speed in the cortical material is 3600 m/s (Njeh *et al.*, 1999; Williams, 1992). The trabecular region medium density is calculated based on micro-CT measured BVF [Eq. (8)], and ultrasound speed in the trabecular region is averaging 1550 m/s measured in our lab.

By selecting the BVF and cortical plate thickness using the average values observed, a typical relationship between the cortical plate transmission coefficient and frequency can be established [Fig. 2(a)]. As a result, the attenuation coefficient, which has an inverse relationship with the transmission coefficient, was positively linearly proportional to frequency in the range 0.3–0.7 MHz [Fig. 2(b)]. The slope of attenuation to frequency curve was a positive number, which became the artifact imposed on the BUA from the trabecular region. This artifact can be obtained individually for each bone sample by setting the corresponding BVF and cortical end-plates thickness in the numerical model. The averaging BUA artifact calculated by the numerical model was 7.5±3.3 dB/MHz for these samples. This was comparable to the BUA offset observed in the experiment (8.5±3.6 dB/MHz). A significant correlation exists between experiment data and numerical simulation ($R^2 = 0.63, p < 0.0001$) (Fig. 3). Paired two-tail t tests showed that there was no significant difference between experimental and model calculated data ($p = 0.069$).

IV. DISCUSSION

A significant BUA artifact, contributed by the cortical end-plates, which account for approximately 15% of the BUA measured from intact samples, was observed. Good agreement has been achieved between the numerical model and the actual experimentally measured BUA artifact induced by the cortical plates. It has been demonstrated that the artifact was primarily contributed by the frequency-dependent attenuation of the ultrasound wave in the cortical end-plates. This finding was in agreement with published results by Langton *et al.* (1996). The analytical model has the capability to predict the BUA artifact induced by the cortical end-plate by showing a significant correlation between experimental and model calculated values. This model indicates that the frequency-dependent energy transmission is the result of wave cancellation inside the cortical layer. This can

be demonstrated by examining the relationship between transmission coefficient T_{icl} and the thickness to wavelength ratio (L_l/λ) in the cortical plate [Fig. 4(a)]. The percentage of energy that can be transferred through the cortical plate, represented by T_{icl}, decreased from 0.72 (72%) to 0.42 (42%) with the increase of L_l/λ from 0.05 to 0.1. This is the real L_l/λ range in this study with the following calculation. In our study, the acoustic speed c in the cortical plate is set at 3600 m/s. For a finite cortical plate thickness, e.g., the averaged thickness of the medial side cortical plate L_l is 0.6 mm, in the frequency range where BUA is calculated ($f = 0.3-0.7$ MHz), the wavelength, $\lambda = c/f$, is between 5 and

(a)

(b)
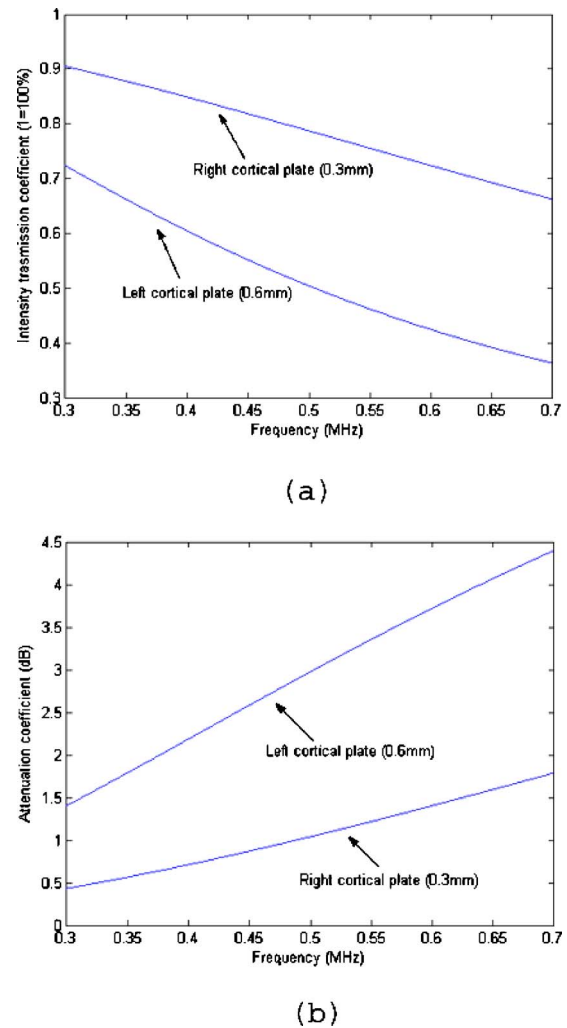
FIG. 2. (Color online) The (a) ultrasound transmission coefficient over frequency in the left and right cortical plates and the (b) resultant ultrasound attenuation coefficient. A linear relationship can be observed between ultrasound attenuation coefficient and frequency. The slope of this linear curve is BUA.

Xia, Lin, and Qin: Cortical shell influence in acoustic assessment
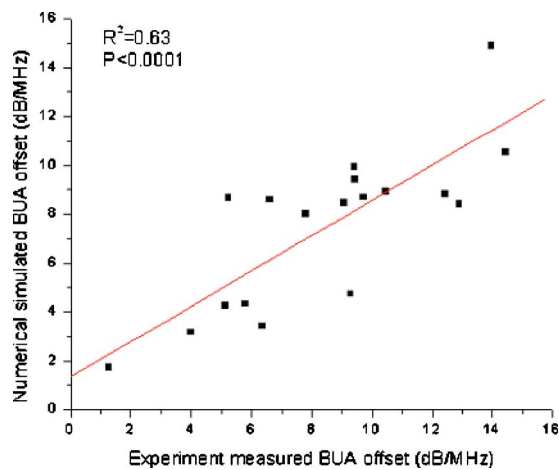
FIG. 3. (Color online) Correlation between the experimentally measured BUA artifact induced by the cortical shell and model prediction. A high correlation exists between them.

12 mm, which determines the L_l/λ ratio in the range of 0.05 to 0.1. In this range, the transmission coefficient has the trend to linearly descend, which is the curving area indicated by the arrow in Fig. 4(a). The higher the frequency, with higher value of L_l/λ for constant L_l, the smaller the portion of energy that will go through. This will eventually lead to a quasilinear ascending attenuation coefficient curve, which is the logarithm of the inverse transmission coefficient, in the same L_l/λ range [Fig. 4(b), indicated by the arrow]. The slope of the curve area indicated by the arrow in Fig. 4(b) is a positive number, and this is the artifact induced by the cortical end-plate in the BUA measurement. All these suggest that the thin layer of cortical plate is acting like a sound intensity modulator, in which stronger interference and phase cancellation between waves may take place when the wavelength of higher frequency components in the ultrasound wave are short enough to approach the thickness of the cortical plate. This will lead to the decreasing energy transfer with increasing frequency, and may be the origin of the BUA artifact in the cortical plate. Thus, if the influence of the cortical end-plates in the BUA measurement can be quantified, the accuracy of the BUA assessment in the trabecular region can be greatly improved.

Figure 5 showed the simulation results by (a) fixing BVF at 18% and varying cortical shell thickness, L_l, from 0.2 to 1 mm with a 0.2-mm interval and (b) fixing the thickness at 0.5 mm and varying BVF from 5% to 30% with 5% intervals. For a typical BVF value, 18% as the averaged value, varying cortical plate thickness leads to different frequency-attenuation coefficient relations; the slope of the curve increases with thicker cortical end-plate from 0.2 to 1 mm. At the same time, the curvature of the relation is concave when thickness is smaller than 0.6 mm and convex when thickness is bigger than 0.6 mm. It is clear from the simulation result [Fig. 5(a)] that the attenuation coefficient is positively quasilinear proportional to the frequency in the range from 0.3 to 0.7 MHz with thickness in the range presented in this simulation. If cortical plate thickness is fixed, e.g., 0.5 mm in this case, varying structural parameter, e.g., BVF, in the trabecular region can also change the slope

(a)

(b)
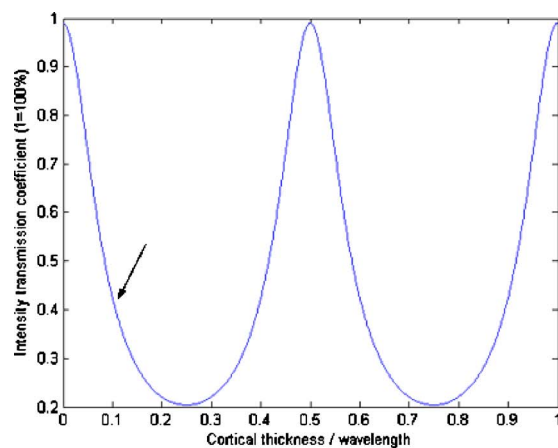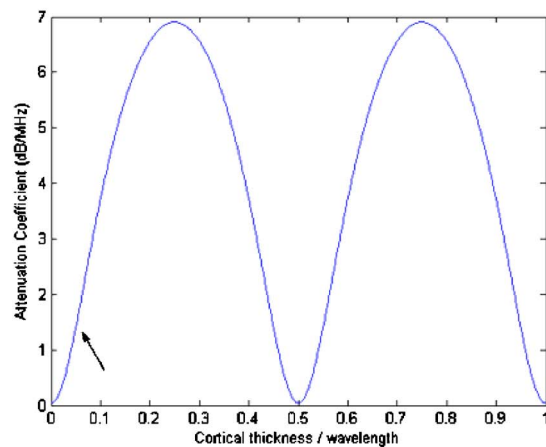
FIG. 4. (Color online) The change of (a) cortical plate intensity transmission coefficient and (b) attenuation coefficient with L/λ ratio from 0 to 1. In our study, the L/λ ratio is in the range of 0.05–0.1, which is indicated by the arrow in the picture.

of the frequency-attenuation coefficient curve, i.e., increased slope with decreased BVF value. However, the slope variation in this curve, introduced by varying cortical plate thickness (e.g., from fairly flat to relatively steep), was changed only slightly dependent on BVF changes (e.g., all the curves clustered together with small slope difference). This has suggested that the wave attenuation as dependence of frequency is more likely the result of wave interference and phase cancellation inside the cortical plate when the wavelength approaches the cortical plate thickness. Change of cortical thickness directly leads to the change of the L_l/λ ratio and a more dramatic change in frequency-attenuation curve can be observed.

The simulation did not take into account scattering and refraction factors by assuming these are negligible. The essentially solid cortex of calcaneus should not have significant scattering effect. The angled wave being incident can happen if the cortex surface is not normal to the incident ultrasound beam due to the surface curvature. However, since the incident focused ultrasound beam at each scanning point has a very small beam diameter (BD$\cong$2.5 mm), there will be very
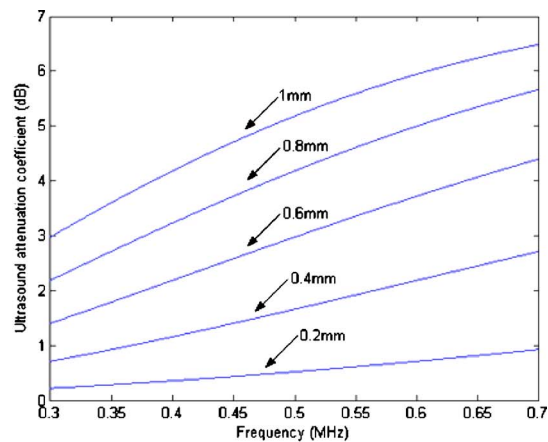
(a)
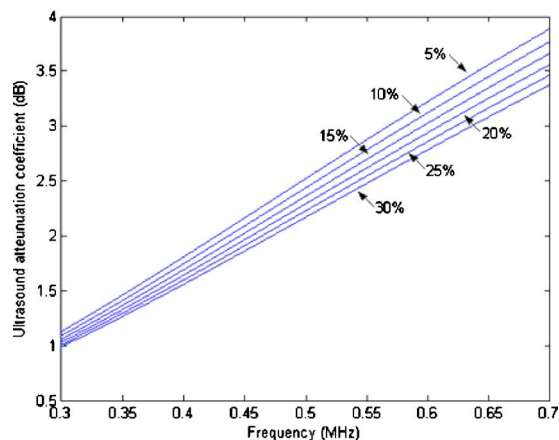
(b)

FIG. 5. (Color online) The relationship between cortical shell induced ultrasound attenuation coefficient and frequency by (a) fixing BVF at 18% and varying cortical shell thickness, L_l and L_r, from 0.2 to 1 mm with 0.2 mm interval and (b) fixing thickness at 0.5 mm and varying BVF from 5% to 30% with 5% interval. Ultrasound attenuation coefficients are approximately linearly proportional to the frequency in all of these cases, from which the BUA artifact can be generated.

limited curvature under such a small beam area and no significant angled wave can be generated. Hence, refraction and shear wave effect may also be assumed to be neglected.

There are very limited studies focusing on the influence of the cortical end-plate on ultrasound BUA measurement. Most *in vitro* calcaneus ultrasound studies are dedicated only on trabecular bone samples. Results from these studies are directly used to explain *in vivo* experiment results with the assumption that the influence of the cortical end-plate can be neglected. Our study shows that a significant BUA artifact can be induced by the cortical plate and the value of this artifact is sensitive to the cortical shell thickness. The variability of calcaneus cortical shell thickness between subjects will induce a random BUA artifact. This may degrade the accuracy of clinical ultrasound in osteoporosis diagnosis, which relies primarily on the ultrasound BUA measurement. The model presented in this study provides a way to quantify the cortical-end-plate-induced BUA artifact. This is helpful to extract the BUA value only from the trabecular region and

potentially improve the accuracy of ultrasound in the assessment of bone quality and predict the risk of fracture particularly in the trabecular region.

The numerical model presented in this study is based on the bold assumption that the cortical end-plates can work as acoustical mediators, providing coefficients ranging from 0 to 1 in dependence on the thickness-to-wavelength ratio. In practice, there is a scope of factors that can compromise this model valid for ideal resonators. These may include nonuniformity of the thickness in projection of ultrasonic beam, absence of a strong border between the cortical plate and the underlying spongy structure, as well as the curvature of the bone surface. Hence, the model is limited in the following aspects. First, as the first-order approximation in this model, a distinct trabecular-cortical interface was assumed, e.g., no interface gradient between cortical shell and trabecular bone. This was not necessarily true in reality. It could be expected that the existence of an interface with certain structural gradients could lead to the overestimation of the BUA artifact by the model. Second, only normal incident wave transmission and reflection were considered in the model. The origin of the BUA artifact was the wave interference inside the cortical plate. However, the limited cortical plate surface curvature may also induce a BUA artifact, due primarily to the phase cancellation of the ultrasound pulse resulting from inconsistencies in propagation time across the ultrasound beam (Langton and Subhan, 2001). The surface-curvature-induced BUA artifact could be the mechanism involved which was not considered in our model. Third, cortical material property was different between subjects, which may be influenced by age, height, and disease, such as osteoporosis. Use of a unified value to represent cortical material in the numerical model may degrade the predictive accuracy of this model. Fourth, the clinical application of this model *in vivo* depends on the accessibility of individual cortical end-plate thickness. This is not easy to achieve under the current technique. Researches performed to investigate the cortical end-plate are necessary to understand and quantify its influence on ultrasound measurement. However, the good agreement between the model simulation data with experimental data demonstrated that this model may be valuable in understanding the interaction between an ultrasound wave and the cortical end-plate.

Barkmann, R., Kantorovich, E., Singal, C., Hans, D., Genant, H. K., Heller, M., and Gluer, C. C. (**2000**). "A new method for quantitative ultrasound measurements at multiple skeletal sites: first results of precision and fracture discrimination," J. Clin. Densitom. **3**, 1–7.

Bauer, D. C., Gluer, C. C., Cauley, J. A., Vogt, T. M., Ensrud, K. E., Genant, H. K., and Black, D. M. (**1997**). "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older

1806   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Xia, Lin, and Qin: Cortical shell influence in acoustic assessment

women. A prospective study. Study of Osteoporotic Fractures Research Group," Arch. Intern Med. **157**, 629–634.

Bouxsein, M. L., and Radloff, S. E. (**1997**). "Quantitative ultrasound of the calcaneus reflects the mechanical properties of calcaneal trabecular bone," J. Bone Miner. Res. **12**, 839–846.

Cameron, J. R., James, G. S., and Roderick, M. G. (**1999**). *Physics of the body* (Medical Physics, Madison, WI).

Chaffai, S., Padilla, F., Berger, G., and Laugier, P. (**2000**). "*In vitro* measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289.

Chaffai, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**). "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Cortet, B., Boutry, N., Dubois, P., Legroux-Gerot, I., Cotten, A., and Marchandise, X. (**2003**). "Does quantitative ultrasound of bone reflect more bone mineral density than bone microarchitecture?" Calcif. Tissue Int. **74**, 60–67.

Drozdzowska, B., and Pluskiewicz, W. (**2002**). "The ability of quantitative ultrasound at the calcaneus to identify postmenopausal women with different types of nontraumatic fractures," Ultrasound Med. Biol. **28**, 1491–1497.

Gluer, C. C., Wu, C. Y., Jergas, M., Goldstein, S. A., and Genant, H. K. (**1994**). "Three quantitative ultrasound parameters reflect bone structure," Calcif. Tissue Int. **55**, 46–52.

Grimm, M. J., and Williams, J. L. (**1997**). "Assessment of bone quantity and 'quality' by ultrasound attenuation and velocity in the heel," Clin. Biomech. (Bristol, Avon) **12**, 281–285.

Hadji, P., Hars, O., Gorke, K., Emons, G., and Schulz, K. D. (**2000**). "Quantitative ultrasound of the oscalcis in postmenopausal women with spine and hip fracture," J. Clin. Densitom. **3**, 233–239.

Huopio, J., Kroger, H., Honkanen, R., Jurvelin, J., Saarikoski, S., and Alhava, E. (**2004**). "Calcaneal ultrasound predicts early postmenopausal fractures as well as axial BMD. A prospective study of 422 women," Osteoporosis Int. **15**, 190–195.

Khaw, K. T., Reeve, J., Luben, R., Bingham, S., Welch, A., Wareham, N., Oakes, S., and Day, N. (**2004**). "Prediction of total and hip fracture risk in men and women by quantitative ultrasound of the calcaneus: EPIC-Norfolk prospective population study," Lancet **363**, 197–202.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics* (Wiley, Hoboken, NJ).

Langton, C. M., and Subhan, M. (**2001**). "Computer and experimental simulation of a cortical end-plate phase cancellation artifact in the measurement of BUA at the calcaneus," Physiol. Meas **22**, 581–587.

Langton, C. M., Palmer, S. B., and Porter, R. W. (**1984**). "The measurement of broadband ultrasonic attenuation in cancellous bone," Eng. Med. **13**, 89–91.

Langton, C. M., Njeh, C. F., Hodgskinson, R., and Currey, J. D. (**1996**). "Prediction of mechanical properties of the human calcaneus by broadband ultrasonic attenuation," Bone (N.Y.) **18**, 495–503.

Laugier, P., Fournier, B., and Berger, G. (**1996**). "Ultrasound parametric imaging of the calcaneus: in vivo results with a new device," Calcif. Tissue Int. **58**, 326–331.

Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.

Lin, J. T., and Lane, J. M. (**2004**). "Osteoporosis: a review," Clin. Orthop. Relat. Res. **425**, 126–134.

Lin, W., Qin, Y. X., and Rubin, C. (**2001**). "Ultrasonic wave propagation in trabecular bone predicted by the stratified model," Ann. Biomed. Eng. **29**, 781–790.

Melton, L. J. III, Atkinson, E. J., O'Connor, M. K., O'Fallon, W. M., and Riggs, B. L. (**1998**). "Bone density and fracture risk in men," J. Bone Miner. Res. **13**, 1915–1923.

National Osteoporosis Foundation (2005). Disease facts, http://www.nof.org/osteoporosis/diseasefacts.htm

Nicholson, P. H., Muller, R., Cheng, X. G., Ruegsegger, P., Van Der Perre, G., Dequeker, J., and Boonen, S. (**2001**). "Quantitative ultrasound and trabecular architecture in the human calcaneus," J. Bone Miner. Res. **16**, 1886–1892.

Njeh, C. F., Fuerst, T., Diessel, E., and Genant, H. K. (**2001**). "Is quantitative ultrasound dependent on bone structure?" Osteoporosis Int. **12**, 1–15.

Njeh, C. F., Hans, D., Fuerst, T., Cluer, C. C., and Genant, H. K. (**1999**). "Quantitative ultrasound, Assessment of osteoporosis and bone status," (Martin Dunitz, London), p. 56.

Qin, Y. X., Lin, W., and Rubin, C. (**2001**). "Interdependent relationship between trabecular bone quality and ultrasound attenuation and velocity using a scanning confocal acoustic diagnostic system." J. Bone Miner. Res. **16**, S470–70.

Turner, C. H., Rho, J., Takano, Y., Tsui, T. Y., and Pharr, G. M. (**1999**). "The elastic properties of trabecular and cortical bone tissues are similar: Results from two microscopic measurement techniques," J. Biomech. **32**, 437–441.

Welch, A., Camus, J., Dalzell, N., Oakes, S., Reeve, J., and Khaw, K. T. (**2004**). "Broadband ultrasound attenuation (BUA) of the heel bone and its correlates in men and women in the EPIC-Norfolk cohort: a cross-sectional population-based study," Osteoporosis Int. **15**, 217–225.

Williams, J. L. (**1992**). "Ultrasonic wave propagation in cancellous and cortical bone: prediction of some experimental results by Biot's theory," J. Acoust. Soc. Am. **91**, 1106–1112.

Signal processing strategies that improve performance and understanding of the quantitative ultrasound SPECTRAL FIT algorithm

Timothy A. Bigelow and William D. O'Brien, Jr.[a]

Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

Quantifying the size of the tissue microstructure using the backscattered power spectrum has had limited success due to frequency-dependent attenuation along the propagation path, thus masking the frequency dependence of the scatterer size. Previously, the SPECTRAL FIT algorithm was developed to solve for total attenuation and scatterer size simultaneously [Bigelow *et al.*, J. Acoust. Soc. Am. **117**, 1431–1439 (2005)]. Herein, the outcomes from signal processing strategies on the SPECTRAL FIT algorithm are investigated. The signal processing methods can be grouped into two categories, viz., methods that improve the performance of the algorithm and methods that provide insight. The methods that improve the performance include compensating for the windowing function used to gate the time-domain signal, averaging the spectra in the normal frequency domain rather than the log domain to improve the precision of the scatterer size and attenuation estimates, improving the selection of the usable frequency range for the SPECTRAL FIT algorithm, and improving the compensation for electronic noise. The methods that provide insight demonstrate that the anomalous rapid fluctuations of the backscattered power spectrum do not affect the SPECTRAL FIT algorithm, and accurate attenuation estimates can be obtained even when the correct scatterer geometry (i.e., form factor) is not known. © 2005 Acoustical Society of America. [DOI: 10.1121/1.2000752]

LIST OF SYMBOLS

a_{eff} = effective radius of scatterer;

$a_{\text{eff}\,j}$ = estimated effective radius of scatterer found from one set (i.e., 25 averaged rf echoes) of simulated backscatter wave forms;

$\bar{a}_{\text{eff}}$ = mean value of estimated effective radius from all sets of backscattered wave forms (i.e., $\bar{a}_{\text{eff}} = \Sigma_{\forall j} a_{\text{eff}\,j} / \Sigma_{\forall j} j$);

ASD = average squared difference term minimized by SPECTRAL FIT algorithm;

c = speed of sound in the tissue;

c_0 = speed of sound in water;

$E[\ ]$ = expected value with respect to scattering random process;

$E_N[\ ]$ = expected value with respect to electronic noise random process;

f = frequency;

f_0 = frequency corresponding to the spectral peak of the Gaussian spectrum, i.e.,

$$\exp\left(-\left(\frac{f-f_0}{\sqrt{2}\sigma_\omega}\right)^2\right);$$

$F_\gamma(f, a_{\text{eff}})$ = form factor related to the scatterer geometry and size;

$f\#$ = f-number for spherically focused transducer;

G_{corr} = windowing correction term for spectrum;

g_{win} = windowing function used to gate the time-domain wave forms;

g_{win2} = windowing function used for homomorphic filtering;

G_{win} = Fourier transform of $g_{\text{win}}(t)$;

$H(f)$ = filtering characteristics of ultrasound source;

k = wave number in tissue;

L = length of the windowing function;

n = frequency exponent when Gaussian form factor is varied [i.e., $F_\gamma(f, a_{\text{eff}}) = \exp(-0.827 \cdot (ka_{\text{eff}})^n)$];

$N(f)$ = additive electronic noise spectrum;

N_{dB} = minimum value allowed for N_{Floor} when no electronic noise has been added;

N_{Floor} = noise floor of system used when selecting usable frequencies;

$P_n, P_p, \bar{P}_n$ = terms used to fit Gaussian distribution to spectrum in log domain;

P_{ref} = reference spectrum [i.e., $P_{\text{ref}}(f) = k_0^4 |V_{\text{inc}}(\omega)|^2 |H(\omega)|^4$];

P_{scat} = $E[|V_{\text{refl}}|^2]$ estimated from set of wave forms;

SNR = signal-to-noise ratio;

s_z = integration variable of field along beam axis;

t = time;

T_{cep} = value used to set the amount of homomorphic filtering;

T_{win} = length windowing function applied to time-domain wave form (i.e., $T_{\text{win}} = 2L/c$);

[a]Electronic mail: wdo@uiuc.edu

$V_{\text{cep}i}$ = rf echo expressed in cepstral domain;

V_{inc} = voltage spectrum applied to ultrasound source;

V_j = voltage spectrum produced by a single back-scattered rf echo;

v_{noise} = example noise signal voltage in time domain (i.e., no signal transmitted by source);

V_{plane} = voltage spectrum produced by backscatter from rigid plane placed at the focal plane [i.e., proportional to $V_{\text{inc}}(\omega)(H(\omega))^2 e^{i2k_0 z_T}$];

V_{refl} = voltage spectrum produced by backscatter from tissue containing scatterers;

$v_{\text{refl}i}$ = voltage of a rf echo in time domain;

w_x, w_y, w_z = equivalent Gaussian dimensions of velocity potential field in focal region;

$X, \bar{X}$ = terms used in minimization scheme for SPECTRAL FIT algorithm;

z_T = distance from ultrasound source's aperture plane to focal plane;

α = total effective attenuation coefficient for tissue between ultrasound source and region containing the scatterers (i.e., assumed linear frequency dependence of the form $\alpha = \alpha_0 f$);

α_1 = local attenuation coefficient of the tissue in the region containing the scatterers;

α_0 = slope of total attenuation coefficient versus frequency;

$(\alpha_0 z_T)_j$ = estimated attenuation along the propagation path for single data set;

$\overline{(\alpha_0 z_T)}$ = mean value for attenuation along the propagation path from all sets of backscattered wave forms [i.e., $\overline{(\alpha_0 z_T)} = \Sigma_{\forall j}(\alpha_0 z_T)_j / \Sigma_{\forall j} j$];

$\sigma_{a_{\text{lower}}}$ = percent deviation in values of scatterer size for sizes smaller than the mean size (i.e., $a_{\text{eff }j} < \bar{a}_{\text{eff}}$);

$\sigma_{a_{\text{upper}}}$ = percent deviation in values of scatterer size for sizes larger than the mean size (i.e., $a_{\text{eff }j} > \bar{a}_{\text{eff}}$);

σ_g = bandwidth term for the Gaussian distribution approximating the windowing function (i.e., $|G_{\text{win}}(f)|^2 \propto e^{f^2/2\sigma_g^2}$);

$\sigma_{\alpha_{\text{lower}}}$ = deviation in dB/MHz in values of attenuation for attenuations smaller than the mean attenuation [i.e., $(\alpha_0 z_T)_j < \overline{(\alpha_0 z_T)}$];

$\sigma_{\alpha_{\text{upper}}}$ = deviation in dB/MHz in values of attenuation for attenuations greater than the mean attenuation [i.e., $(\alpha_0 z_T)_j > \overline{(\alpha_0 z_T)}$];

σ_ω = bandwidth term for Gaussian distribution, i.e.,

$$\exp\left(-\left(\frac{f - f_0}{\sqrt{2}\sigma_\omega}\right)^2\right)$$

τ_{cep} = time values in cepstral domain.

I. INTRODUCTION

In recent years, an explosion of new medical imaging technologies has occurred. For each technology, quantitative information is the desired outcome to further assist the clinician in making a diagnosis or monitoring a therapy. One method for obtaining quantitative ultrasound (QUS) information is to use the frequency dependence of the *in vivo* backscattered power spectrum to estimate a characteristic size of the tissue microstructure (Lizzi *et al.*, 1983; 1986; Insana *et al.*, 1990). These scatterer size estimates, and related processing strategies, have shown potential for differentiating tumor types (Oelze *et al.*, 2004) as well as assessing the effects of ultrasound therapy (Lizzi *et al.*, 1997a). Unfortunately, the full diagnostic potential of scatterer size estimates has not been realized due to frequency-dependent attenuation along the propagation path between the transducer and the scattering region (i.e., total attenuation).

Current size estimation algorithms assume that the total attenuation is known *a priori*. Errors in the assumed attenuation result in inaccuracies in the scatterer size estimate of the same magnitude as the errors in the assumed attenuation. Although tissues with drastically different characteristic sizes can be distinguished based on the scatterer size (using an assumed total attenuation), subtle differences, such as assessing the disease state of liver (Lu *et al.*, 1999), will probably be missed. Excellent clinical results have been obtained for ophthalmologic tumors (Lizzi *et al.*, 1983; 1997a; 1997b; Silverman *et al.*, 2003) as well as carotid plaque (Noritomi *et al.*, 1997; Lee *et al.*, 1998) due to the negligible or known total attenuation (i.e., attenuation of water). In general, however, the uncertainties in the total attenuation reduce the applicability of scatter size for clinical applications.

In addition to limiting the applicability of the size estimates, uncertainties in total attenuation also limit other QUS methods. For example, tissue elasticity can be determined by applying a known external force at the skin surface and estimating the resulting internal tissue displacement (Ophir *et al.* 1991; Skovoroda *et al.*, 1994; O'Donnell *et al.*, 1994; Skovoroda *et al.*, 1995). A related technique, acoustic radiation force impulse imaging (ARFI), uses ultrasound to generate the force within the body that then produces the desired tissue displacement (Nightingale *et al.*, 2000; Lizzi *et al.*, 2003). Unfortunately, the force applied using ARFI is currently unknown due to uncertainties in the radiation force, and hence, the elasticity of the tissue cannot be quantified. However, an estimate of the radiation force could potentially be obtained from an estimate of the total attenuation along the propagation path. Therefore, an accurate estimate of the total attenuation might allow ARFI to quantify tissue elasticity as well.

A new method has been proposed to estimate scatterer size and total attenuation simultaneously using an algorithm termed SPECTRAL FIT (SF) (Bigelow *et al.*, 2005). Using that algorithm, accurate estimates of size and attenuation were obtained for different attenuation values and signal-to-noise levels for larger window lengths. Unfortunately, the precision of the estimates was poor limiting the resolution when forming QUS images of tissue. The poor precision meant that a large number of rf echoes were required before reliable estimates could be obtained increasing the size of the resolution cell. The formation of improved QUS images could potentially allow for early detection of tumors, identification

of a tumor type, or other pathology. The goal of the work reported herein is to investigate different signal processing strategies capable of providing more accurate (i.e., lower mean error) and precise (i.e., smaller deviation) estimates of scatterer size and total attenuation. Successful signal processing techniques included averaging of the power spectra in the normal frequency domain rather than the log domain to improve the precision, correcting for the windowing of the time-domain wave forms to improve the accuracy at smaller window lengths, improving the automatic selection of the frequency range used to obtain the estimates, and improving the compensation of electronic system noise to allow more accurate estimates. Other signal processing strategies did not improve the performance of the estimation scheme; however, these methods are included in this paper because they provide insight into the problem that could guide future investigations.

II. METHODS THAT IMPROVED PERFORMANCE

The signal processing methods that improved the performance of the estimation algorithm will be considered first. A brief review of the SF algorithm is provided before discussing the proposed signal processing strategies.

A. Overview of SPECTRAL FIT algorithm

The expected squared voltage produced by the backscatter returned from a region of interest (ROI) that contains randomly positioned weak scatterers is proportional to (Bigelow and O'Brien, 2004)

$$E[|V_{\text{refl}}(f)|^2] \propto f^4 |V_{\text{inc}}(f)|^2 |H(f)|^4 e^{-4\alpha(f)z_T} F_\gamma(f, a_{\text{eff}})$$
$$\times \int_{-L/2}^{L/2} ds_z (g_{\text{win}}(s_z) e^{-4s_z^2/w_z^2(f)} e^{4\alpha_1(f)s_z}) \quad (1)$$

while the squared voltage produced by the backscatter returned from a rigid plane placed at the focal plane is proportional to (Bigelow and O'Brien, 2004)

$$|V_{\text{plane}}(f)|^2 \propto |V_{\text{inc}}(f)|^2 |H(f)|^4. \quad (2)$$

If the transducer is assumed to be weakly focused (e.g., f/4), the time gate used to window the backscattered rf echoes is relatively small compared to the depth of focus (i.e., local plane wave approximation), the attenuation in the focal region does not significantly alter the spectrum over the length of the window, and the total attenuation has a linear dependence of frequency (i.e., $\alpha = \alpha_0 f$), then the integral in Eq. (1) can be simplified, and the expected squared voltage produced by the backscatter from the tissue's ROI is approximately proportional to

$$E[|V_{\text{refl}}(f)|^2] \propto f^4 |V_{\text{plane}}(f)|^2 e^{-4\alpha_0 z_T f} F_\gamma(f, a_{\text{eff}}). \quad (3)$$

Hence, by comparing the expected voltage from the ROI to the voltage from a rigid plane placed at the focal plane, an estimate for the total attenuation along the propagation path α as well as the scatterer size a_{eff} can be obtained. In the SF algorithm, the comparison is performed by finding the values of α_0 and a_{eff} that minimize the average squared difference (ASD) given by

$$\text{ASD} = \underset{f}{\text{mean}}[(X(f, a_{\text{eff}}, \alpha_0) - \bar{X}(a_{\text{eff}}, \alpha_0))^2], \quad (4)$$

where

$$X(f, a_{\text{eff}}, \alpha_0) = \ln\left(\frac{P_{\text{scat}}(f)}{\max_f (P_{\text{scat}}(f))}\right)$$
$$- \ln\left(\frac{P_{\text{ref}}(f) F_\gamma(f, a_{\text{eff}}) e^{-4\alpha_0 f z_T}}{\max_f (P_{\text{ref}}(f) F_\gamma(f, a_{\text{eff}}) e^{-4\alpha_0 f z_T})}\right), \quad (5)$$

$$\bar{X}(a_{\text{eff}}, \alpha_0) = \underset{f}{\text{mean}}[X(f, a_{\text{eff}}, \alpha_0)].$$

P_{scat} is an estimate of $E[|V_{\text{refl}}(f)|^2]$ obtained from a set of backscattered wave forms, and P_{ref} is given by

$$P_{\text{ref}}(f) = k_0^4 |V_{\text{plane}}(\omega)|^2. \quad (6)$$

Subtracting by $\bar{X}(a_{\text{eff}}, \alpha_0)$ removes the effects of multiplicative constants.

B. Verifying calculation of P_{scat}

When using the SF algorithm, the first requirement is to obtain a good approximation of $E[|V_{\text{refl}}(f)|^2]$ (i.e., P_{scat}) from a set of backscattered wave forms. In our earlier paper (Bigelow et al., 2005) P_{scat} was calculated by averaging the spectra of 25 rf echoes in the log domain, that is,

$$P_{\text{scat}}(f) = \exp\left(\frac{1}{25} \sum_{j=1}^{25} \ln(|V_j(f)|^2)\right). \quad (7)$$

The averaging was performed in the log domain because it was assumed that the transmitted pulse is convolved with the impedance properties of the random medium to generate the reflected rf echo. However, traditionally P_{scat} has been calculated by averaging the wave forms in the normal frequency domain (Insana et al., 1990), that is,

$$P_{\text{scat}}(f) = \frac{1}{25} \sum_{j=1}^{25} |V_j(f)|^2. \quad (8)$$

Therefore, the first signal processing analysis involved determining which method [i.e., Eq. (7) or (8)] yields better performance for the SF algorithm.

The comparison was performed using simulated rf echoes from an infinite half-space placed flush against the aperture plane of a spherically focused f/4 transducer with a focal length of 5 cm. The transducer was excited by an impulse function [i.e., $V_{\text{inc}}(f) = 1$] that had filtering characteristics given by

$$H(f) = \frac{|f| \exp\left(-\left(\frac{f - 8 \text{ MHz}}{6 \text{ MHz}}\right)^2\right)}{\max_{\forall f}\left(|f| \exp\left(-\left(\frac{f - 8 \text{ MHz}}{6 \text{ MHz}}\right)^2\right)\right)}. \quad (9)$$

The half-space had an attenuation of 0.3 dB/cm MHz and contained randomly positioned scatterers with Gaussian impedance distributions (i.e., Gaussian scatterers) with effective radii of 25 μm at a density of 35/mm^3. The veloc-

ity potential field from the source was approximated as a three-dimensional Gaussian distribution in the focal region (Bigelow and O'Brien, 2004) where the equivalent Gaussian dimensions were given by

$$w_x = w_y = 0.87\lambda f\# \,,$$

$$w_z = 6.01\lambda (f\#)^2. \tag{10}$$

In the simulations, 1000 rf echoes were generated corresponding to 1000 independent scatterer distributions. Each echo was sampled at a sampling rate of 53 MHz when "digi-

tizing" the signal. The 1000 echoes were then grouped into sets of 25 wave forms so that 40 independent estimates of scatterer size and total attenuation could be obtained. For each rf echo, the wave form was windowed in the time domain using a Hamming windowing function where the length of the time gate was calculated from $T_{win}=2L/c$. The same length Hamming window (i.e., $T_{win}=2L/c_0$) was also used to window the signal returned from the rigid plate placed at the focus when obtaining the reference spectrum. The ASD from Eq. (4) was minimized over the set of all frequencies corresponding to signal values satisfying

$$\left\{10\log\left(\frac{P_{scat}(f)}{\max_{\forall f}(P_{scat}(f))}\right)\right\}_{\substack{polynomial\\fit}} > \max\left(\left[-30\ \text{mean}\left(\left\{10\log\left(\frac{P_{scat}(f)}{\max_{\forall f}(P_{scat}(f))}\right)\right\}_{\substack{polynomial\\fit}}\right)\right]\right), \tag{11}$$

where the spectrum was fit by a polynomial of large degree (i.e., 50) to reduce the impact of spectral variations on the selected frequency range.

The accuracy (error in mean value) and precision (estimate deviation) of the size and total attenuation estimates for calculating P_{scat} according to Eqs. (7) and (8) are shown in Fig. 1. In these plots, the errors in the mean value (i.e., plots I) were the errors between the mean value of all 40 estimates and the true value. The deviations (i.e., plots II) were found by adding the standard deviation values of estimates above and below the mean value. The standard deviations were given by

$$\sigma_{a_{upper}} = \frac{100}{a_{eff}|_{Theory}} \sqrt{\frac{\sum_{\forall a_{eff\,j}>\bar{a}_{eff}}(a_{eff\,j}-\bar{a}_{eff})^2}{\sum_{\forall a_{eff\,j}>\bar{a}_{eff}}j}},$$

$$\sigma_{a_{lower}} = \frac{100}{a_{eff}|_{Theory}} \sqrt{\frac{\sum_{\forall a_{eff\,j}<\bar{a}_{eff}}(a_{eff\,j}-\bar{a}_{eff})^2}{\sum_{\forall a_{eff\,j}<\bar{a}_{eff}}j}} \tag{12}$$

for the deviation in a_{eff} and

$$\sigma_{\alpha_{upper}} = \sqrt{\frac{\sum_{\forall(\alpha_0 z_T)_j>\overline{(\alpha_0 z_T)}}((\alpha_0 z_T)_j-\overline{(\alpha_0 z_T)})^2}{\sum_{\forall(\alpha_0 z_T)_j>\overline{(\alpha_0 z_T)}}j}},$$

$$\sigma_{\alpha_{lower}} = \sqrt{\frac{\sum_{\forall(\alpha_0 z_T)_j<\overline{(\alpha_0 z_T)}}((\alpha_0 z_T)_j-\overline{(\alpha_0 z_T)})^2}{\sum_{\forall(\alpha_0 z_T)_j<\overline{(\alpha_0 z_T)}}j}} \tag{13}$$

for the deviation in total attenuation. The deviations were given by $(\sigma_{\alpha_{upper}}+\sigma_{\alpha_{lower}})$ and $(\sigma_{\alpha_{upper}}+\sigma_{\alpha_{lower}})$. The mean values of the estimates using Eqs. (7) and (8) are almost identical, but the deviations in the estimates are slightly larger when averaging in the log-domain [i.e., Eq. (7)]. Hence, the precision of the SF algorithm is improved when the spectra from the rf echoes are averaged in the normal frequency domain [i.e., Eq. (8)] rather than in the log domain [i.e., Eq. (7)].

C. Windowing function compensation

In Fig. 1, the accuracy of the estimates from the SF algorithm degrades quickly for window lengths less than 4 mm due to the convolution effects of the windowing function (i.e., windowing in the time domain translates to convolution in the frequency domain) (Akita and Ueda, 1988). This is restrictive because the convolution effects limit the resolution of the QUS image. Hence, the next signal process-
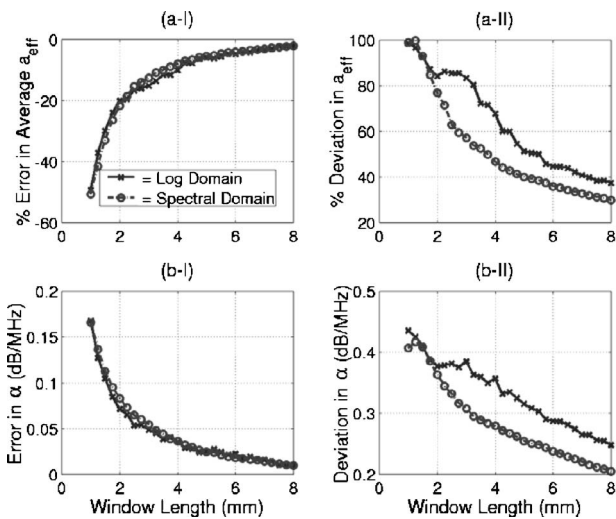
FIG. 1. A comparison between averaging in the normal spectral domain and the log spectral domain to reduce spectral noise due to random scatterer spacing for a half-space with an attenuation of 0.3 dB/cm MHz in the absence of electronic noise: (a) compares the scatterer size estimates and (b) compares the total attenuation estimates. (I) The errors in the mean value for the estimates and (II) the total deviation about the mean estimated values.

ing strategy endeavored to compensate for the effects of windowing so that smaller windowing functions could be used. One possibility would be to use a reference phantom instead of a rigid plane at the focus to obtain the reference spectrum (Gerig *et al.*, 2003) and then use the same Hamming window on both. However, a reference phantom increases the uncertainty of the estimation technique (Gerig *et al.*, 2003). Furthermore, performing a comprehensive evaluation of the reference phantom technique is difficult because of the need to test materials with attenuation, scatterer type, and scatterer sizes differing from the reference phantom. Hence, the reference phantom technique was not evaluated at this time. Instead, an attempt was made to compensate directly for the windowing effects on the backscattered wave forms while still using a plane placed at the focus as a reference.

It has been shown that windowing of the signals in the time domain is equivalent to a convolution of the power spectra given by (Wear, 2002)

$$E[|V_{refl}|^2]_{windowed} \propto E[|V_{refl}|^2]_{ideal} * |G_{win}(f)|^2, \quad (14)$$

where $G_{win}(f)$ is the Fourier transform of the windowing function. Also, if the windowing function and expected backscattered spectra can be reasonably described by Gaussian distributions given by

$$E[|V_{refl}|^2]_{ideal} \propto e^{-(f-f_0)^2/2\sigma_\omega^2},$$
$$|G_{win}(f)|^2 \propto e^{-f^2/2\sigma_g^2}, \quad (15)$$

then the windowed backscattered spectrum is approximately given by (Wear, 2002)

$$E[|V_{refl}|^2]_{windowed} \propto e^{-(f-f_0)^2/2(\sigma_\omega^2+\sigma_g^2)}. \quad (16)$$

Windowing increases the bandwidth of the backscattered wave form. In Bigelow *et al.* (2005), the scatterer size was related qualitatively to a decrease in the backscattered bandwidth. Hence, the increase in the bandwidth due to windowing is responsible for the underestimate of the scatterer size as was observed in Figs. 1 and 2. Furthermore, the bandwidth broadening, as described in this manner, can be compensated by multiplying the windowed spectrum by

$$G_{corr} = e^{-(f-f_0)^2\sigma_g^2/2(\sigma_\omega^2+\sigma_g^2)\sigma_\omega^2}. \quad (17)$$

This correction for windowing was also tested using the rf echoes from the half-space described previously. The backscattered spectra from 25 independent rf echoes windowed with Hamming windows without additive electronic noise were averaged in the normal frequency domain [i.e., Eq. (8)] yielding an estimate $E[|V_{refl}|^2]_{windowed}$. The resulting average spectrum was then fit by a Gaussian distribution in the log domain, that is,

$$[f_0, (\sigma_\omega^2 + \sigma_g^2)]$$
$$= \min_{\substack{\forall f_0 \\ \forall (\sigma_\omega^2+\sigma_g^2)}} \{mean((P_n(f) - P_p(f, f_0, (\sigma_\omega^2 + \sigma_g^2)) - \bar{P}_n)^2)\},$$

$$(18)$$

where
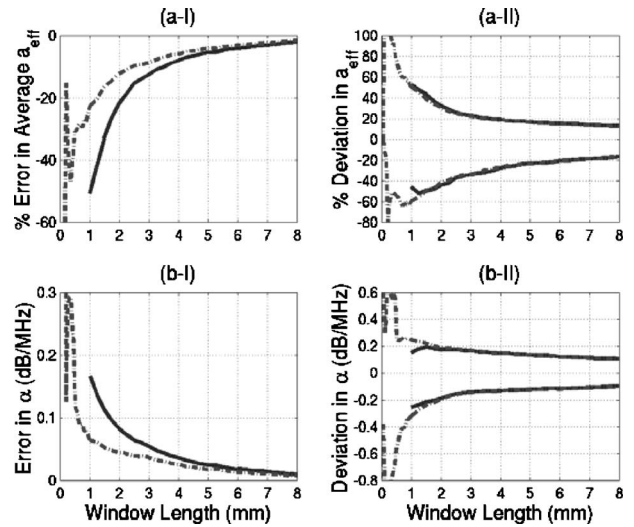
FIG. 2. A comparison between (I) the mean values for the estimates (accuracy) and (II) the deviations about the mean values (precision) for (a) scatterer size and (b) total attenuation when using the same Hamming window to window the reference and backscattered wave forms, (—), and using G_{corr} to correct for the effects of windowing, (---). In the simulations, the half-space had an attenuation of 0.3 dB/cm MHz and no electronic noise was added to the backscattered wave forms.

$$P_n(f) = \ln\left(\frac{P_{scat}(f)}{\max_{\forall f}(P_{scat}(f))}\right),$$

$$P_p(f, f_0, (\sigma_\omega^2 + \sigma_g^2)) = -\frac{(f-f_0)^2}{2(\sigma_\omega^2 + \sigma_g^2)}, \quad (19)$$

$$\bar{P}_n = mean(P_n(f) - P_p(f, f_0, (\sigma_\omega^2 + \sigma_g^2))),$$

to obtain values for f_0 and $(\sigma_\omega^2 + \sigma_g^2)$. The Fourier transform of the Hamming window was also calculated and fit by a Gaussian distribution in the normal frequency domain to obtain the value of σ_g^2. G_{corr} was then calculated and applied from these parameters. Because the effects of windowing on the backscattered spectrum were directly compensated, no windowing was applied to the reference spectrum from the rigid plane. The results of this simulation are shown in Fig. 2 with the results using a Hamming window for both the reference and backscattered signals from Fig. 1. The correction term allows for more accurate estimates of total attenuation and scatterer size (i.e., better than 20%) down to window lengths of 1 mm. However, the precision of the estimates continues to degrade for the smaller window lengths. The loss of precision at smaller window lengths probably results from a reduction in the number of independent samples in the least-squares fit in Eq. (4) due to a loss in the frequency resolution of the spectrum (Wear, 2001).

In this section, a new windowing compensation method was developed and applied to the SF algorithm. Other researchers have developed methods for windowing function compensation for the traditional estimation algorithm where the attenuation was known *a priori* (Akita and Ueda, 1988). Because their work assumed that the total attenuation was known, the performance of the different windowing compensation methods cannot be compared in this study.

T. A. Bigelow and W. D. O'Brien, Jr.: Signal Processing SPECTRAL FIT

D. Improving selection of frequency range

In the earlier simulations, the frequency range used by the SF algorithm was dictated by Eq. (11). However, this definition of the frequency range is not robust and depends significantly on the sampling rate used to acquire the data. Therefore, before the SF algorithm can be reliably implemented, a better definition of the usable frequency range needs to be obtained. Because the principal purpose in selecting the frequency range is to exclude spectral regions dominated by system noise, a suitable alternative would be to select only those frequencies whose signal levels were reasonably larger than the noise floor. Hence, a more generalized selection criteria is given by the set of all frequencies corresponding to signal levels greater than

$$
N_{\text{Floor}} = \max\left(\left[\, N_{\text{dB}} \;\text{mean}\left(10\log\left(\frac{P_{\text{scat}}(f_{N-200}:f_N)}{\max_{\forall f}(\{P_{\text{scat}}(f)\}_{\text{Gaussian}\atop\text{fit}})}\right)\right) + 6\,\right]\right),
\tag{20}
$$

where N_{dB} sets a limit on the frequency range in the absence of any electronic noise and f_N refers to the largest positive frequency after taking the Discrete Fourier Transform (DFT) of the sampled wave forms. Also, the signal must be sufficiently oversampled so that at least the last 200 frequency samples would correspond to the additive white electronic system noise. The selection of the last 200 samples, rather than another large number of frequency samples (i.e., last 100 samples), is arbitrary. Six decibels were added to the average of the last 200 samples to ensure that N_{Floor} was always greater than the noise. Smaller values (i.e., 3 dB) were attempted but some of the fluctuations of the noise about its mean value influenced the estimates. A value other than 6 dB may be used on a case-by-case basis by analyzing the deviation of the noise in the last 200 frequency samples, but this possibility is beyond the scope of our current investigation. Furthermore, estimating the noise floor from the last 200 samples is only valid when the noise is white. When the noise is not white, then the noise floor over the relevant frequency range should be estimated from a sample noise signal obtained in the absence of a transmitted signal as was discussed in Bigelow et al. (2005).

E. Noise compensation revisited

In another paper (Bigelow et al., 2005), a method to compensate for the additive electronic system noise was introduced. The addition of electronic noise artificially broadened and also reduced (i.e., cannot use information at frequencies below the noise floor) the usable spectrum. Previously, the compensation involved dividing the received spectrum by the term $(1+E_N[|N(f)|^2]/P_{\text{scat}}(f))$ where $E_N[|N(f)|^2]$ was estimated from a reference noise signal in the absence of a transmitted signal in order to eliminate the artificial broadening. However, this compensation method was originally derived when the spectra of the rf echoes were averaged in the log domain rather than the normal frequency domain (i.e., see Sec. II B). Therefore, we need to assure that the same noise compensation is still valid when averaging the spectra in the normal frequency domain. The value of P_{scat} in the presence of electronic noise when averaging in the normal frequency domain is given by

$$
P_{\text{scat}}(f) = \frac{1}{25}\sum_{i=1}^{25}\left(|V_i(f)|^2\left(1 + \frac{E_N[|N(f)|^2]}{|V_i(f)|^2}\right)\right)
$$

$$
= P_{\text{scat}}(f)_{\text{ideal}}\left(1 + \frac{E_N[|N(f)|^2]}{P_{\text{scat}}(f)_{\text{ideal}}}\right).
\tag{21}
$$

The artificial broadening is still given by $(1+E_N[|N(f)|^2]/P_{\text{scat}}(f)_{\text{ideal}})$, and dividing the measured P_{scat} by $(1+E_N[|N(f)|^2]/P_{\text{scat}}(f))$ should still reduce the impact of the electronic noise by eliminating the artificial broadening.

Also, $E_N[|N(f)|^2]$ used in compensating for the electronic noise might be obtainable from $P_{\text{scat}}(f_{N-200}:f_N)$, if the spectrum is sufficiently oversampled and the noise is white, removing the need to record the noise in the absence of a transmitted signal. To test this possibility, computer simulations were performed using the same f/4 transducer described when evaluating the calculation of P_{scat} previously in this paper. The half-space for these simulations had an attenuation of 0.05 dB/cm MHz, similar to the half-space used to evaluate the electronic noise compensation by Bigelow et al. (2005), and contained randomly positioned Gaussian scatterers with effective radii of 25 μm at a density of 35/mm³. The rf echoes from 1000 independent scatterer distributions were generated and grouped into 40 sets of 25 wave forms per set for processing. After generating the rf echoes, three different levels of electronic noise were added to the wave forms corresponding to average SNRs of 9, 23, and 36 dB. The SNR values were calculated from

$$
\text{SNR} = \frac{1}{25}\sum_{j=1}^{25}\left(10\cdot\log\left(\frac{\int(g_{\text{win}}(t)v_{\text{refl}_j}(t))^2 dt}{\int(g_{\text{win}}(t)v_{\text{noise}}(t))^2 dt}\right)\right),
\tag{22}
$$

where v_{refl_j} are the rf echoes from each group of 25 used in the estimate before the noise was added, and v_{noise} is the noise signal used to obtain the estimate for $E_N[|N(f)|^2]$ as described by Bigelow et al. (2005). The noise added to each wave form was randomly generated for that wave
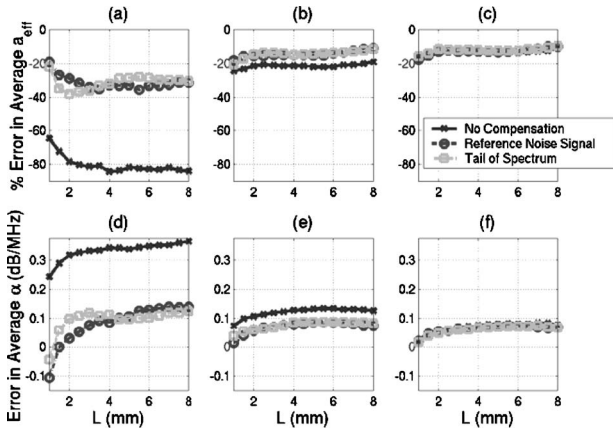
FIG. 3. Comparison of errors in scatterer size estimates for SNRs of (a) 9 dB, (b) 23 dB, and (c) 36 dB as well as attenuation estimates for SNRs of (d) 9 dB, (e) 23 dB, and (f) 36 dB without noise compensation and with different methods for estimating $E_N[|N(f)|^2]$ for a half-space with an attenuation of 0.05 dB/cm MHz.

form while the noise power was set relative to the reference signal from the rigid plane placed at the focal plane. As a result, the SNR for each wave form was slightly different. Hence, Eq. (22) gives the average SNR for the 25 wave forms in a set. The generated noise was zero-mean white with a Gaussian probability density function. The wave forms were analyzed and P_{scat} was compensated for noise using $E_N[|N(f)|^2]$ obtained from both a reference noise signal as well as by averaging the last 200 values of P_{scat}. The convolution effects of windowing were compensated and the frequency range used in the minimization was given by Eq. (20) with N_{dB} set to −20 dB.

The results for both methods of obtaining $E_N[|N(f)|^2]$ are shown in Fig. 3 along with the results obtained without electronic noise compensation. Both methods for obtaining an estimate for $E_N[|N(f)|^2]$ perform about the same and give significant improvement in the accuracy compared to when the electronic noise is not compensated. Due to the comparable performance, in the future $E_N[|N(f)|^2]$ will be found from the last 200 values of P_{scat}, removing the need to record the noise in the absence of a transmitted signal when the noise is white.

III. METHODS THAT PROVIDED INSIGHT

In addition to the signal processing strategies that improved the performance of the SF algorithm, different approaches were attempted that provided no improvement (Bigelow, 2004). Two of the failed approaches are included herein because although they did not work, they provided useful information.

A. Homomorphic deconvolution

One of the signal processing strategies investigated used homomorphic deconvolution (Proakis and Manolakis, 1996) to smooth the backscattered spectrum and hopefully to compensate for the effect of random scatterer spacing thereby improving the precision of the estimates. Each of the back-

scattered wave forms was gated using a Hamming window $g_{win}(t)$, converted to the frequency domain, and then filtered according to

$$|V_i|^2 = |\exp(\text{Re}(\text{IFFT}\{V_{cepi}(\tau_{cep}) \cdot g_{win2}(\tau_{cep})\}))|^2, \quad (23)$$

where

$$V_{cepi}(\tau_{cep}) = \text{FFT}\{\ln(|\text{IFFT}\{v_{refli}(t) \cdot g_{win}(t)\}|)\} \quad (24)$$

and

$$g_{win2} = \begin{cases} 1 & |\tau_{cep}| \leq T_{cep} \\ 0 & |\tau_{cep}| > T_{cep}. \end{cases} \quad (25)$$

The filtered wave forms were then averaged in the normal spectral domain according to Eq. (8), and the effect of convolution due to the Hamming window was compensated.

To test the SF algorithm, simulations were performed using the spherically focused f/4 transducer placed adjacent to an infinite half-space as described previously. This time, 2000 random scatterer distributions were generated. The wave forms from each distribution were then initially grouped in sets of 40 to yield 50 independent estimates of scatterer size and total attenuation for each Hamming window length (varied from 1 to 8 mm). The wave forms were initially taken in sets of 40 instead of 25 to attempt to further improve the precision. (The relationship between varying the number of wave forms in each set and precision will be discussed later in this section.) In addition, a sampling rate of 160 MHz was used when "digitizing" the signal so that the features of the wave forms could be observed better in the cepstral domain. The attenuation of the half-space region was 0.3 dB/cm MHz. The range of frequencies used by the minimization was selected based on Eq. (20) with N_{dB} set to −20 dB.

In the simulations, the value of T_{cep} used to set the amount of filtering was varied as 0.15, 0.31, 0.46, and 0.62 μs. The locations of all of these cutoffs relative to the main spectral peak are shown with an example signal plotted in the cepstral domain in Fig. 4 along with the original unfiltered spectrum. After applying the filter in the cepstral domain, the example spectra were significantly smoothed (Fig. 5). The smaller the value of T_{cep}, the greater the filtering, and the smoother the resulting spectra. The simulation results for the four different filter lengths are shown in Fig. 6. For a filter length of 0.15 μs, the accuracy of the scatterer size estimates was seriously degraded, especially for small window lengths. Hence, the 0.15 μs filter length will not be considered in our detailed analysis of the accuracy and precision given in Fig. 7. In addition to simulating each of the four filtered cases, the simulation was also run without filtering (i.e., $T_{cep}=\infty$). Hence, the performance of the filtered cases could be compared with the performance available from the unfiltered SF algorithm with 40 wave forms in each set averaged in the normal spectral domain while compensating for the convolution effects of windowing.

In Fig. 7, the accuracy and precision for filter lengths of 0.31, 0.46, and 0.62 μs are compared to the results obtained with no filtering. The accuracy comparison is done by dividing the mean value of the percent error in the scatterer size estimate with filtering by the mean value of the percent error
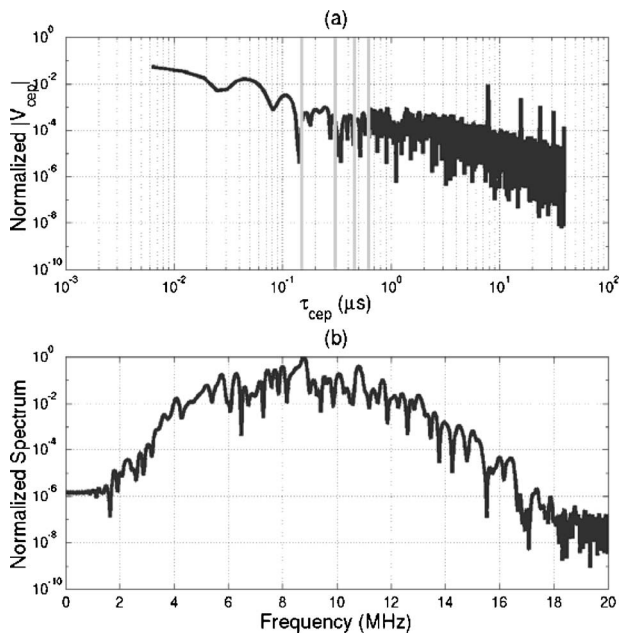
FIG. 4. (a) Example of a spectrum for a single rf echo in cepstral domain showing locations of T_{cep} cutoffs (vertical gray lines) used in the homomorphic deconvolution with (b) the corresponding normalized spectra before any homomorphic processing has been applied.

FIG. 6. Percent error in scatterer size estimate for different window lengths after homomorphic deconvolution with (a) $T_{cep}=0.15$ μs, (b) $T_{cep}=0.31$ μs, (c) $T_{cep}=0.46$ μs, and (d) $T_{cep}=0.62$ μs.

in the scatterer size estimate without filtering [Fig. 7(a)]. Likewise, the precision comparison is done by dividing the upper [Fig. 7(b)] and lower [Fig. 7(c)] σ values as given by Eq. (12) with filtering by the corresponding values without filtering. An improvement in accuracy or precision would be indicated by a ratio less than one, and degradation in accuracy or precision would be indicated by a ratio greater than one.

From Figs. 6 and 7, the accuracy of the scatterer size estimates is degraded by small filter lengths (0.15 and 0.31 μs). The loss in accuracy at smaller filter lengths is probably due to relevant spectral information being ignored (smoothed out) as a result of the filtering. This is similar to

the poor accuracy that was observed when spectral information was ignored by fitting the spectrum to a Gaussian distribution (Bigelow et al., 2005). At larger filter lengths, the accuracy of the scatterer size estimates is slightly improved, especially at larger window lengths [Fig. 7(a)]. In terms of

FIG. 5. Example of a normalized spectrum for a single rf echo after applying homomorphic deconvolution with (a) $T_{cep}=0.15$ μs, (b) $T_{cep}=0.31$ μs, (c) $T_{cep}=0.46$ μs, and (d) $T_{cep}=0.62$ μs.
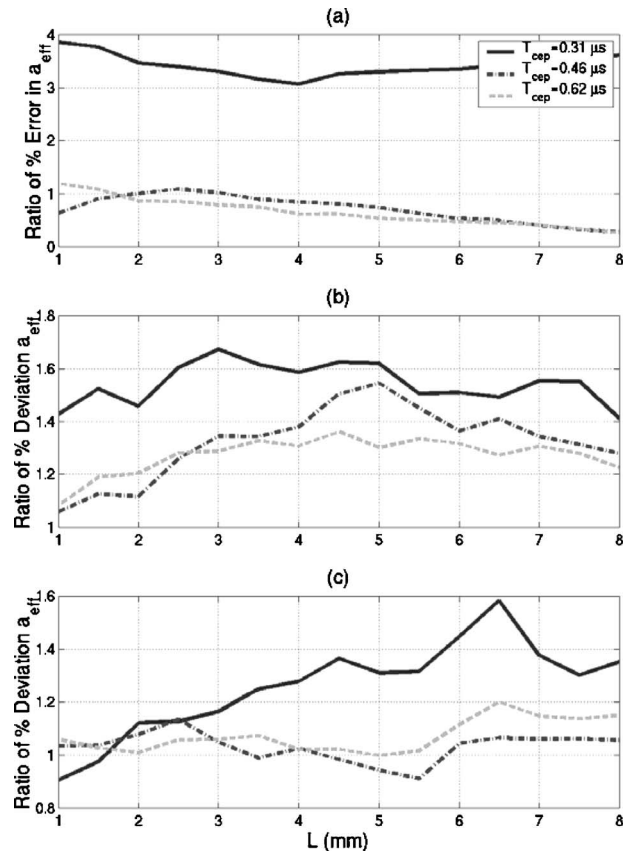
FIG. 7. Simulation results comparing the performance of different homomorphic filter lengths to the performance of the basic SPECTRAL FIT algorithm by plotting (a) the ratio of the average percentage error of the filtered and unfiltered results, (b) the ratio of $\sigma_{a_{upper}}$ of the filtered and unfiltered results, and (c) the ratio of $\sigma_{a_{lower}}$ of the filtered and unfiltered results.
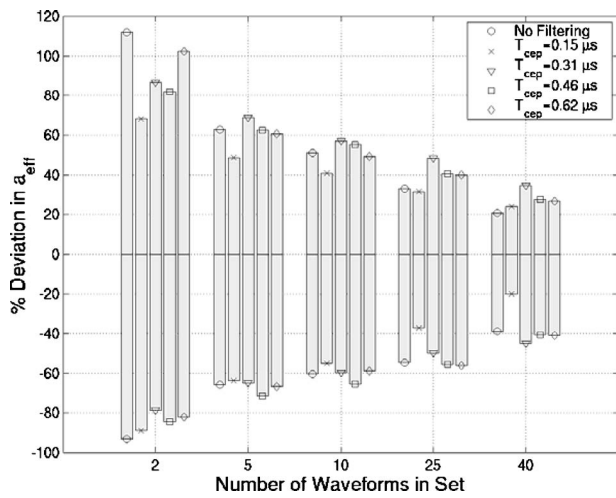
FIG. 8. $\sigma_{a_{upper}}$ and $\sigma_{a_{lower}}$ values when a different number of wave forms are averaged in each set before performing the estimate for the different homomorphic deconvolution parameters. The first bar in each group corresponds to the case where no homomorphic filtering was applied to the wave forms while the second, third, fourth, and fifth bars correspond to results for filter lengths of 0.15, 0.31, 0.46, and 0.62 μs, respectively.

FIG. 9. A comparison between the basic SPECTRAL FIT algorithm ($n=2$) and varying the assumed form factor ($n=2-3.6$) and then extrapolating to find the final estimate for (a) the percent error in the average scatterer size, (b) the percent deviation in the scatterer size, (c) the error in the average total attenuation, and (d) the deviation in the total attenuation for different window lengths.

precision, however, the filtering consistently degrades the performance. The ratio in Fig. 7(b) is always greater than one while the ratio in Fig. 7(c) is greater than or approximately equal to one. Because it is the precision of the estimates that serves as the limiting factor of the SF algorithm, and not the accuracy, using homomorphic deconvolution to smooth each spectrum before averaging provides no benefit to the algorithm when there are 40 wave forms per set.

Although filtering does not provide any benefit when there are 40 wave forms per set, some improvement may occur when there are fewer wave forms available for averaging in the spectral domain. Hence, the simulations were re-evaluated for a Hamming window length of 3 mm with the wave forms grouped into independent sets of 2, 5, 10, and 25 wave forms. For the sake of comparison, a total of 50 estimates were made for each of the different wave form groupings. Hence, not all 2000 of the previously generated wave forms were needed. Figure 8 shows the precision results for each grouping of the wave forms where the upper and lower σ values from Eq. (12) are plotted above and below the axis, respectively. The first bar in each group corresponds to the case where no homomorphic filtering was applied to the wave forms while the second, third, fourth, and fifth bars correspond to results for filter lengths of 0.15, 0.31, 0.46, and 0.62 μs, respectively.

A systematic improvement in precision occurs as the number of independent wave forms used to obtain the estimate is increased (Fig. 8). However, for a given number of wave forms in a set, there is no improvement in precision provided by homomorphic deconvolution. The only possible exception occurs when there are only two wave forms per set, but the overall precision for this case is so poor that the amount of improvement is insignificant. Hence, homomorphic deconvolution does not allow for fewer wave forms to be used in the estimation scheme for the same level of precision.

From this investigation, homomorphic deconvolution

clearly does not provide a benefit to the SF algorithm. The poor precision of the SF algorithm is probably related to the slowly varying spectral fluctuations (i.e., small τ_{cep} values) rather than the fast spectral fluctuations that were effectively removed by processing the signals in the cepstral domain. Hence, future algorithms should not be concerned about smoothing the spectra, but rather the emphasis should be placed on reducing the impact of gradual changes to the spectra resulting from the random scatterer spacing. Unfortunately, removing the smooth perturbations is a much more challenging problem.

B. Varying form factor

Another signal processing strategy investigated to improve the precision of the SF algorithm that did not work involved varying the assumed frequency dependence of the form factor describing the scatterer geometry. The correct form factor when using scatterers with a Gaussian impedance distribution is given by

$$F_\gamma(f, a_{eff}) = \exp(-0.827 \cdot (ka_{eff})^n), \tag{26}$$

where $n=2$ (Insana *et al.*, 1990). Different estimates for scatterer size and total attenuation can be found by varying the value of n and solving the minimization routine of the SF algorithm for each value of n. A final estimate is then obtained by fitting a line to the estimates versus n values and extrapolating the scatterer size for an n value of 2. In the fit, the n values were weighted by 1 over the minimum value of ASD from Eq. (4). Hence, the estimates corresponding to the better fits from the minimization routine would have a larger impact on the final estimate.

Some simulation results using varying form factors are presented in Fig. 9 along with the results for the basic SF algorithm. These results were obtained using a half-space with an attenuation of 0.3 dB/cm MHz containing Gaussian scatterers with effective radii of 25 μm while varying the
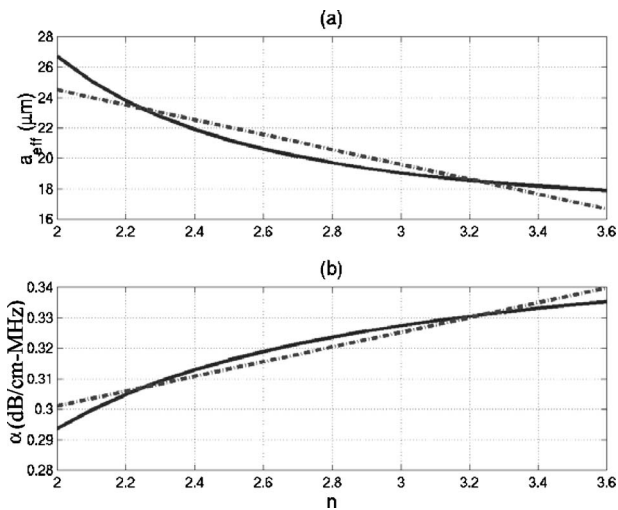
FIG. 10. Example scatterer size and total attenuation estimates at each value of the assumed form factor (solid line) along with the corresponding extrapolation line (dashed line).

assumed value of n from 2 to 3.6 in steps of 0.1. Smaller values of n (i.e., less than 2) were also considered, but the results were not qualitatively different from the ones shown. Once again, 1000 wave forms were generated corresponding to 1000 independent scatterer distributions. The wave forms were grouped into sets of 25 (yielding 40 final estimates), windowed with Hamming windows, and averaged in the normal spectral domain. The convolution impact of windowing was compensated. The percent deviation was also found by adding the upper and lower σ values from Eqs. (12) and (13). Varying the frequency dependence of the form factor from 2 to 3.6 provides an insignificant amount of improvement to the precision of the estimates [Figs. 9(b) and 9(d)] while degrading the accuracy of the estimates [Figs. 9(a) and 9(c)] by about 8%.

The lack of improvement can be understood by considering the results for each assumed form factor along with the linear fit for an example spectrum shown in Fig. 10. A Hamming window with a length of 3 mm was used to obtain these results. The curve is smooth. Hence, variations in the assumed frequency dependence of the form factor do not produce statistically independent estimates and as a result did not improve the precision of the estimation scheme.

The curves shown in Fig. 10 also demonstrate that if the correct form factor (i.e., scatterer geometry) was not known, then reasonable estimates for the total attenuation would still be obtainable. For example, if a frequency dependence of 2.25 is selected (similar to the frequency dependence of a spherical shell type scatterer over some frequency bandwidths), then the estimated attenuation value would differ by only about 0.01 dB/cm MHz from the value given for the true frequency dependence of 2. Unfortunately, the scatterer size will still vary with the form factor because the constant term (i.e., 0.827 for Gaussian scatterers) will vary depending upon the scatterer geometry. However, applications that only require the *in vivo* incident power spectrum, obtainable from the total attenuation, and not the scatterer size, would still benefit from the estimation scheme based on these results even when the correct form factor is not known. This con-

clusion is important because the correct form factor for biological tissue remains to be determined.

In order to further validate these results, an experiment was performed using a glass-bead phantom (i.e., Gaussian form factor does not apply). The radius of the glass beads in the phantom varied from 22.5 to 26.5 μm and the beads were placed at a density of approximately 48/mm³ (Madsen *et al.*, 1999). The phantom had an attenuation of 0.826 ± 0.1 dB/cm MHz as measured in our lab monitoring the decay in the backscatter with distance as well as insertion loss measurements. A single-element f/4 spherically focused transducer (Valpey Fisher Instruments, Inc., Hopkinton, MA) with a focal length of 5.54 cm was used to obtain the backscattered signals. The transducer was shock excited by a Panametrics 5900 (Waltham, MA) pulser/receiver operating in pulse/echo mode. $H(f)$ for the transducer followed a Rayleigh function similar to the form given in Eq. (9) and was given by

$$|H(f)| = \frac{|f|\exp\left(-\left(\dfrac{f-6.86\ \mathrm{MHz}}{9.04\ \mathrm{MHz}}\right)^2\right)}{\max_{\forall f}\left(|f|\exp\left(-\left(\dfrac{f-6.86\ \mathrm{MHz}}{9.04\ \mathrm{MHz}}\right)^2\right)\right)}. \quad (27)$$

In the experiment, the focus of the transducer was positioned 15 mm into the phantom and 875 independent wave forms were acquired using a microprecision positioning system (Daedal Inc., Harrison City, PA) where each wave form was separated by ~1 beamwidth. Each wave form was averaged 1000 times to remove noise by a digital oscilloscope (LeCroy 9354 TM, Chestnut Ridge, NY) using a sampling frequency of 100 MHz before being transferred to a computer for later processing. As a result, the noise floor for the experiment was around −30 dB. The averaged independent wave forms were then windowed by a Hamming window with a length of 3 mm [corrected for windowing using Eq. (17)], grouped into sets of 25, and averaged in the normal frequency domain to obtain 35 measurements of $P_{scat}(f)$. The values of P_{scat} were corrected for the mylar sheet boundary of the phantom so that the attenuation results could be directly compared to earlier experimental results of the phantom's attenuation. The −20 dB bandwidth of the P_{scat} signals was then used in the SF algorithm along with a P_{ref} obtained from the reflection returned by a smooth Plexiglas™ plane placed at the focal plane to obtain 35 estimates of scatterer size and total attenuation.

In the earlier implementations of the SF algorithm involving simulated Gaussian scatterers, a simple minimization routine could be utilized because the ASD surface given by Eq. (4) only had one clearly defined minimum. In the phantom experiment, the ASD surface could have multiple minima with one minimum being significantly lower in value than the others. In this work, the problem of multiple minima was avoided by mapping out the entire ASD surface over a range of possible scatterer sizes and attenuation values, and then, finding the scatterer size and attenuation value that gave the global minimum. The ASD surface was determined for scatterer sizes from 0 μm to two times the scatterer size
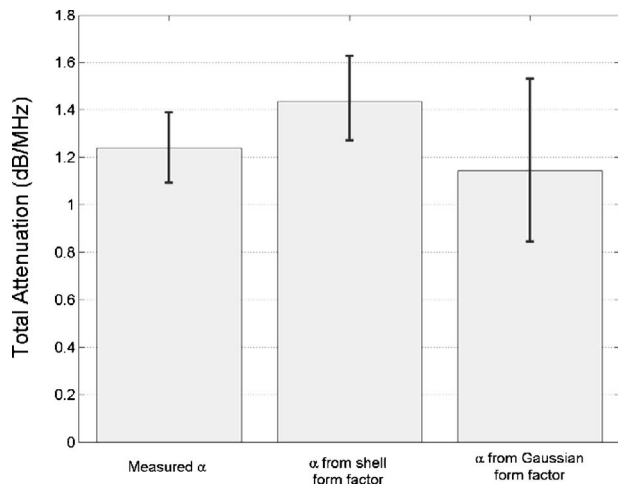
FIG. 11. Comparison of the measured total attenuation for the phantom to the total attenuation found using the SPECTRAL FIT algorithm assuming a spherical shell form factor and a Gaussian form factor.

found when using the traditional size-estimation scheme (Insana *et al.*, 1990) where the attenuation was known (i.e., measured phantom attenuation). The range of attenuation values for the ASD surface was then determined by finding the attenuation value that minimized the ASD when the scatterer size was set at 0 μm and two times the scatterer size from the traditional estimation scheme, respectively.

In order to show the independence of the attenuation estimate on the exact form factor, the analysis using the SF algorithm was performed twice. First, the glass-bead scatterers were described using the correct spherical shell form factor [i.e., $F_\gamma(f, a_{eff}) = (\text{sinc}(2ka_{eff}/\pi))^2$] (Insana *et al.*, 1990). Using the correct form factor, the mean value for the scatterer radius was 22.5 μm, $\sigma_{a_{upper}}$ was 2.2 μm, and $\sigma_{a_{lower}}$ was 3.7 μm which is in the range of the actual sizes of the glass beads. Then, the analysis was repeated using a Gaussian form factor [i.e., $F_\gamma(f, a_{eff}) = \exp(-0.827 \cdot (ka_{eff})^2)$]. Using the Gaussian form factor, the mean value for the scatterer effective radius was 40.0 μm, $\sigma_{a_{upper}}$ was 7.3 μm, and $\sigma_{a_{lower}}$ was 10.4 μm. The total attenuation results for both form factors are shown with the measured attenuation results for the phantom in Fig. 11. The mean values for the total attenuation using both form factors are within 0.2 dB/MHz of the average measured value. Hence, knowing the exact form factor is not critical when estimating attenuation along the propagation path however, it is critical when estimating the scatterer size.

IV. CONCLUSION

Six signal processing strategies were evaluated. The best precision was obtained when averaging rf spectra in the normal frequency domain rather than in the log domain. The convolution effects of windowing were reduced by approximating the windowing function as a Gaussian transformation on an ideal backscattered Gaussian spectrum. Although the use of Gaussian approximations when solving for the scatterer size and total attenuation had previously been shown to be invalid (Bigelow *et al.*, 2005), a Gaussian approximation of the backscattered spectrum was sufficiently valid to correct for the effects of windowing. After investigating the windowing correction, an improved method for selecting the frequency range used in the minimization for the SF algorithm was discussed. Also, the correction for electronic noise introduced in Bigelow *et al.* (2005) was improved by oversampling the backscattered rf wave form and using the information at higher frequencies to estimate the noise level, thus removing the need to acquire a reference noise signal in the absence of a transmitted signal.

Two approaches were also presented that improved our understanding of the problem. First, homomorphic deconvolution removed the rapid spectral fluctuations from the averaged backscattered power spectrum. The homomorphic-processed signals had no improvement over the basic SF algorithm. The rapid fluctuations on the spectrum are probably not important when solving for the scatterer size and total attenuation using the SF algorithm. Instead, the gradual frequency dependencies resulting from the random scatterer spacing probably corrupt the estimates from the SF algorithm leading to the poor precision.

The second signal processing strategy that failed to improve the precision of the SF algorithm involved varying the assumed form factor and then extrapolating to obtain size and attenuation estimates at the correct value of the form factor. The extrapolated values had poorer accuracy and the same precision as directly estimating the size and attenuation using the correct form factor. It was observed that the estimate for total attenuation had only a weak dependence on the form factor. This conclusion was further validated using a phantom experiment. As a result, the SF algorithm should yield accurate estimates of total attenuation when the correct form factor is not known.

Akita, M., and Ueda, M. (**1988**). "The effects of windowing on spectral estimation of echoes scattered by a random medium," J. Acoust. Soc. Am. **107**, 1243–1248.

Bigelow, T. A. (**2004**). "Estimating the medical ultrasound *in vivo* power spectrum," Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Theoretical approximations and simulation analysis," J. Acoust. Soc. Am. **116**, 578–593.

Bigelow, T. A., Oelze, M. L., and O'Brien, W. D., Jr. (**2005**). "Estimation of total attenuation and scatterer size from backscattered ultrasound waveforms," J. Acoust. Soc. Am. **117**, 1431–1439.

Gerig, A., Zagzebski, J., and Varghese, T. (**2003**). "Statistics of ultrasonic scatterer size estimation with a reference phantom," J. Acoust. Soc. Am. **113**, 3430–3437.

Insana, M. F., Wagner, R. F., Brown, D. G., and Hall, T. J. (**1990**). "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192.

Lee, D. J., Sigel, B., Swami, V. K., Justin, J. R., Gahtan, V., O'Brien, S. P., Dwyer-Joyce, L., Feleppa, E. J., Roberts, A. B., and Berkowitz, H. D. (**1998**). "Determination of carotid plaque risk by ultrasonic tissue characterization," Ultrasound Med. Biol. **24**, 1291–1299.

Lizzi, F. L., Astor, M., Liu, T., Cheri, D., Coleman, D. J., and Silverman, R. H. (**1997a**). "Ultrasonic spectrum analysis for tissue assays and therapy evaluation," Int. J. Imaging Syst. Technol. **8**, 3–10.

Lizzi, F. L., Greenebaum, M., Feleppa, E. J., and Elbaum, M. (**1983**). "The-

oretical framework for spectrum analysis in ultrasonic tissue characterization," J. Acoust. Soc. Am. **73**, 1366–1373.

Lizzi, F. L., Kalisz, A., Astor, M., Coleman, D. J., Silverman, R. H., and Reinstein, D. Z. (**1997b**). "Very-high frequency ultrasonic imaging and spectral assays of the eye," Acoust. Imaging **23**, 107–112.

Lizzi, F. L., Muratore, R., Deng, C. X., Ketterling, J. A., Alam, K. A., Mikaelian, S., and Kalisz, A. (**2003**). "Radiation-force technique to monitor lesions during ultrasonic therapy," Ultrasound Med. Biol. **29**, 1593–1605.

Lizzi, F. L., Ostromogilsky, M., Feleppa, E. J., Rorke, M. C., and Yaremko, M. M. (**1986**). "Relationship of ultrasonic spectral parameters to features of tissue microstructure," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 319–329.

Lu, Z. F., Zagzebski, J. A., and Lee, F. T. (**1999**). "Ultrasound backscatter and attenuation in human liver with diffuse disease," Ultrasound Med. Biol. **25**, 1047–1054.

Madsen, E. L., Dong, F., Frank, G. R., Garra, B. S., Wear, K. A., Wilson, T., Zagzebski, J. A., Miller, H. L., Kirk, Shung K., Wang, S. H., Feleppa, E. J., Liu, T., O'Brien, W. D., Jr., Topp, K. A., Sanghvi, N. T., Zaitsev, A. V., Hall, T. J., Fowlkes, J. B., Kripfgans, O. D., and Miller, J. G. (**1999**). "Interlaboratory comparison of ultrasonic backscatter, attenuation, and speed measurements," J. Acoust. Soc. Am. **18**, 615–631.

Nightingale, K. L., Nightingale, R. W., Palmer, M. L., and Trahey, G. E. (**2000**). "A finite element model of remote palpation of breast lesions using radiation force: Factors affecting tissue displacement," Ultrason. Imaging **22**, 35–54.

Noritomi, T., Sigel, B., Swami, V., Justin, J., Gahtan, V., Chen, X., Feleppa, E. J., Roberts, A. B., and Shirouzu, K. (**1997**). "Carotid plaque typing by multipile-parameter ultrasonic tissue characterization," Ultrasound Med. Biol. **23**, 643–650.

O'Donnell, M., Skovoroda, A. R., Shapo, B. M., and Emelianov, S. Y. (**1994**). "Internal displacement and strain imaging using ultrasonic speckle tracking," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 314–325.

Oelze, M. L., O'Brien, W. D. Jr., Blue, J. P., and Zachary, J. F. (**2004**). "Differentiation and characterization of rat mammary fibroadenomas and 4T1 mouse carcinomas using quantitative ultrasound imaging," IEEE Trans. Med. Imaging **23**, 764–771.

Ophir, J., Céspedes, I., Ponnekanti, H., Yazdi, Y., and Li, X. (**1991**). "Elastography: A quantitative method for imaging the elasticity of biological tissues," Ultrason. Imaging **13**, 111–134.

Proakis, J. G., and Manolakis, D. G. (**1996**). "*Digital Signal Processing: Principles, Algorithms, and Applications*," 3rd ed. (Prentice Hall, Upper Saddle River, NJ), Chap. 4, pp. 365–367.

Silverman, R. H., Folberg, R., Rondeau, M. J., Boldt, H. C., Lloyd, H. O., Chen, X., Lizzi, F. L., Weingeist, T. A., and Coleman, D. J. (**2003**). "Spectral parameter imaging for detection of prognostically significant histological features in uveal melanoma," Ultrasound Med. Biol. **29**, 951–959.

Skovoroda, A. R., Emelianov, S. Y., Lubinski, M. A., Sarvazyan, A. P., and O'Donnell, M. (**1994**). "Theoretical analysis and verification of ultrasound displacement and strain imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 302–313.

Skovoroda, A. R., Emelianov, S. Y., and O'Donnell, M. (**1995**). "Tissue elasticity reconstruction based on ultrasonic displacement and strain images," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 747–765.

Wear, K. A. (**2001**). "Fundamental precision limitations for measurements of frequency dependence of backscatter: Applications in tissue mimicking phantoms and trabecular bone," J. Acoust. Soc. Am. **110**, 3275–3282.

Wear, K. A. (**2002**). "A Gaussian framework for modeling effects of frequency-dependent attenuation, frequency-dependent scattering, and gating," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1572–1582.

Underwater hearing sensitivity of a male and a female Steller sea lion (*Eumetopias jubatus*)

Ronald A. Kastelein[a]
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands

Robbert van Schie
Bunschotenmeen 11, 3844 HE Harderwijk, The Netherlands

Wim C. Verboom
TNO Observation Systems, Dept. Underwater Technology, P.O. Box 96864, 2509JG The Hague, The Netherlands

Dick de Haan
Netherlands Institute for Fisheries Research (RIVO), P.O. Box 68, 1970 AB IJmuiden, The Netherlands

The unmasked underwater hearing sensitivities of an 8-year-old male and a 7-year-old female Steller sea lion were measured in a pool, by using behavioral psychophysics. The animals were trained with positive reinforcement to respond when they detected an acoustic signal and not to respond when they did not. The signals were narrow-band, frequency-modulated stimuli with a duration of 600 ms and center frequencies ranging from 0.5 to 32 kHz for the male and from 4 to 32 kHz for the female. Detection thresholds at each frequency were measured by varying signal amplitude according to the up–down staircase method. The resulting underwater audiogram (50% detection thresholds) for the male Steller sea lion showed the typical mammalian U-shape. His maximum sensitivity (77 dB *re*: 1 μPa, rms) occurred at 1 kHz. The range of best hearing (10 dB from the maximum sensitivity) was from 1 to 16 kHz (4 octaves). Higher hearing thresholds (indicating poorer sensitivity) were observed below 1 kHz and above 16 kHz. The maximum sensitivity of the female (73 dB *re*: 1 μPa, rms) occurred at 25 kHz. Higher hearing thresholds (indicating poorer sensitivity) were observed for signals below 16 kHz and above 25 kHz. At frequencies for which both subjects were tested, hearing thresholds of the male were significantly higher than those of the female. The hearing sensitivity differences between the male and female Steller sea lion in this study may be due to individual differences in sensitivity between the subjects or due to sexual dimorphism in hearing.
© 2005 Acoustical Society of America. [DOI: 10.1121/1.1992650]

I. INTRODUCTION

There are nine species of sea lions and fur seals living in both hemispheres. Of these Otariid pinnipeds, the Steller sea lion (*Eumetopias jubatus*) is the largest and shows the most marked sexual dimorphism. Females weigh up to 350 kg and adult males may be up to three times heavier. The Steller sea lion occurs in the subarctic waters of the North Pacific Ocean. It leads an amphibious life and spends much time both resting on land and in the water during activities such as migration, foraging, and courtship (Loughlin, 2002). Since the mid-1970s the western population of Steller sea lions has declined and the species was declared endangered in 1997 (Trites and Larkin, 1996; Merrick *et al.*, 1997). One of the factors causing this decline could be disturbance by increased anthropogenic underwater noise (Akamatsu *et al.*, 1996).

Steller sea lions produce both aerial and underwater vocalizations. The loud aerial vocalizations produced by bulls are described as belches, growls, snorts, scolds, and hisses; these signals seem to have a social function mainly related to territorial behavior during the breeding season. The aerial sounds emitted by females and their pups are described as bellows and as resembling the bleating of sheep (Ono, 1965; Gentry, 1968; Orr and Poulter, 1967). Steller sea lions also produce a large variety of underwater vocalizations described as belching, barking, and clicks (Orr and Poulter, 1967; Schusterman *et al.*, 1970; Poulter and del Carlo, 1971).

To determine the importance of sound for Steller sea lions during activities such as communication, reproduction, predator avoidance, and navigation, and the potential for disturbance by anthropogenic noise, information is needed on the species' hearing sensitivity both in air and underwater. However, neither aerial nor underwater hearing sensitivities of the Steller sea lion have been tested. Underwater audiograms have so far been obtained for only nine of the 34 pinniped species of the three families (*Phocidae, Otariidae,* and *Odobenidae*). To date, hearing profiles are available for only two of nine *Otariids*. These profiles are based on information from only one or two animals per species: a 5–6-
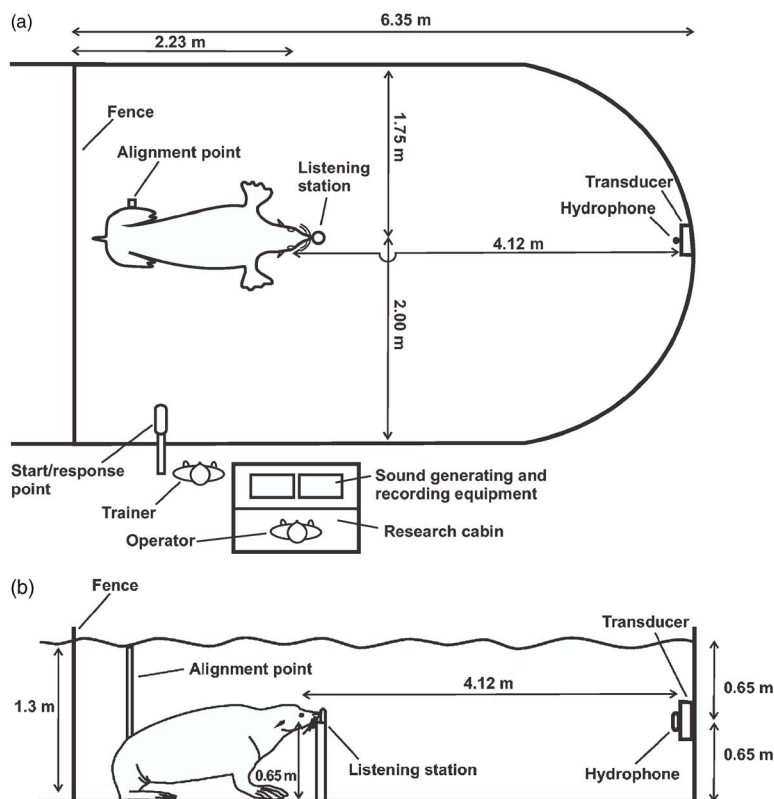
FIG. 1. The study area, showing one of the Steller sea lions under water in the correct position at the listening station. (a) top view and (b) side view, both to scale.

year-old male California sea lion, *Zalophus californianus* (Schusterman *et al.*, 1972), a 12-year-old male California sea lion (Kastak and Schusterman, 2002) and two 2–3-year-old female Northern fur seals, *Callorhinus ursinus* (Moore and Schusterman, 1987). Due to the small sample size per species, very little is known about the hearing demographics in pinniped species including differences between individuals that might be related to age, size, and sex. Steller sea lion hearing may not resemble that of the *Otariids* which have been tested. A recent study of the underwater hearing sensitivity of the large Pacific walrus (*Odobenus rosmarus divergens*) shows a limited frequency hearing range for this species (Kastelein *et al.*, 2002). The limited frequency range may be due to the large size of the hearing organ structures in the walrus (Kastelein *et al.*, 1996) compared to those of the smaller pinniped study subjects of which the hearing sensitivity has been tested. Because Steller sea lions show a large sexual difference in body size, large differences in the sizes of the structures of the hearing organs probably exist. This may lead to sexual differences in hearing sensitivity or frequency range of hearing, as is found in terrestrial mammals (Heffner *et al.*, 2001). It is not clear if sizes of the middle and inner ears are related to the hearing sensitivity in marine mammals.

In order to measure hearing sensitivity of Steller sea lions, and to evaluate how hearing sensitivity varies between individuals of different sizes and sexes, we tested, under identical conditions, the underwater hearing (frequency range of hearing and frequencies of best hearing sensitivity) of two captive Steller sea lions of similar age: an adult male and an adult female.

II. MATERIALS AND METHODS

A. Study animals

The study animals were a male (code EjZH021) and a female (code EjZH022) Steller sea lion which were collected for research purposes in British Columbia, Canada, at the age of a few weeks and housed at the Vancouver Aquarium until the age of 4 years, after which they were sent to Dolfinarium Harderwijk, The Netherlands. During the current experiment, the animals were healthy. The male was 8 years old and his body weight varied between 590 and 960 kg depending on the season. The female was 7 years old and her body weight varied between 213 and 228 kg. Veterinary records showed that the animals had not been exposed to ototoxic medication. The male received between 20 and 40 kg and the female between 7 and 15 kg of thawed fish (herring, *Clupea harengus*; mackerel, *Scomber scombrus*; scad, *Trachurus trachurus*, and sprat, *Sprattus sprattus*) per day depending on the season (Kastelein *et al.*, 1990), divided over four to seven meals.

B. Study area

The experiment was conducted at one end of a C-shaped outdoor concrete pool with an adjacent haul-out space. The portion of the pool used for the experiment was 9 m(l) ×3.75 m(w), 1.3 m deep (Fig. 1). The water level was kept constant. The average monthly water temperature varied between 3 and 22 °C, and the salinity was approximately 2.5% NaCl. The water circulation pump was switched off 10 min before and during sessions, so that there was no pump noise or water current in the pool during the experiments. No other

TABLE I. The mean 50% detection thresholds of an 8-year-old male Steller sea lion for eight narrow-band FM signals (based on total number of reversals), mean session threshold range, number of sessions conducted, number of reversals used to calculate the mean detection threshold, and prestimulus response rate based on the number of prestimulus responses in all trials (signal-present+signal-absent trials). SPL in dB *re*: 1 μPa, rms.

Center frequency (kHz)	Frequency modulation range (kHz)	Mean 50% detection threshold (SPL)	Mean session threshold range (SPL)	No. of sessions	Total no. of reversals	Prestimulus response rate (%)
0.5	0.495–0.505	100	93–107	4	40	23
1	0.99–1.01	77	72–80	10	122	16
2	1.98–2.02	81	77–83	10	176	7
4	3.96–4.04	82	77–89	10	150	7
8	7.92–8.08	87	78–92	5	58	4
16	15.84–16.16	85	77–88	6	76	3
25	24.75–25.25	90	86–92	4	70	2
32	31.68–32.32	99	94–101	3	40	4

animals were present in the pool during the tests, as the one or two pool mates (one of which was the other study subject) were trained to stay ashore during sessions. The equipment used to produce the sound stimuli was housed out of sight (above) of the study animals, in an observation and data collection cabin that was located 4 m away from the animals and trainer (Fig. 1).

C. Test stimuli production and calibration

Narrow-band sinusoidal frequency-modulated (FM) signals were produced by a waveform generator (Hewlett Packard, model 33120A). The modulation range of the signal was ±1% of the center frequency (the frequency around which the signal fluctuated symmetrically), and the modulation frequency was 100 Hz. For example, if the center frequency was 10 kHz, the frequency fluctuated 100 times per second between 9.9 and 10.1 kHz. Tables I and II show the frequency ranges of the signals.

In most previous studies of pinniped hearing, except in an experiment in which the hearing sensitivity of a Pacific walrus was tested (Kastelein *et al.*, 2002), pure tones have been used as test signals. In the present study, narrow-band FM signals were used because such signals reduce the level of constructive and destructive interference effects (standing waves) on the signals reaching the animal in a reverberant

pool. Sound measurements showed that the exposure level was more constant when using FM signals than when using pure tones. Within a measurement session, the maximum variation between SPL measurements of the FM signals varied per frequency between 0.2 and 6 dB, but was generally around 3 dB. In humans, FM signals tend to have a slightly higher arousal effect than pure tones, which may lead to slightly lower (<5 dB depending on center frequency and modulation frequency) hearing thresholds (Morgan *et al.*, 1979). The test signals had no harmonics with sound pressure levels (SPLs) near the sea lions' hearing thresholds.

A modified audiometer used for testing human aerial hearing (Midimate, model 602) was used to control the duration and amplitude of signals (Fig. 2). During the experiment, the stationary portion of the signal was 600 ms in duration. The onset and offset of the signal were delayed with a rise and fall time (each 50 ms) to prevent transients. This signal duration was expected to be acceptable in relation to the hearing system's integration time (based on the integration time of the harbor seal, *Phoca vitulina*; Terhune, 1988). The SPL at the sea lion's head while at the listening station could be varied in 5-dB increments (this step size was determined by the audiometer; 5-dB steps are generally used in human audiometry). Before each session the sound generating system was checked in two ways. The voltage output

TABLE II. The mean 50% detection thresholds of a 7-year-old female Steller sea lion for five narrow-band FM signals (based on total number of reversals), mean session threshold range, number of sessions conducted, number of reversals used to calculate the mean detection threshold, and prestimulus response rate based on the number of prestimulus responses in all trials (signal-present+signal-absent trials). SPL in dB *re*: 1 μPa, rms.

Center frequency (kHz)	Frequency modulation range (kHz)	Mean 50% detection threshold (SPL)	Mean session threshold range (SPL)	No. of sessions	Total no. of reversals	Prestimulus response rate (%)
4	3.96–4.04	81	76–83	3	38	11
8	7.92–8.08	80	78–82	2	24	2
16	15.84–16.16	79	72–83	2	22	0
25	24.75–25.25	73	71–74	2	24	2
32	31.68–32.32	79	77–80	2	24	7
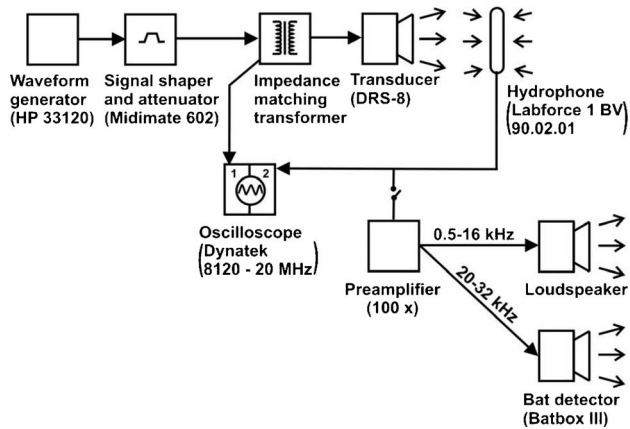
FIG. 2. Block diagram of the signal generation and listening system used in the Steller sea lion underwater hearing study.

level of the system, at the input of the impedance-matching transformer (while the attenuator was at the same setting as during calibrations), was checked with an oscilloscope (Dynatek 8120, 20 MHz; Fig. 2). In addition, the voltage output of a hydrophone (Labforce 1 BV, model 90.02.01) placed in front of the transducer was checked when a signal was produced.

The signals were projected by an underwater piezoelectric transducer (Ocean Engineering Enterprise, model DRS-8; 25 cm diameter) with its impedance-matching transformer (Fig. 2). The transducer was fixed to the pool wall in a protective stainless-steel cage during the entire study period. It was 4 m in front of the study animals when they were positioned at the listening station. The transducer was positioned with the acoustic axis of the projected sound beam pointed at the center of the animal's head.

The male Steller sea lion's underwater hearing sensitivity was measured for signals with center frequencies of 0.5, 1, 2, 4, 8, 16, 25, and 32 kHz. The female's hearing was tested for signals with center frequencies of 4, 8, 16, 25, and 32 kHz. The female's hearing was tested for fewer frequencies than the male's because she gave birth and therefore had to be separated from the male.

Before each session, the test signal (at a sufficient SPL) was checked aurally by the signal operator via the hydrophone which was mounted just in front of the transducer. The hydrophone's output was connected to an amplifier and loudspeaker or, for the ultrasonic signals (>20 kHz), to a bat detector (Batbox III; Stag Electronics, Steyning, UK).

The root-mean-square (rms) SPL (dB *re*: 1 μPa) of each test frequency was measured approximately once each month at the sea lions' typical head position when the animals were at the listening station during the tests (Fig. 1). The sea lions were not in the pool during these calibrations. During trials, the location of the sea lion's head (while at the listening station) relative to the transducer was carefully checked by the trainer to ensure that head position was consistent to within a few centimeters.

The calibration equipment used for all signals consisted of a broadband hydrophone [Brüel & Kjaer (B&K) 8101], with flat frequency response (within 1 dB) in the tested frequency range, a conditioning amplifier (B&K, Nexus 2690),

a computer with a data acquisition card (National Instruments, PCI-MIO-16E-1, 12-bit resolution), and a coaxial module to receive the input signals (National Instruments, model BNC-2090). The system was calibrated with a pistonphone (B&K, 4223). The signals were digitized at a sample rate of 512 kHz and fast Fourier transformed (FFT) into the frequency domain using a Hanning window. The highest peak in the spectrum was selected to determine the SPL, and five consecutive 0.2-s time blocks were used to calculate the average SPL per calibration session. The maximum SPL variation between calibration sessions varied per frequency between 0 and 4 dB, but was generally around 2 dB. The average SPL per frequency was calculated based on all calibration sessions. These averages were used to determine the session thresholds. The SPLs 20 cm in all six directions from the auditory meatus of the animals varied by 0−2 dB.

The SPL was calibrated at a level 10−30 dB above the threshold levels found in the present study. The linearity of the attenuation of the audiometer was checked three times during the study and was precise. The spectra of the signals were monitored to detect potential surface reflections causing cancellations. Multipath arrivals and standing waves introduce both temporal and spatial variations in the observed SPL at the listening station. The use of frequency-modulated stimuli generally resulted in smaller amplitude variations (generally 3 dB) at the listening station than pure tones.

D. Background noise

No activity took place near the pool during sessions, and the water pump in a nearby engine room was switched off. Underwater background noise levels were measured under the same conditions as during the sessions.

The equipment used to measure the background noise in the pool differed from the equipment used to calibrate the signals. The equipment consisted of a broadband (0−100 kHz) hydrophone (B&K, 8101), a voltage amplifier system (TNO TPD, 0−300 kHz), and an analyzer system (Hewlett Packard 3565, controlled by a Toshiba Pro 4200 notebook computer; frequency range 0−80 kHz, sample frequency 260 kHz, df=31 Hz, FFT measurement converted to 1/3-octave bandwidths). The total system was calibrated with a pistonphone (B&K, 4223) and a white noise "insert voltage signal" into the hydrophone preamplifier. Measurements were corrected for the frequency sensitivity of the hydrophone and the frequency response of the measurement equipment.

Background noise levels are given as "equivalent sound-pressure spectrum levels" (L_{eq} method), i.e., the time-averaged levels of the fluctuating noise (Hassall and Zaveri, 1988). An equivalent sound-pressure level would produce the same acoustic energy, over a stated time period, as a specified time-varying noise. Background noise was converted to "spectrum level" (dB *re*: 1 μPa/$\sqrt{\text{Hz}}$) using the formula $L_{sp}=L_{eq}-10 \log$ BW, in which L_{sp}=equivalent sound spectrum level (in dB *re*: 1 μPa), L_{eq}=equivalent sound pressure level, and BW=bandwidth of 1/3-octave bands (in Hz). Figure 3 shows the background noise levels in the pool, converted to spectrum level (L_{eq}, 1-Hz bandwidth,
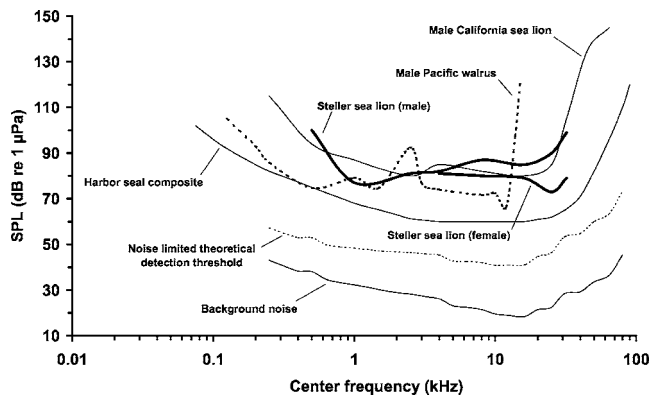
FIG. 3. The mean 50% detection thresholds in dB *re*: 1 μPa (rms) for narrow-band FM signals obtained for the male and female Steller sea lion in the present study (for details, see Tables I and II). Also shown are the underwater audiograms of a harbor seal [a composite of data from Møhl (1968a), Turnbull and Terhune (1990), and Kastak and Schusterman (1998), with potentially masked thresholds adapted], a California sea lion (Schusterman *et al.*, 1972), and a male Pacific walrus (Kastelein *et al.*, 2002). The background noise in the pool of the present study between 250 Hz and 80 kHz is plotted in dB *re*: 1 μPa/√Hz (derived from 1/3-octave band levels). Also shown is the noise-limited theoretical detection threshold level based on the background noise level, an estimated sea lion critical ratio (derived from Moore and Schusterman, 1987; Southall *et al.*, 2000, 2003) and an assumed directivity index for sea lions.

range 250 Hz–80 kHz). Also shown in this figure is the noise-limited theoretical detection threshold curve based on the background noise level in the pool, an estimated sea lion critical ratio [derived from Moore and Schusterman, 1987 (Northern fur seal), and Southall *et al.*, 2000, 2003 (California sea lion)] and an assumed directivity index for sea lions. (The noise-limited theoretical detection threshold =background noise+critical ratio−directivity index).

E. Experimental procedure

Training the sea lions for the go/no-go underwater hearing procedure took 1 month. Operant conditioning using positive reinforcement was used to condition the animals to produce a response (go) in the presence of a signal and to withhold the response (no-go) in the absence of the signal. Experimental sessions consisted of multiple trials. A trial began when one of the animals was positioned at a start and response point on land next to the trainer [Fig. 1(a)]. When the trainer gave the animal the vocal command accompanied by a gesture (pointing downward), the animal descended to the listening station (a nylon ball on a wooden pillar), so that its external auditory meatus was 412 cm from the sound source and about 65 cm below the water surface [Fig. 1(b)]. Each animal was trained to position its nose against the listening station so that its head axis was in line with the projected beam axis. To assist them to reach the correct alignment with the transducer's beam, the animals were trained to put their left hind flipper against a bar on a fence behind them.

Two trial types were presented during each experimental session: signal-present trials and signal-absent trials. In signal-present trials, the stimulus was presented unpredictably between 4 and 10 s after the animal was positioned correctly at the listening station. A minimum waiting time of

4 s was chosen because it took about 4 s for the waves, created by the animal's descent, to dissipate. If the animal detected the sound, it responded by leaving the listening station (go response) at any time during the signal's duration and returning to the start and response point [Fig. 1(a)]. The signal operator then informed the trainer (via a headset radio) that the response was correct (a hit), after which the trainer gave a vocal signal and the sea lion received a fish reward. If the animal did not respond to the signal, the signal operator informed the trainer that the animal had failed to detect the signal (a miss). The trainer then signaled to the animal (by tapping the fish bucket on the floor) that the trial had ended, thus calling the animal back to the start and response point. No reward was given following a miss. If the animal moved away before a signal was produced (a prestimulus response or false alarm), the signal operator told the trainer to end the trial without reinforcement.

On signal-absent, or catch trials, the signal operator told the trainer to end the trial after a random interval of 4 to 10 s after the sea lion had stationed. The trial was terminated when the trainer blew a whistle. If the animal responded correctly by remaining at the listening station until the whistle was blown (a correct rejection), it then returned to the start and response point and received a fish reward. If the sea lion left the listening station before the whistle was blown (a prestimulus response or false alarm), the signal operator told the trainer to end the trial without rewarding the animal. The same amount of fish was given as a reward for correct go and no-go responses. In both signal-present and signal-absent trials, the trainer was unaware of the trial type until the animal had responded.

A session generally consisted of 20–24 trials and lasted for about 20 min. Each session consisted of 50% signal-present and 50% signal-absent trials presented in random order based on a pseudorandom series table (Gellermann, 1933; maximum of three consecutive similar trials), with the modification that the first trial in a session was always a signal-absent trial. Each day one of four data collection sheets with different random series was used. Over the 18-month study, six different sets of four data collection sheets were used.

In each session, the signal frequency was held constant and the signal amplitude was varied according to a modified up/down staircase psychometric technique (Robinson and Watson, 1973). This is a variant of the method of limits, which results in a 50% correct detection threshold (Levitt, 1971). Prior to testing at a given frequency, an estimated threshold was determined during preliminary sessions, in which the rough hearing threshold per test frequency was determined using the up–down staircase method.

During subsequent experimental sessions, the starting SPL of the signal was 10–15 dB above the estimated threshold. Following each hit, the signal amplitude on the next signal-present trial was reduced by 5 dB. Following each miss, the signal level was increased on the next signal-present trial by 5 dB. Prestimulus responses (false alarms) did not lead to a change in signal amplitude for the next trial.

TABLE III. Two-way ANOVA to test for differences in hearing sensitivity between the male and female Steller sea lion. There is a significant difference between the hearing thresholds of the two sea lions, and the signal frequency has a significant effect on the hearing threshold.

Source of variation	Degrees of freedom	Adjusted means square	Test statistic (F)	Probability (P)
Sea lion	1	9147.6	292.89	0.000
Signal frequency	4	734.3	23.51	0.000
Sea lion × signal frequency	4	1239.3	39.68	0.000
Error	500	31.2		
Total	509			

A switch in the sea lion's response from a detected signal (a hit) to an undetected signal (a miss), or vice versa, is called a reversal.

Thresholds were determined for eight frequencies for the male sea lion and five frequencies for the female sea lion. In order to prevent the animals' learning process from affecting the shape of the audiogram, the test frequency was varied from session to session. The sequence of frequencies, tested in successive sessions, were neighbors in the following list: 0.5, 1, 2, 4, 8, 16, 25, and 32 kHz. This way, the difference in frequency between sessions was limited, reducing the potential need for adaptation to the frequency.

One experimental session was conducted daily (maximum 5 days/week) during the first meal of the day between 0800 and 0820 h, when the sea lions had not been fed for 16 h, and the park was still closed to visitors. Data for the male were collected between June and August 2001, and between May and November 2002. Both years, no tests were conducted during the Steller sea lion's rutting period in July due to loss of food motivation (Kastelein *et al.*, 1990). Data for the female were collected between July and August 2001. It was the intention to conduct at least ten sessions per frequency and to test frequencies below 0.5 kHz and above 32 kHz. However, the study period was shorter than previously planned because access to the animals for research was not granted after October 2002.

F. Analysis

Sound levels at which reversals took place were taken as data points. The mean 50% detection threshold per frequency was defined for each animal as the mean amplitude of all the reversal data points obtained in all available sessions after the mean session thresholds leveled off, which usually occurred after two or three sessions. Sessions with more than 20% prestimulus responses (four out of the usual 20 trials per session) were not included in this analysis. These occurred only about ten times during the entire study, and usually coincided with obvious transient background noises.

The hearing thresholds obtained for each animal were compared across frequencies using a one-way ANOVA to evaluate within-subject effects of signal frequency on sensitivity. Tukey tests were used to compare sensitivities at the various signals frequencies. The hearing thresholds of the two animals at the five frequencies they both were tested for were compared using a two-way ANOVA to identify individual differences in sensitivity.

All statistical analysis was carried out on MINITAB for Windows version 13 (Ryan and Joiner, 1994) with a significance level of 5%, using the hearing threshold as the dependent variable. Test assumptions were met in all tests (Zar, 1984).

III. RESULTS

The underwater audiogram (50% detection thresholds) for the male Steller sea lion shows the typical mammalian U-shape (Fig. 3 and Table I). His maximum sensitivity (77 dB *re*: 1 μPa, rms) occurred at 1 kHz. The range of best hearing (10 dB from the maximum sensitivity) was from 1 to 16 kHz (4 octaves), and sensitivity fell below 1 kHz and above 16 kHz. After the initial two or three sessions of a frequency, which were not included in the analysis, the animal's sensitivity for each test frequency was stable over the 1.5-year study period. The male's average prestimulus response rate (for both signal-present and signal-absent trials) varied between 2% and 23%, depending on the frequency (Table I). Most prestimulus responses occurred during tests with low-frequency signals. There was a significant difference in the hearing threshold of the male Steller sea lion due to frequency ($F_{7,724}=164.15, P<0.001$). Tukey tests showed that the thresholds for the frequencies could be ordered as follows: $1<2=4<16=8<25<32=0.5$ kHz.

The female's maximum sensitivity (73 dB *re*: 1 μPa, rms) occurred at 25 kHz (Fig. 3 and Table II). After the initial two or three (depending on the test frequency) sessions, which were not included in the analysis, the female's sensitivity for each test frequency was stable over the 3-month study period. The female's average prestimulus response rate (for both signal-present and signal-absent trials) varied between 0% and 11%, depending on the frequency (Table II). There was a significant difference in the hearing threshold of the female Steller sea lion due to frequency ($F_{4,111}=9.67, P<0.001$). Tukey tests showed that the thresholds fell into two significantly different groups, such that the threshold for 25 kHz was significantly lower than that for 4, 8, and 32 kHz. The threshold for 16 kHz was not significantly different from those of either group.

At frequencies for which both subjects were tested, hearing thresholds of the male were significantly higher than those of the female (Table III).

IV. DISCUSSION AND CONCLUSIONS

A. Evaluation of the data

It is important to know whether the audiograms of the present study are absolute audiograms or if the signals were masked by the ambient noise in the pool. A calculation [detection threshold=background noise+critical ratio −directivity index] shows that the audiograms found in the present study are considerably (more than 10 dB) above the noise limited theoretical detection threshold, and thus represent absolute audiograms (Fig. 3).

In contrast to the slightly lower hearing thresholds found in humans when using FM signals instead of pure-tone signals (Morgan *et al.*, 1979), the use of FM signals instead of pure tones probably had little effect on the thresholds found in the present study. This assumption is based on the results from a hearing test with 250-Hz signals on a Pacific walrus. No difference was found in thresholds between the same narrow-band FM signals as used in the present study (frequency modulation only 1% of the center frequency), and pure-tone signals (Kastelein *et al.*, 2002). In addition, a study with California sea lions, in which much wider band FM signals were used (1-octave sweeps in 10–50 ms: 3.2→6.4 kHz and 6.4→3.2 kHz; 5-ms rise and fall times), showed similar evoked potential thresholds to those obtained with pure tones (Bullock *et al.*, 1971).

Signal duration can influence hearing thresholds. In humans, for instance, aerial hearing thresholds begin to increase when signal duration drops below a certain value. However, several threshold durations are mentioned in literature: 10 s (Garner, 1947); 1 s (Plomp and Bouman, 1959); 300 ms (Watson and Gengel, 1969). From Johnson (1968) 1 s can be derived as the relevant value for bottlenose dolphins (*Tursiops truncatus*). In harbor seals, phocid pinnipeds, sharp increases in hearing thresholds only occur when signal duration becomes less than 50 ms (Terhune, 1988; 1989). Kastelein *et al.* (2002) found no statistical differences between hearing thresholds in a Pacific walrus for signals of 1500 and 300 ms. The integration time of the walrus is probably also shorter than 300 ms. As the hearing system integration time is frequency dependent and decreases for increasing frequencies, it may be expected that the 600-ms signal duration used in the present study was more than the integration time of the Steller sea lion's hearing system. A short test signal duration (but above the integration time) has two advantages: it reduces the chance of the animal being able to adjust the position of its head in order to find regions of slightly higher SPL due to multipath effects, and it reduces the chance of the animal producing a false alarm during signal presentations below the hearing threshold. The 600-ms signal used in the present study was long enough to yield an accurate best hearing sensitivity curve but short enough to avoid the disadvantages of longer signals.

For some frequencies the sample size was relatively small due to the unexpected early termination of the study. However, when the study was interrupted for 6 months and then resumed, the mean hearing thresholds of the male from before the break were similar to those after the break (varying 2–4 dB depending on the test frequency).

B. Comparison with other pinniped hearing studies

Comparing the findings of the present study to those of other pinniped hearing studies is difficult, since the accuracy of the studies often cannot be established. Information on the calibration methodology, threshold calculation, and variation in the threshold data between sessions is often limited, and results are not always free from masking influences. In addition, most studies have used different methodology, stimuli parameters such as signal type (pure tone versus FM signal) and signal duration, ways to report SPL (peak-to-peak or rms, causing a 9-dB difference), the level calculation method is often not specified, and the studies were conducted at different depths. Also, the physical effort an animal has to exert to show response probably affects the threshold level. In some cases animals only had to press a pedal next to them as a response, while in the present study the sea lion had to move its entire body several meters. This probably led to a conservative audiogram, but most other audiograms on pinniped underwater hearing are probably conservative as well.

Despite these limitations, general comparisons can be made between the underwater audiogram of the Steller sea lions and the underwater audiograms of the other pinnipeds of which the underwater hearing has been tested. This is the case for two *Otariid* species: California sea lion (Schusterman *et al.*, 1972; Kastak and Schusterman, 1998; 2002), Northern fur seal (Schusterman and Moore, 1978; Moore and Schusterman, 1987), six *Phocid* species: harbor seal (Møhl, 1968a, b; Terhune, 1988; Terhune and Turnbull, 1995; Turnbull and Terhune, 1994; Kastak and Schusterman, 1998), ringed seal, *Pusa hispida* (Terhune and Ronald, 1975), harp seal, *Pagophilus groenlandicus* (Terhune and Ronald, 1972), gray seal, *Halichoerus grypus* (Ridgway and Joyce, 1975), Hawaiian monk seal, *Monachus schauinslandi* (Thomas *et al.*, 1990), Northern elephant seal, *Mirounga angustirostris* (Kastak and Schusterman, 1998; 1999), and one *Odobenid* species: the Pacific walrus, *Odobenus rosmarus divergens* (Kastelein *et al.*, 2002). Of these studies, the latter was conducted with the same equipment and methodology as the present study and can therefore easily be compared.

The hearing sensitivity of the male Steller sea lion in the present study is fairly similar to that of the male California sea lion tested by Schusterman *et al.* (1972). However, the hearing sensitivity of the female Steller sea lion in the present study was higher than that of the male California sea lion for frequencies above 20 kHz (Fig. 3). For the frequencies that were tested in both species, the hearing sensitivity of the two much smaller female Northern fur seals (Moore and Schusterman, 1987) was higher than that of the female Steller sea lion in the present study. At 1 kHz, the hearing of the male Steller sea lion in the present study was 10–20 dB better than that of the male California sea lion and the female Northern fur seals (Schusterman *et al.*, 1972; Schusterman and Moore, 1978; Moore and Schusterman, 1987; Kastak and Schusterman, 1998). In three studies with seals (ringed seals, harp seals, and a female Northern elephant seal), the underwater hearing sensitivity was tested for frequencies below 1 kHz (Terhune and Ronald, 1972; 1975; Kastak and Schusterman, 1998). The Northern elephant seal can hear

lower frequencies better than the smaller *Phocids* that have been tested, and, at least the female Northern elephant seal, can hear high frequencies well (Kastak and Schusterman, 1998; 1999).

Maybe per pinniped group and within species a correlation exists between body size and hearing sensitivity. The differences in hearing sensitivity between the male and the female Steller sea lion in the present study may be due to individual differences in sensitivity between the subjects or due to sexual dimorphism in hearing. In the latter, the hearing differences may be due to differences in the sizes of structures in the hearing organs. This idea is based on the limited hearing range found in the largest pinniped of which the hearing has been tested so far: a male Pacific walrus (Kastelein *et al.*, 2002). The walrus has large middle- and inner-ear structures compared to the other pinnipeds of which the hearing has been tested (Kastelein and Gerrits, 1990; Kastelein *et al.*, 1996). The physics of sound may relate to the size of ear structures; the larger the size, the more suitable for low-frequency sound reception (Wartzok and Ketten, 1999). It would therefore be of interest to test the underwater hearing sensitivity of an adult male Northern elephant seal, which is the only pinniped in which the male is as big as, or bigger than, an adult male walrus.

C. Ecological significance and suggestions for future research

The frequency range of underwater vocalizations made by Steller sea lions is unknown, as the equipment used by earlier investigators could only record audible sounds up to 20 kHz or less (Orr and Poulter, 1967; Schusterman *et al.*, 1970; Poulter and del Carlo, 1971). However, the maximum underwater hearing sensitivity range (800 Hz–25–30 kHz) obtained in the present study for Steller sea lions overlaps with the frequency range of the underwater vocalizations recorded in those studies. Vocalizations and hearing in males are very important during territorial disputes in the breeding season. One could speculate that for this reason their hearing is more tuned to low-frequency sounds. Females need to find their pups among many other pups in a breeding colony when they return from feeding trips at sea. One could speculate that their hearing is more tuned towards the higher frequencies of pup calls.

The hearing range found in the present study also suggests that Steller sea lions can hear many of the social calls of one of their two main predators, the killer whale (*Orcinus orca*), which are between 500 Hz and 35 kHz (Schevill and Watkins, 1966; Awbrey *et al.*, 1982; Ford, 1989). Steller sea lions can probably also hear some of the killer whale's echolocation clicks which are between 4 and 50 kHz (Schevill and Watkins, 1966; Diercks *et al.*, 1971; Awbrey *et al.*, 1982; Barrett-Lennard *et al.*, 1996). In addition, the frequency of the click repetition rate of killer whales' echolocation signals (10–330 Hz) may sometimes be within the Steller sea lion's range of hearing. It is possible that the relatively high-frequency underwater hearing is related to predator avoidance rather than communication between conspecifics.

Most anthropogenic noise contains energy up to 1 kHz. The lower-frequency components of anthropogenic noise travel further than the higher components. The results from the present study show that Steller sea lions can hear some underwater anthropogenic noise. The effect of audible anthropogenic noise on marine mammals is highly variable in type and magnitude (Severinsen, 1990; Cosens and Dueck, 1993; Richardson *et al.*, 1995), and Steller sea lions are sensitive to acoustic disturbance in certain contexts (Hoover, 1988). A controlled acoustic experiment showed that Steller sea lions sometimes hauled out of the water as a reaction to certain loud sounds (Akamatsu *et al.*, 1996). Anthropogenic noise might reduce the time Steller sea lions forage in particular areas, thus reducing the physiological condition of the animals, and thus their reproductive success.

Therefore, bottom trawl fishing, tanker routes, and drilling platforms should be planned far enough away from areas that are important in Steller sea lion ecology. In addition to the hearing sensitivity of the Steller sea lion, the radius of the discomfort zone depends on several other factors such as the general ambient noise level, water depth, ocean floor sediment, and the spectrum, source level, and duration of the anthropogenic noise. Pinniped hearing is generally more sensitive to anthropogenic noise than is odontocete hearing, which is more sensitive to higher frequencies. This is because odontocetes use echolocation.

To estimate at what distances Steller sea lions can hear conspecifics, echolocation clicks of killer whales, and anthropogenic noise, additional information is needed. Information needs to be obtained on how Steller sea lions hear in the presence of masking noise (critical ratios, critical bands), how they hear sounds of different durations (especially pulsed sounds), and how they perceive the spatial positioning of sounds coming from different directions (receiving beam pattern). Also, the effect of loud pulsed signals on hearing thresholds for pure tones could be studied, as has already been done with California sea lions (Finneran *et al.*, 2003). The effect of loud noise on Steller sea lion hearing should also be tested as has already been done in three other pinniped species (Kastak *et al.*, 1999)

Akamatsu, T., Nakamura, K., Nitto, H., and Watabe, M. (**1996**). "Effects of underwater sounds on escape behaviour of Steller sea lions," Fish. Sci. **62**, 503–510.

J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 1, September 2005

Kastelein *et al.*: Underwater hearing in Steller sea lions   1827

Awbrey, F. T., Thomas, J. A., Evans, W. E., and Leatherwood, S. (**1982**). "Ross sea killer whale vocalizations: Preliminary description and comparison with those of some Northern Hemisphere killer whales," Rep. Int. Whal. Commn. **32**, 667–670.

Barrett-Lennard, L. G., Ford, J. F. B., and Heise, K. A. (**1996**). "The mixed blessing of echolocation: Differences in sonar use by fish-eating and mammal-eating killer whales," Anim. Behav. **51**, 553–565.

Bullock, T. H., Ridgway, S. H., and Suga, N. (**1971**). "Acoustically evoked potentials in midbrain auditory structures in sea lions," Z. Vergl. Physiol. **74**, 372–387.

Cosens, S. E., and Dueck, L. P. (**1993**). "Icebreaker noise in Lancester sound, N.W.T., Canada: Implications for marine mammal behavior," Marine Mammal Sci. **9**, 285–300.

Diercks, K. J., Trochta, R. T., Greenlaw, C. F., and Evans, W. E. (**1971**). "Recordings and analysis of dolphin echolocation signals," J. Acoust. Soc. Am. **49**, 1729–1732.

Finneran, J. J., Dear, R., Carder, D. A., and Ridgway, S. H. (**2003**). "Auditory and behavioral responses of California sea lions (*Zalophus californianus*) to single underwater impulses from an arc-gap transducer," J. Acoust. Soc. Am. **114**, 1667–1677.

Ford, J. K. B. (**1989**). "Acoustic behavior of resident killer whales (*Orcinus orca*) off Vancouver Island, Britisch Columbia," Can. J. Zool. **67**, 727–745.

Garner, W. R. (**1947**). "The effect of frequency spectrum in the ear," J. Acoust. Soc. Am. **19**, 808–815.

Gellermann, L. W. (**1933**). "Chance orders of alternating stimuli in visual discrimination experiments," J. Gen. Psychol. **42**, 206–208.

Gentry, R. L. (**1968**). "Territoriality and reproductive behavior in make Steller sea lions," Am. Zool. **8**, 739.

Hassall, J. R., and Zaveri, . (**1988**). Acoustic Noise Measurements, Brüel & Kjaer documentation.

Heffner, R. S., Koay, G., and Heffner, H. E. (**2001**). "Audiograms of five species of rodents: Implication for the evolution of hearing and the perception of pitch," Hear. Res. **157**, 138–152.

Hoover, A. A. (**1988**). "Steller sea lion, *Eumetopias jubatus*," in *Selected Marine Mammal of Alaska: Species Accounts with Research and Management Recommendations*, edited by J. W. Lentfer (Marine Mammal Commission, Washington, D.C.), pp. 159–193.

Johnson, C. S. (**1968**). "Relation between absolute threshold and duration of tone pulses in the bottlenosed dolphin," J. Acoust. Soc. Am. **43**, 757–763.

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., and Schusterman, R. J. (**1999**). "In-air and underwater hearing of a northern elephant seal (*Mirounga angustirostris*)," Can. J. Zool. **77**, 1751–1758.

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**, 1142–1148.

Kastak, D., and Schusterman, R. J. (**2002**). "Changes in auditory sensitivity with depth in a free-diving California sea lion (*Zalophus californianus*)," J. Acoust. Soc. Am. **112**, 329–333.

Kastelein, R. A., and Gerrits, N. M. (**1990**) "The anatomy of the Walrus head (*Odobenus rosmarus*). I: The skull," Aquatic Mammals **16**, 101–119.

Kastelein, R. A., Vaughan, N., and Wiepkema, P. R. (**1990**). "The food consumption of Steller sea lions (*Eumetopias jubatus*)," Aquat. Mammals **15**, 137–144.

Kastelein, R. A., Dubbeldam, J. L., Bakker, M. A. G. de, and Gerrits, N. M. (**1996**). "The Anatomy of the walrus head (*Odobenus rosmarus*). IV. The ears and their function in aerial and underwater hearing," Aquat. Mammals **22**, 95–125.

Kastelein, R. A., Mosterd, P., van Santen, B., Hagedoorn, M., and de Haan, D. (**2002**). "Underwater audiogram of a Pacific walrus (*Odobenus rosmarus divergens*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 2173–2182.

Levitt, H. (**1970**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Loughlin, T. R. (**2002**). "Steller's Sea Lion, *Eumetopias jubatus*", in Encyclopedia of Maine Mammals, edited by W. F. Perrin, B. Würsig., and J. G. M. Thewissen (Academic, New York), pp. 1181–1185.

Merrick, R. L., Chumbley, M. K., and Byrd, G. V. (**1997**). "Diet diversity of Steller sea lions (*Eumetopias jubatus*) and their population decline in Alaska: A potential relationship," Can. J. Fish. Aquat. Sci. **54**, 1342–1348.

Møhl, B. (**1968a**). "Auditory sensitivity of the common seal in air and water," J. Aud Res. **8**, 27–38.

Møhl, B., (**1968b**). "Hearing in seals," in *The Behavior and Physiology of Pinnipeds*, edited by R. J. Harrison, R. C. Hubbard, R. S. Peterson, C. E. Rice, and R. J. Schusterman (Appleton-Century-Crofts, New York), pp. 172–195.

Moore, P. W. B., and Schusterman, R. J. (**1987**). "Audiometric assessment of Northern fur seals, *Callorhinus ursinus*," Marine Mammal Sci. **3**, 31–53.

Morgan, D. E., Dirks, D. D., and Bower, D. R. (**1979**). "Suggested threshold sound pressure levels for frequency-modulated (warble) tones in the sound field," J. Speech Hear Disord. **44**, 37–54.

Ono, K. A. (**1965**). "Mother–pup interactions in the Steller sea lion (*Eumetopias jubatus*)," Proceedings of 2nd annual conference on the biological sonar and diving mammals, Stanford Research Institute, Fremont, California, 9–20.

Orr, R. T., and Poulter, T. C. (**1967**). "Some observations on reproduction, growth, and social behavior in the Steller sea lion," Proc. Calif. Acad. Sci. **35**, 193–225.

Plomp, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.

Poulter, T. C., and del Carlo, D. G. (**1971**). "Echo ranging signals: Sonar of the Steller sea lion," J. Aud Res. **11**, 43–52.

Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego).

Ridgway, S., and Joyce, P. L. (**1975**). "Studies on seal brain by radiotelemetry," Int. Council for the Exploration of the Sea **169**, 81–91.

Robinson, D. E., and Watson, C. S. (**1973**). "Psychophysical methods in modern Psychoacoustics," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. 2, pp. 99–131.

Ryan, B. F., and Joiner, B. L. (**1994**). MINITAB *Handbook*, 3rd ed. (Duxbury, Belmont, CA).

Schevill, W. E., and Watkins, W. A. (**1966**). "Sound structure and directionality in *Orcinus* (killer whale)," Zoologica (N.Y.) **51**, 70–76.

Schusterman, R. J., and Moore, P. (**1978**). "Underwater audiogram of the northern fur seal (*Callorhinus ursinus*)," J. Acoust. Soc. Am. **64**, S 87.

Schusterman, R. J., and Moore, P. (**1978**). "The upper limit of underwater auditory frequency discrimination in the California sea lion," J. Acoust. Soc. Am. **63**, 1591–1595.

Schusterman, R. J., Balliet, R. F., and Nixon, J. (**1972**). "Underwater audiogram of the California sea lion by the conditioned vocalization technique," J. Exp. Anal Behav. **17**, 339–350.

Schusterman, R. J., Balliet, R. F., and St. John, S. (**1970**). "Vocal displays under water by the gray seal, the harbor seal, and Stellar sea lion," Psychonomic Sci. **18**, 303–305.

Severinsen, T. (**1990**). "Effects of disturbance on marine mammals," in "Environmental Atlas Gipsdalen, Svalbard," edited by T. Severinsen and R. Hansson, Norwegian Polar Research Institute Report No. 66. Vol. **3**, pp. 41–63.

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.

Southall, B. L., Schusterman, R. J., and Kastak, D. (**2003**). "Auditory masking in three pinnipeds: Aerial critical ratios and direct critical bandwidth measurements," J. Acoust. Soc. Am. **114**, 1660–1666.

Terhune, J. M. (**1988**). "Detection thresholds of a harbor seal to repeated underwater high-frequency, short duration sinusoidal pulses," Can. J. Zool. **66**, 1578–1582.

Terhune, J. M. (**1989**). "Underwater click hearing thresholds of a harbor seal," Aquat. Mamm. **15**, 22–26.

Terhune, J. M., and Ronald, K. (**1972**). "The harp seal, *Pagophilus groenlandicus* (Erxleben,1777). III. The underwater audiogram," Can. J. Zool. **50**, 565–569.

Terhune, J. M., and Ronald, K. (**1975**). "Underwater hearing sensitivity of two ringed seals (*Pusa hispida*)," Can. J. Zool. **53**, 227–231.

Terhune, J., and Turnbull, S. (**1995**). "Variation in the psychometric functions and hearing thresholds of a harbour seal," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden), pp. 81–93.

Thomas, J., Moore, P., Withrow, R., and Stoermer, M. (**1990**). "Underwater audiogram of a Hawaiian monk seal (*Monachus schauinslandi*)," J. Acoust. Soc. Am. **87**, 417–420.

Trites, A. W., and Larkin, P. A. (**1996**). "Changes in abundance of Steller sea lions (*Eumetopias jubatus*) in Alaska from 1956 to 1992; How many were there?," Aquatic Mammals **22**, 153–166.

Turnbull, S. D., and Terhune, J. M. (**1990**). "White noise and pure tone

masking of pure tone thresholds of a harbour seal listening in air and underwater," Can. J. Zool. **68**, 2090–2097.

Turnbull, S. D., and Terhune, J. M. (**1994**). "Descending frequency swept tones have lower thresholds than ascending frequency swept tones for a harbor seal and human listeners," J. Acoust. Soc. Am. **96**, 2631–2636.

Wartzok, D., and Ketten, D. R. (**1999**). "Marine Mammal Sensory Systems," in *Biology of Marine Mammals*, edited by J. Reynolds and S. Rommel (Smithsonian Institution Press, Washington, D.C.), pp. 117–175.

Watson, Ch. S., and Gengel, R. W. (**1969**). "Signal duration and signal frequency in relation to auditory sensitivity," J. Acoust. Soc. Am. **46**, 989–997.

Zar, J. H. (**1984**). *Biostatistical Analysis*, 2nd ed., (Prentice-Hall International, Englewood Cliffs, NJ).

Types, distribution, and seasonal occurrence of sounds attributed to Bryde's whales (*Balaenoptera edeni*) recorded in the eastern tropical Pacific, 1999–2001

Sara L. Heimlich,[a] David K. Mellinger, and Sharon L. Nieukirk
Cooperative Institute for Marine Resources Studies, Oregon State University, and NOAA Pacific Marine Environmental Laboratory, 2030 SE Marine Science Drive, Newport, Oregon 97365

Christopher G. Fox
National Geophysical Data Center, National Oceanographic and Atmospheric Administration, 325 Broadway, E/GC, Boulder, Colorado 80305-3328

Vocalizations resembling known Bryde's whale sounds were recorded on autonomous hydrophones at seven sites in the eastern tropical Pacific. Five short (<3 s) low-frequency (<80 Hz) "phrase" types were observed. "Swept alternating tonal" phrases included a 37-Hz tone and often a 25–16-Hz downswept tone, while "nonswept alternating tonal" phrases had a predominant tone at 29 Hz and often additional tones at 16 and 47 Hz. Alternating tonal phrases were found in 79% of the total hours in which phrases were detected, and occurred primarily at the eastern hydrophone sites. "Burst-tonal" phrases included tones that were often preceded by a wideband burst of noise. The "low burst-tonal" phrase contained tones at 19 and 30 Hz, and was detected at five of the hydrophone sites. The "high burst-tonal" phrase included a 42-Hz tone and was observed only on the northwestern hydrophones. A single "harmonic tone" phrase type was observed that included a fundamental tone at 26 Hz and at least two harmonics; this phrase was observed exclusively at the eastern hydrophone stations. This opportunistic survey has shown that acoustics is an effective means of studying this poorly understood, pelagic balaenopterid. [DOI: 10.1121/1.1992674]

I. INTRODUCTION

Cetacean surveys in the eastern tropical Pacific (ETP) have typically been conducted visually (Wade and Gerrodette, 1993). In recent years, acoustic surveys have successfully examined the occurrence (Oleson *et al.*, 2003) and seasonal distributions (Stafford *et al.*, 1999b) of several balaenopterid species and populations in the region. Bryde's whales (*Balaenoptera edeni*) have a pan-tropical distribution (Leatherwood *et al.*, 1982; Reeves *et al.*, 2003) and are particularly abundant in the ETP east of 110° W and west of 150° W between 20° N and 10° S (Wade and Gerrodette, 1993). The Bryde's whale is not listed as endangered (IWC, 2000, 2002; NMFS, 2004), but the status of populations, and stocks and their numbers (Hilton-Taylor, 2000) is unclear. In the ETP, the International Whaling Commission (IWC) recognizes one stock of Bryde's whales, the cross-equatorial or "Peruvian" stock (Donovan, 1991; IWC, 2000, 2002), with a population estimate of approximately 13 000 animals (95% CI=8900–19 900; Wade and Gerrodette, 1993).

Although Bryde's whales have often been observed in this region, little was published about their vocalizations until Thompson *et al.* (1979) briefly described recordings made by Cummings in the 1960's. Cummings *et al.* (1986) gave a more detailed account, reporting that sounds recorded in the presence of Bryde's whales in the Sea of Cortez were very short (~0.4 s), low-frequency (~124 Hz) tonal sounds that were slightly frequency modulated upward or downward (~15 Hz). These sounds varied in amplitude with the whales' distance from the hydrophone. Edds *et al.* (1993) recorded sounds from a stranded juvenile in the Gulf of Mexico and an adult and adult–calf pair in the Sea of Cortez. In these recordings, sounds were somewhat longer in duration (~1.5 s) and included pulses, moans, and pulsed moans between 90 and 900 Hz. Recently, Oleson *et al.* (2003) examined acoustic recordings of Bryde's whale sounds collected opportunistically during visual surveys for cetaceans in the ETP. These authors reported six distinct types of vocalizations which were determined, by the directions of sightings and acoustic bearings, to be from Bryde's whales. These calls were described as ~3 s, low-frequency (<60 Hz) "tone bursts," and were typically longer and much lower in frequency than those previously reported for Bryde's whales.

Other low-frequency cetacean vocalizations have been reported in multiyear, wide-scale surveys conducted using hydrophone instruments deployed in the ETP (Stafford *et al.*, 1999a, b). These reports led us to believe that a similar effort to examine data from these hydrophones for Bryde's whale vocalizations might be fruitful, as the instruments were located in the central part of Bryde's whale range and were near areas where Bryde's whale sounds were recorded from ships (Oleson *et al.*, 2003). Here, we describe the basic characteristics, general distribution, and seasonal occurrence of several sounds found in these recordings that are believed to be from Bryde's whales.

---

[a]Electronic mail: sara.heimlich@oregonstate.edu
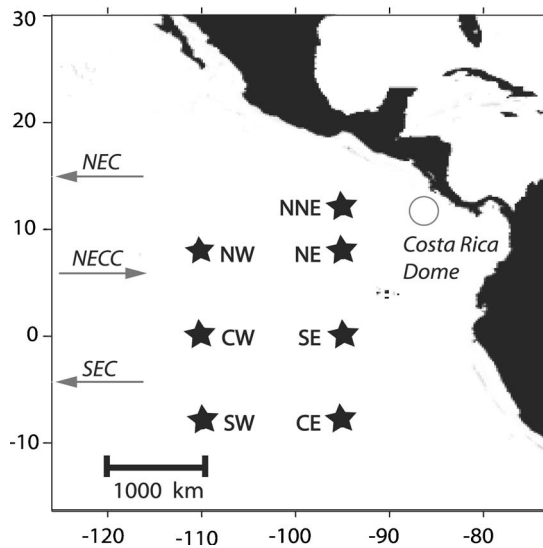
FIG. 1. Locations of autonomous hydrophones (stars) moored in the eastern tropical Pacific, and major oceanographic features including North Equatorial Current (NEC), North Equatorial Counter Current (NECC), and South Equatorial Current (SEC).

II. METHODS

An array of autonomous hydrophone instruments (Fox *et al.*, 2001) was deployed in the ETP, as part of a seismic experiment, within 12° N–8° S and 95°–110° W (Fig. 1) near the axis of the deep-sound channel (700–900 m). Water depth varied among sites, but was over 3400 m in all cases. These instruments recorded 12-bit samples continuously at a sample rate of 250 Hz. The antialiasing low-pass filter had a flat (±5 dB) frequency response in the 15–110-Hz band of interest. Five instruments operated from November 1999 to November 2000, and seven operated in the following year, November 2000 through November 2001.

Recordings were examined by visual inspection of continuously scrolling spectrograms made with custom software written in Interactive Data Language (IDL). Spectrogram parameters for this visual scan were 4.096-s (1024-sample) frame and FFT size, 76% overlap, and Hamming window, for a filter bandwidth of 0.99 Hz. During routine data processing, vocalizations were detected that did not have the identifying characteristics of the previously studied fin (*B. physalus*), blue (*B. musculus*) (Stafford *et al.*, 1999b), or humpback (*Megaptera novaeangliae*) whale sounds. These short-duration, low-frequency, swept tonal and tone-burst vocalizations were most similar in time/frequency characteristics to several sounds described by Oleson *et al.* (2003) for Bryde's whales.

A systematic survey for these sounds was conducted: on every fourth day of available data, all 24 hours were examined for vocalizations having these characteristics. Vocalizations were categorized into types by identifying specific parts distinctive to each type (Fig. 2, Table II). A group of such parts separated by at most a few seconds was considered a *phrase*, following the terminology of Payne and McVay (1971) for humpback whales and Mellinger and Clark (2003) for blue whales. The acoustic characteristics of phrases were examined using the analysis software OSPREY (Mellinger,

2000), with spectrogram parameters of 0.512-s (128-sample) frame size, 1.024-s (256-sample) FFT size, 87.5% overlap, and Hamming window, for a filter bandwidth of 7.9 Hz. These settings gave time and frequency resolutions similar to those used by Oleson *et al.* (personal communication, 2005). Acoustic characteristics of each phrase type were measured for a limited number of phrases with high signal-to-noise ratios. Measurements were made of beginning frequency, end frequency, and duration, as well as the interval between the end of each phrase and the start of the next. Statistics for these measurements are reported as mean plus or minus 1 standard deviation. Each type was named based on its characteristic structure rather than the nondescriptive nomenclature used previously (Oleson *et al.*, 2003). These measurements were used as the basis for comparing vocalizations between seasons, hydrophone sites, and previous acoustic surveys.

Presence or absence of each phrase type was noted for each hour of data from each site examined. The number of hours in which any Bryde's whale-like vocalization was observed was termed *hours with phrases* (HWP). A phrase type was considered present in an hour regardless of the number of times it occurred within that hour. For a given phrase type, the number of hours in which that type occurred was termed *phrase type hours* (PTH). The proportion of time in which any Bryde's whale phrase type was recorded at each of the seven hydrophone sites was determined by dividing HWP by the total number of data hours that were examined for that site. A monthly proportion for individual phrase types was determined by dividing PTH during which a vocalization type was seen in that month by the total number of data hours that were examined for that month.

III. RESULTS

Data from approximately 8 days per month were analyzed from five instruments that operated from November 1999 through November 2000, and seven instruments from November 2000 through November 2001. The NW and NNE instruments were not operational from November 1999 though November 2000. Vocalizations were detected in 2953 (11%) of the total 27 209 total hours examined. Of the five instruments that operated in both years, the CE hydrophone had the greatest number and variation of vocalizations, while the SW hydrophone had the fewest (Table I).

A. Phrase characteristics

Five distinct phrase types were observed (Fig. 2). All five were tonal in nature and of short duration, and showed similarities to Bryde's whale vocalizations described by Oleson *et al.* (2003), in particular their types Be1, Be2, and Be3 (Table II). Similarities included the nature of the sounds (tonal, sometimes with slight upward or downward frequency modulation), duration, approximate frequency band, and time/frequency structure as described below.

Each phrase type consisted of several parts, some of which might be absent in any given phrase (Fig. 2 and Table II). All phrase types had a tonal component, the *predominant tone*, which always occurred at a frequency that was charac-
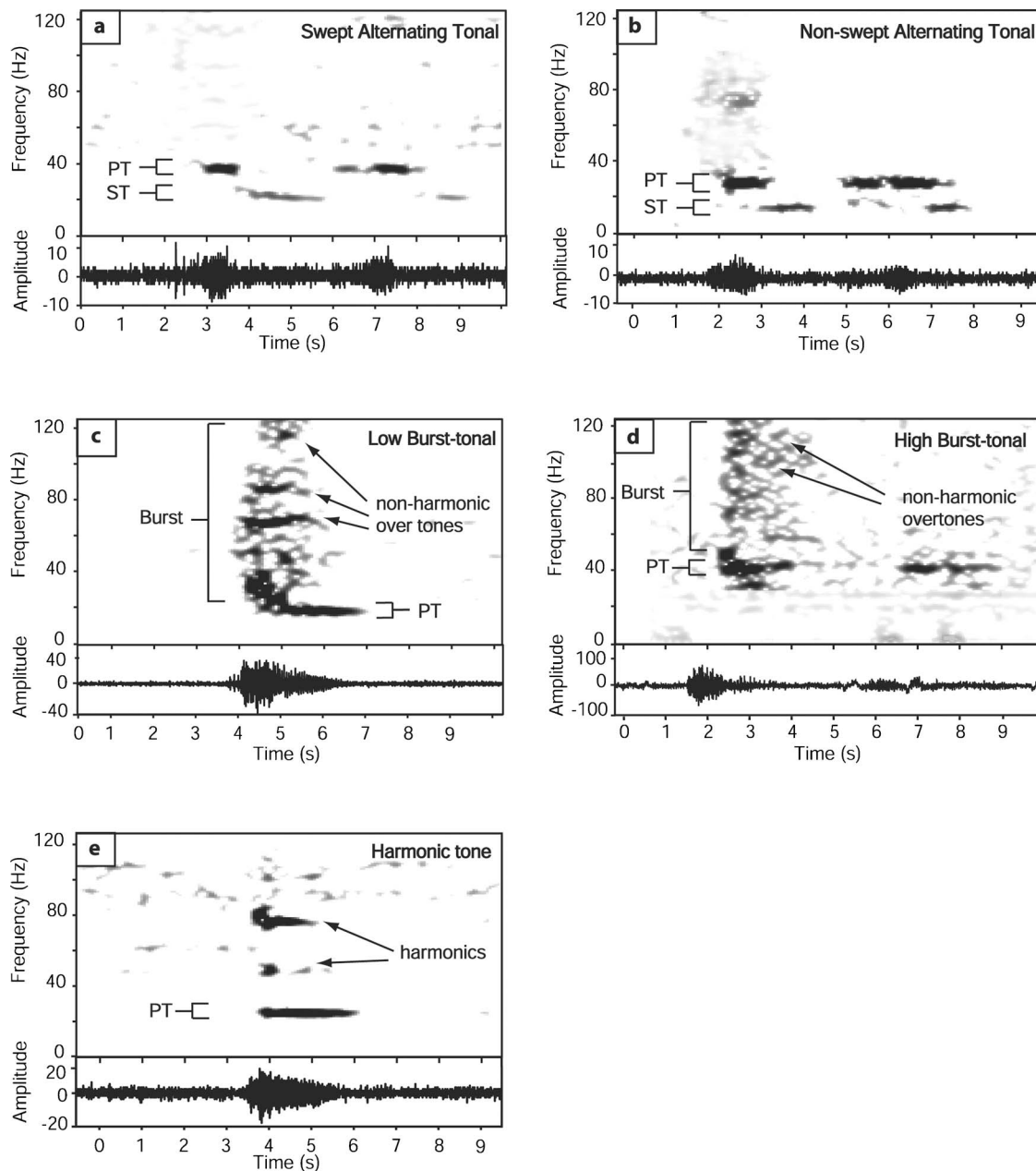
FIG. 2. Spectrograms and time series of five phrase types attributed to Bryde's whales in the eastern tropical Pacific. PT=predominant tone, ST= secondary tone. (a) Typical swept alternating tonal phrase showing characteristic repetitive structure. (b) Typical nonswept alternating tonal phrase showing characteristic repetitive parts. (c) Typical low burst-tonal phrase showing characteristic nonrepetitive parts. (d) Typical high burst-tonal phrase illustrating burst, noise, and repetitive parts. (e) Typical harmonic tone phrase showing fundamental and harmonics. Spectrogram parameters: 0.512-s (128-sample) frame size, 1.024-s (256 sample) FFT size, 87.5% overlap, and Hamming window, for a filter bandwidth of 7.9 Hz.

teristic for the phrase type. As this tone was usually the highest amplitude part of the phrase, it made the phrase type readily identifiable. Each phrase also had *secondary tones*, which did not always occur, at frequencies markedly above and below the predominant tone (Fig. 2).

The structures of the five phrase types fell into three general categories. *Alternating tonal phrases* had a constant, low-frequency tone (the predominant tone for this category) which temporally overlapped an even lower tone of lesser amplitude [Figs. 2(a) and 2(b)]. These upper-plus-lower tone pairs were often repeated, with the tones thus appearing as alternating parts. *Burst-tonal phrases* had a burst of wideband noise and, at approximately the same instant, several

overlapping, frequency-modulated, nonharmonically related tones; these were quickly followed by a distinct, high-amplitude lower-frequency tone (the predominant tone for this category) that temporally overlapped the higher tones [Figs. 2(c) and 2(d)]. *Harmonic tone phrases* had a fundamental tone (the predominant tone for this phrase type) and one or more harmonics [Fig. 2(e)].

The lower-frequency tones of alternating tonal and burst-tonal phrases sometimes occurred repeatedly, at intervals of 3–5 s. In alternating tonal phrases, repetition of parts was common and could occur at the frequencies of both the predominant tone and the lower, secondary tone [Figs. 2(a) and 2(b)], while in burst-tonal phrases, repetition was infre-

TABLE I. Summary statistics for the seven hydrophone sites surveyed in the eastern tropical Pacific. Hours with phrases (HWP) values represent the number of hours in which any Bryde's whale vocalization of that type was observed. "Rare" phrase types were those observed 3% or less.

Site	N hrs examined	Hours w/phrases (HWP)	% total HWP	Phrase type observed
NNE	2322	54	1.8	Harmonic tone
NW	2420	196	6.6	High burst-tonal (rare: Low burst-tonal)
NE	4442	355	12.0	Swept alternating tonal Low burst-tonal Harmonic tone (rare: Nonswept alternating tonal)
CW	4541	82	2.8	Low burst-tonal
CE	4553	1968	66.6	Swept alternating tonal Nonswept alternating tonal Low burst-tonal (rare: Harmonic tone)
SW	4484	71	2.4	Low burst-tonal (rare: Swept alternating tonal) (rare: Nonswept alternating tonal)
SE	4445	227	7.7	Nonswept alternating tonal (rare: Swept alternating tonal) (rare: Harmonic tone)
Total	27 209	2953	100	5 types

TABLE II. Comparison of summary statistics for Bryde's whale vocalizations reported by Oleson *et al.* (2003) and those reported in this paper (**in bold**). Mean, standard deviation, range, and percent occurrence within that phrase type are reported for each of the parts that are characteristic of each phrase type described in this manuscript. An asterisk (*) indicates a harmonic of the fundamental. Values in italics are phrase durations for cases when all parts were observed together.

Phrase type	Geographic region	N measured	Frequency (Hz)	% occurrence	Duration (s)	Intercall or Interphrase interval (min)	Overtones	% total within region
Be1	ET	37	21.2 (20.0–23.7)		2.7 (1.1–6.7)	1.2 (0.2–4.4)	No	27.4
Swept alternating tonal	**ETP**	**418**	**36.7±0.6 (33.5–39.0)** **21.7±1.1 (14.8–26.0)**[a]	**100** **68**	**1.1±0.5 (0.3–3.0)** **1.8±0.6 (0.4–3.2)** *2.0±1.1 (0.3–5.4)*	**2.9±1.0 (1.0–21.8)**	**No**	**46.3**
Nonswept alternating tonal	**ETP**	**302**	**28.6±0.8 (24.6–32.3)** **15.7±1.5 (13.5–20.7)** **46.7±3.3 (37.1–53.9)**	**100** **48** **41**	**1.1±0.5 (0.30–3.4)** **1.4±0.6 (0.3–3.2)** **1.4±0.5 (0.3–3.0)** *2.0±0.9 (0.3–5.6)*	**3.0±1.9 (1.2–12.6)**	**No**	**33.1**
Be2	ETP	36	41.7 (40.7–43.3)		1.3 (0.5–3.0)	2.9 (0.3–7.2)	Harmonics	26.7
High burst–tonal	**ETP**	**128**	**41.7±1.0 (39.9–46.6)** **31.0±1.2 (24.8–37.9)** **49.4±1.1 (47.5–56.4)**	**100** **58** **38**	**1.0±0.5 (0.3–2.5)** **0.9±0.3 (0.4–1.8)** **0.5±0.1 (0.3–1.0)** *1.0±0.5 (0.3–2.8)*	**3.9±2.8 (1.4–16.1)**	**Nonharmonically related tones**	**5.9**
Low burst–tonal	**ETP**	**140**	**18.5±0.8 (16.8–26.7)** **29.8±2.5 (24.1–36.1)** **67.2±1.6 (62.6–78.8)**	**100** **100** **78**	**1.4±0.5 (0.4–2.8)** **1.0±0.3 (0.3–2.5)** **1.4±0.3 (0.4–2.0)** *2.1±0.8 (0.4–4.0)*	**5.8±4.2 (1.7–33.8)**	**Nonharmonically related tones**	**6.5**
Be3	ETP	18	25.6 (24.0–26.9)		1.7 (1.0–4.0)	2.3 (0.4–8.6)	Harmonics	13.3
Harmonic tone	**ETP**	**114**	**25.4±0.6 (20.7–25.8)** ***77.3±1.1 (75.1–79.6)** ***51.3±4.41 (46.8–66.3)**	**100** **47** **24**	**2.1±0.5 (0.5–3.2)** **1.2±0.5 (0.4–2.3)** **1.6±0.3 (0.5–2.2)** *2.3±0.8 (0.5–7.0)*	**6.3±6.1 (2.00–44.3)**	**Harmonics**	**8.2**
Be7	Caribbean	51	44.3 (43.7–48.7)		1.6 (1.0–4.0)	2.8 (0.4–7.1)	Harmonics	100
Be8a	Japan	18	44.6 (43.0–48.0)		0.35 (0.25–0.6)	0.05 (0.03–0.09)	Harmonics	81.8

[a]The frequency values of the swept tone at 21.7 Hz are for the mean of highest and lowest frequencies in the tone.

quent and, when it did occur, was at only the predominant tone's frequency [Fig. 2(d)]. Repetition was believed not to be a result of multipath propagation because the interval between the upper two tones was sometimes different than the interval between the lower two tones; a repetition due to multipath propagation would delay the two tones equally. In addition, the repetition periods of 3.3–4.3 s are longer than usually seen for multipath effects. These tones occurred with decreasing amplitude on each repetition. When repetitions were not observed, it was not merely because the later, repeating tones were below the level of background noise: phrases were observed in which the level of the initial upper tone (the predominant tone) was far enough (10–12 dB) above the level of background noise that any later repetitions would have been visible in a spectrogram.

B. Phrase types

The *swept alternating tonal* phrase type [Fig. 2(a)] had a predominant tone averaging 36.7 Hz, which always occurred, and a lower secondary tone of lesser amplitude that occurred in 68% of the 418 measured phrases. The secondary tone swept down 6.4±1.9 Hz from an average frequency of 24.9 Hz to an average frequency of 18.5 Hz. Phrase duration averaged 2.0 s and the interphrase interval averaged 2.9 min. Repetitive parts occurred in 47% of the measured swept alternating tonal phrases. The characteristics of this type most closely resembled those of the Be1 call described by Oleson *et al.* (2003) (Table II).

The *nonswept alternating tonal* phrase type [Fig. 2(b)] had a predominant tone averaging 28.6 Hz that always occurred, and a lower secondary tone averaging 15.7 Hz that occurred in 48% of the 302 measured phrases and began approximately 0.7 s later. A third tone averaging 46.6 Hz occurred in 41% of the phrases measured. Phrase duration averaged 2.0 s and the interphrase interval averaged 3.0 min. (Table II). Repetitive elements occurred in 76% of the measured nonswept alternating tonal phrases.

Both alternating tonal phrase types were typically recorded in extended, irregular *series* over spans of several hours. Within a series, phrases had consistent intervals between the onsets of their predominant and lower tones, and showed consistent amplitude relationships between these tones. On rare occasions, phrases of both types were observed with higher-frequency energy bands above the predominant tone, similar to the structure of the tones seen in the burst-tonal phrases described below. The swept alternating tonal and the nonswept alternating tonal phrases occasionally occurred during the same hour, and occasionally even overlapped temporally. Intervals less than 0.5 s were often observed between phrase onsets when they overlapped.

The *low burst-tonal* phrase type [Fig. 2(c)] had two distinct parts that always occurred: a constant-frequency tone that averaged 18.5 Hz, and a frequency-modulated (FM) tone that averaged 29.8 Hz. A third FM tone at 67.2 Hz was seen in over half (78%) of the 140 measured phrases (Table II). Phrase duration averaged 2.1 s and interphrase interval averaged 5.8 s. Only 22% of the measured low burst-tonal phrases occurred with repetitive elements.

The *high burst-tonal* phrase type [Fig. 2(d)] had a predominant FM tone that occurred at an average of 41.7 Hz. A second, lower FM tone that averaged 31.0 Hz appeared in 50% of the 128 measured phrases. A higher third tone that averaged 49.4 Hz occurred in 38% of the measured phrases. Phrase duration averaged 1.0 s and interphrase interval 3.9 s (Table II). Just 22% of the measured high burst-tonal phrases occurred with repetitive elements. The characteristics of the high burst-tonal phrase type most closely resembled those of the Be2 call type described by Oleson *et al.* (2003) (Table II).

There was only one *harmonic tone* phrase type [Fig. 2(e)]. This had a fundamental frequency that averaged 25.4 Hz. The second harmonic was observed in only 24%, and the third harmonic in 47%, of the 114 measured harmonic tone phrases (Table II). Phrase duration averaged 2.3 s and the interphrase interval averaged 6.3 s Of the measured phrases, 37% had repetitions of the fundamental or harmonics. This phrase type most closely resembled the Be3 call type described by Oleson *et al.* (2003) (Table II).

Like the two alternating tonal phrase types, the burst-tonal and harmonic tone phrase types occurred in series that were detected over periods of several hours. Within a series, phrases had consistent intervals between the onsets of successive repetitive tones and showed consistent amplitude relationships between upper and lower tones.

C. Geographic and seasonal occurrence

Nonswept and swept alternating tonal phrase types were the most abundant of the five types, comprising 1535 (46.3%) and 1097 (33.1%), respectively, of the total of 3315 PTH. Both types occurred almost exclusively at the eastern sites, predominantly at the CE site, throughout the entire study period (Table I). Whereas swept alternating tonal phrases also occurred moderately to the north (8° N) and very little to the south (8° S), nonswept alternating tonal phrases occurred rarely in the northern regions at any time and only moderately to the south in the second year, predominantly at the SE site (Fig. 3). Both alternating tonal phrases also showed the strongest seasonality, with more vocalizing recorded from roughly August through February than in the remainder of the year (Fig. 3).

The low burst-tonal, high burst-tonal, and harmonic tone phrase types occurred least frequently of the five phrase types, comprising 215 (6.5%), 195 (5.9%), and 273 (8.2%), respectively, of the total of 3315 PTH. However, low burst-tonal phrases were the most widely distributed of all phrase types throughout the study area, occurring at five of the seven sites (Table I). They occurred principally at 0° latitude (CE and CW), and were the only type of phrases to occur at CW. The high burst-tonal phrase was not observed at any of the five hydrophone sites which operated during the first year, and exclusively at the NW site during the second year (Table I); this instrument did not operate the first year. Harmonic tone phrases were observed exclusively at the eastern (95° E) locations in both study years, predominantly at the NE and NNE sites (Table I). While none of these phrase types showed a consistent seasonal pattern, low burst-tonal
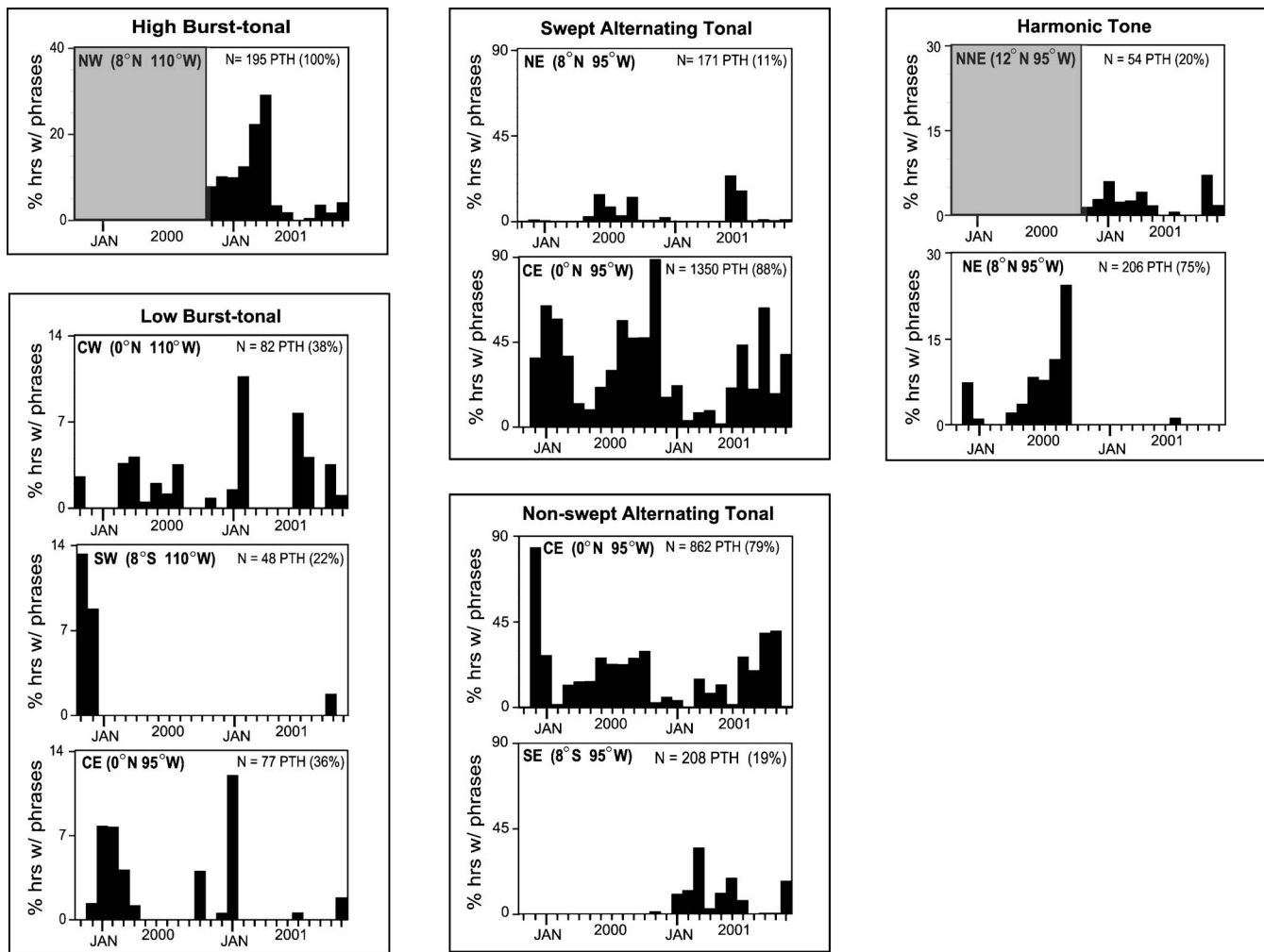
FIG. 3. Seasonal occurrence and geographic distribution of the five phrases identified in this study at locations where occurrence was greater than 3%. Note the *Y*-axis scaling varies between phrase types. Shaded areas indicate periods with no data. *X* axis represents November 1999 through November 2001. PTH=Phrase Type Hours.

phrases occurred in a somewhat bimodal pattern, and high burst-tonal phrases and harmonic tone phrases had somewhat unimodal occurrence patterns (Fig. 3).

Of the five phrase types, the harmonic tone and non-swept alternating tonal phrase types were the only ones to evidence any pronounced year-to-year change during the sampled time period: the harmonic tone phrase at the NE site occurred extensively in the first year but almost not at all in the second year, while the nonswept alternating tonal phrase at the NE and SE sites occurred very little in the first year but moderately often in the second year (Fig. 3).

IV. DISCUSSION

The vocalizations reported in this paper are thought to be from Bryde's whales because of their similarity to the vocalizations described by Oleson *et al.* (2003) from the ETP and their dissimilarity to known sounds from other species in the ETP (Stafford *et al.*, 1999a, b). We further believe them to be produced by balaenopterid whales because of the hier-archical nature of these vocalizations, as seen in humpback (Payne and McVay, 1971), fin (Watkins *et al.*, 1987), blue

(McDonald *et al.*, 1995; Mellinger and Clark, 2003; Stafford *et al.*, 2004), and minke (Mellinger *et al.*, 2000) whales.

In these autonomous hydrophone data, vocalizations that we attribute to Bryde's whales were found in approximately 11% of the hours examined, with detections in over 90% of hours in some months. It was not surprising to find them in these ETP data, given that Bryde's whales are the most com-monly sighted whale in this area (Wade and Gerrodette, 1993) and that vocalizing Bryde's whales have been posi-tively identified there (Oleson *et al.*, 2003), but vocalizations occurred more frequently than expected. These types of vo-calizations may have been overlooked in previous studies because, unlike other baleen whale sounds, Bryde's whale vocalizations can be quite short, usually less than 3 s, and are typically not very loud relative to background noise. These factors can make detection of these vocalizations somewhat difficult.

The five vocalization types we identified as Bryde's whale phrases showed different geographic distributions. Three of the five types (harmonic tone, swept alternating tonal, and nonswept alternating tonal) were observed pre-dominantly in the eastern part of the study region, one (high

burst-tonal) occurred exclusively at the NW site, and one (low burst-tonal) was observed predominantly at the Equator. In addition, the nonswept alternating tonal and swept alternating tonal types exhibited some degree of spatial and temporal correlation. The five types occurred either predominantly at or south of the Equator (low burst-tonal, even alternating tonal and swept alternating tonal types), or predominantly at or north of 8° N (harmonic tone and high burst-tonal types); all types were seen further north (8° and 12° N) than previously reported. Similar north–south and east–west divisions were observed in blue whale vocalizations recorded in the ETP (Stafford, *et al.* 1999b). Such distributions may be influenced by the Intertropical Convergence Zone, which in the ETP tends to be north of the Equator in most years, and the South and North Equatorial currents, the North Equatorial Counter Current, and the Costa Rica Dome (Fig. 1).

Reilly (1990) showed seasonal changes in the historical distribution and habitat differences among dolphins in the ETP, and attributed this to the shape and location of the "biotic province" created by these oceanographic features. Reilly and Fiedler (1994) further suggested these features have major consequences for the ecology of pelagic predators. This would include Bryde's whales and their seasonal, temporal, and geographic distributions in this region.

Although the somewhat bimodal pattern of temporal occurrence in most of the vocalization types suggests seasonal variations may occur, seasonal patterns in these acoustic data tended to vary with vocalization type, hydrophone location, and year. Such patterns may be due to changes in the vocal activity of individuals, the distribution of whales in the area, or oceanographic influences in the ETP. Until we know the meaning of these phrase types, how they relate to other phrase types, how vocalizations relate to behavior, and how vocalization rates are related to the number of animals present, interpreting these seasonal patterns will be difficult at best. A survey covering a larger geographic area or examination of a longer time series may reveal complementary changes in vocalization rates in different areas in different seasons.

Previous opportunistic recordings collected in the ETP identified six different vocalization types (Oleson *et al.*, 2003), of which only three appear similar to those we detected, despite the longer period of data collection in this study (2 years of continuous recording vs opportunistic recording in July and December). Part of this discrepancy may be due to fundamental differences between these two experiments. This dataset's frequency range extended only to approximately 110 Hz, therefore some of the higher frequencies of some of the Oleson *et al.* call types may not have been detected.

Regardless of the differences in data collection among studies, at least eight different vocalization types have so far been attributed to Bryde's whales in the ETP. Why is there such variability in a relatively small area? Oleson *et al.* (2003) observed a relationship between call type and group size or behavior, which could explain some variability. Another explanation is that animals from different stocks or populations are present in the area seasonally. Like some

other balaenopterid whales (Winn *et al.*, 1981; Thompson *et al.*, 1992; Stafford, *et al.* 1999b), Bryde's whales are known to make different sounds in different parts of the world (Oleson *et al.*, 2003). However, correlating vocalization types with geographic or stock parameters is difficult because of a lack of general understanding of the behavioral and ecological context of vocalizations, and the confusion regarding Bryde's whale taxonomy and populations. The IWC recognizes seven stocks in the ETP and only one species of Bryde"s whale worldwide (Donovan, 1991). Then again, others have argued for recognizing two species based on morphology (Rice, 1998) or three based on genetics (Wada *et al.*, 2003). Furthermore, some have reported inshore and offshore forms (Best, 1977). Until these issues are resolved, the relationship between taxonomic and management units and variation in vocal behavior will remain unclear.

The burst-tonal and harmonic tone phrases found in the ETP in this study have characteristics in common with the Be7 and Be8a Bryde's whale vocalizations described from the southern Caribbean and Japan (Table II, this paper; Oleson *et al.*, 2003). Specifically, they share the same approximate duration and the structure of nonharmonically related overtones. While the relationship between the geographic differences in vocalization type occurrence and biological factors is unclear, it seems evident that Bryde's whale sounds have some universal characteristics that make them readily identifiable to species. These include a predominant tone having very little frequency modulation, a duration of $1-3$ s, and a frequency below approximately 120 Hz (Thompson *et al.*, 1979; Cummings *et al.*, 1986; Edds *et al.*, 1993; Oleson *et al.*, 2003).

If Bryde's whale stocks do have distinctive acoustic behaviors, further examination of acoustic data, both from autonomous hydrophone arrays and from towed array surveys, should prove useful in interpretation of the regional differences and seasonal distributions indicated in this paper. Combining acoustic information with genetic data will aid in defining population structure and distribution of Bryde's whales in the ETP, and perhaps elsewhere in the Pacific.

Best, P. B. (**1977**). "Two allopatric forms of Bryde's whale off South Africa," Rep. Int. Whal. Commn. (Special Issue) **1**, 10–38.

Cummings, W. C., Thompson, P. O., and Ha, S. J. (**1986**). "Sounds from Bryde, *Balaenoptera edeni*, and finback, *B. physalus*, whales in the Gulf of California," Fish. Bull. **84**, 359–370.

Donovan, G. P. (**1991**). "A review of IWC stock boundaries," Rep. Int. Whal. Commn. (Special Issue) **13**, 39–68.

Edds, P. L., Odell, D. K., and Tershy, B. R. (**1993**). "Vocalization of a captive juvenile and free-ranging adult-calf pairs of Bryde's whales, *Balaenoptera edeni*," Marine Mammal Sci. **9**, 269–284.

Fox, C. G., Matsumoto, H., and Lau, T.-K. A. (**2001**). "Monitoring Pacific Ocean seismicity from an autonomous hydrophone array," J. Geophys. Res. **106**(B3), 4183–4206.

Hilton-Taylor, C. (**2000**). *IUCN Red List of Threatened Species* (International Union for the Conservation of Nature and Natural Resources, Gland).

IWC (**2000**). "Report of the sub-committee on the revised management procedure," J. Cetacean Res. Manage. **4** (Suppl.), 79–124.

IWC (**2002**). Report of the scientific committee J. Cetacean Res. Manage. **4** (Suppl.), 1–92.

Leatherwood, S., Reeves, R. R., Perrin, W. F., and Evans, W. E. (**1982**). "Whales, dolphins, and porpoises of the eastern North Pacific and adjacent Arctic waters: A guide to their identification," NOAA Technical Rept. NMFS Circular 444.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). "Blue and fin whales observed on a seafloor array in the northeast Pacific," J. Acoust. Soc. Am. **98**, 712–721.

Mellinger, D. K. (**2000**). MOBYSOFT. National Oceanographic and Atmospheric Administration web site http://cetus.pmel.noaa.gov/cgi-bin/MobySoft.pl

Mellinger, D. K., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.

Mellinger, D. K., Carson, C. D., and Clark, C. W. (**2000**). "Characteristics of minke whale (*Balaenoptera acutorostrata*) pulse trains recorded near Puerto Rico," Marine Mammal Sci. **16**, 739–756.

NMFS (**2004**). "Species listed under the Endangered Species Act of 1973," National Marine Fisheries Service web site http://www.nmfs.noaa.gov/prot_res/species/ESA_species.html

Oleson, E. M., Barlow, J., Gordon, J., Rankin, S., and Hildebrand, J. A. (**2003**). "Low frequency calls of Bryde's whales," Marine Mammal Sci. **19**, 160–172.

Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 587–597.

Reeves, R. R., Smith, B. D., Crespo, E. A., and Notarbartolo di Sciara, G., (eds.). (**2003**). *Dolphins, Whales and Porpoises: 2002–2010 Conservation Action Plan for the World's Cetaceans* (IUCN/SSC Cetacean Specialist Group, IUCN, Gland).

Rice, D. W. (**1998**). "Marine mammals of the world—Systematics and distribution," Soc. Mar. Mamm. Special Pub. **4**, 231 pp.

Reilly, S. B. (**1990**). "Seasonal changes in distribution and habitat differences among dolphins in the eastern tropical Pacific," Mar. Ecol.: Prog. Ser. **66**, 1–11.

Reilly, S. B. and Fiedler, P. C. (**1994**). "Interannual variability of dolphin habitats in the eastern tropical Pacific 1: Research vessel surveys, 1986–1990," Fish. Bull. **92**, 434–450.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999a**). "An acoustic link between blue whales in the eastern tropical Pacific and the northeast Pacific," Marine Mammal Sci. **15**, 1258–1268.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999b**). "Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific," J. Acoust. Soc. Am. **106**, 3687–3698.

Stafford, K. M., Bohnenstiehl, D. R., Tolstoy, M., Chapp, E., Mellinger, D. K., and Moore, S. E. (**2004**). "Antarctic-type blue whale calls recorded at low latitudes in the Indian and the eastern Pacific Oceans," Deep-Sea Res., Part I **51**, 1337–1346.

Thompson, T. J., Winn, H. E., and Perkins, P. J. (**1979**). "Mysticete sounds," in *Behavior of Marine Mammals: Current Perspectives in Research*, edited by H. E. Winn and B. L. Olla (Plenum, New York), pp. 403–431.

Thompson, P. O., Findley, L. T., and Vidal, O. (**1992**). "The 20-Hz pulses and other vocalizations of fin whales, *Balaenoptera physalus*, in the Gulf of California, Mexico," J. Acoust. Soc. Am. **92**, 3051–3057.

Wada, S., Oishi, M., and Yamada, T. K. (**2003**). "A newly discovered species of living baleen whale," Nature (London) **426**, 278–281.

Wade, P. R., and Gerrodette, T. (**1993**). "Estimates of cetacean abundance and distribution in the Eastern Tropical Pacific," Rep. Int. Whal. Commn. **43**, 477–493.

Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signals of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.

Winn, H. E., Thompson, T. J., Cummings, W. C., Hain, J., Hudnall, J., Hays, H., and Steiner, W. W. (**1981**). "Song of the humpback whale—population comparisons," Behav. Ecol. Sociobiol. **8**, 41–46.

Session 1aAAa

Architectural Acoustics and ASA Committee on Standards: Comparison of U.S. and International Standards in Architectural Acoustics

David E. Marsh, Chair

Pelton Marsh Kinsella, 1420 W. Mockingbird Ln., Dallas, TX 75247-4932

Chair's Introduction—7:30

Invited Papers

7:35

1aAAa1. Comparison of acoustical standards and guidelines used in the USA, UK and Europe. Edward Adamczyk (PMK Intl., Saeed Tower II, Sh. Zayed Rd., #204, P.O. Box 58123, Dubai, UAE, edward.a@pmkconsultants.ae) and David E. Marsh (Pelton Marsh Kinsella, Dallas, TX 75247)

A comparison of acoustical standards and guidelines used in the USA, UK and some other countries of Europe is presented with examples of applications to projects in the UAE and other Middle Eastern countries. The paper focuses on similarities and differences in acoustical design criteria adopted for building services noise, vibration, airborne and impact sound isolation, environmental noise, room acoustics, and electroacoustics. An example is office speech privacy where the UK uses a different parameter for sound insulation/isolation compared to the USA. Some of the criteria used in different countries are quite onerous and not commercially feasible. This is particularly relevant to hospital and school designs in the UK where large PFI/PPP funded projects require compliance with many standards. Compliance with UK Building Regulations in residential buildings has now become a legal requirement with testing at completion to confirm compliance, as is the case in some other European countries. Alternatively, compliance may require use of a robust standard detail that has been proven by multiple testing without a single failure. The paper concludes with case studies where the financial and legal impact of complying with these standards was a key issue, particularly in the design of mixed-use developments.

7:55

1aAAa2. Comparison summary: Various countries' standards for classroom acoustics. Jack B. Evans (JEAcoust., 1705 West Koenig Ln., Austin, TX 78756)

A comparative summary presentation of many countries' national acoustical standards for classroom acoustics will be presented. Facility renovation or new construction is subject to standards or regulations that control reverberation decay time, sound isolation between interior spaces, intrusive noise from the exterior environment, background noise from building systems and/or user installed classroom equipment. Child learner physical conditions are reviewed, based on many researchers' published results in North America and Europe, in regard to the special building acoustical requirements needed to achieve good classroom environments. In response to the authors' solicitations, colleagues from many nations provided facts, opinions, and reference for their own countries' standards, which are presented in categorical matrix format, including ANSI S12.60-2002, the relatively new American classroom acoustics standard. Summary results from I-INCE TC4 working group study on international standards are also incorporated. [This presentation is derived from the international classroom standards portion (only) of a paper originally written for the 11th International Meeting on Low Frequency Noise and Vibration and its Control, Maastricht, The Netherlands in 2004 with recent updates.]

8:15

1aAAa3. International Organization for Standardization (ISO) standards as viewed from a U.S. perspective. Kenneth P. Roy (Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604)

The U.S. is a participating member in the development of ISO standards through the American National Standards Institute (ANSI). Specific ANSI participation relative to standards on architectural acoustics is within ISO TC43 SC2, which deals with building acoustics, as opposed to noise for instance. The actual standards development and voting responsibilities rest with the ANSI Technical Advisory Group (TAG) to ISO TC43 SC2, which is delegated to ASTM (American Standards for Testing and Materials) committee E 33.06 International Standards. This paper will present the process whereby the entire American continent is involved in the development of ISO standards, since ASTM is an international standard organization, whereas the U.S. voting is of course solely an ANSI responsibility. Both current and in-development ISO standards will be discussed.

8:35

1aAAa4. Harmonizing the American National Standards Institute (ANSI) and International Electrotechnical Commission (IEC) sound level meter specification standards. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Turnpike, Vernon, CT 06066)

It has been over 20 years since the last major revision of the American National Standard Specification for Sound Level Meters, S1.4-1983 (R2002). This is the most extensively referenced ANSI acoustical standard, with wide application to architectural acoustics, soundscapes, environmental noise, and hearing conservation. A new revision is proposed that conforms as closely as possible to the recently developed International Electrotechnical Commission Standard for Electroacoustics-Sound Level Meters Part 1: Specifications, IEC 61672-1 (2002-05). This proposed revision represents an improvement over ANSI S1.4-1983, particularly in its frequency response tolerance limits at high frequencies, and impulse (toneburst) response. Also, it incorporates the high-frequency electrical design goals and tolerance limits for A-weighted response that were specified in Amendment S1.4A-1985, and it maintains the more restrictive low-frequency tolerance limits of the earlier (1983) specification, compared to IEC 61672-1. The proposed revision maintains the laboratory precision Type 0 specification, which is absent in IEC 61672-1, while incorporating the specifications for integrating-averaging sound level meters previously given in ANSI S1.43-1997. Sound level meters which meet the Type 1 and Type 2 specification requirements, respectively, in this S1.4 revision will also meet the Class 1 and Class 2 specification requirements of 61672-1.

8:55

1aAAa5. Mexican laws and standards on acoustics. Omar Saad (Universidad Iberoamericana, Dept. de Arquitectura y Urbanismo, Prol. Paseo de la Reforma 880, Santa Fe 01210, Mexico, saadacustica@prodigy.net.mx)

We explore the current state of Mexican standards and laws on acoustics, current efforts on the way to improve such standard and international standards that society has adopted due to the limited Mexican laws on noise.

Contributed Paper

9:15

1aAAa6. Analysis of the Mexican standard on optimal reverberation times according to enclosure function. Sergio Beristain (IMA, ESIME, IPN. P.O. Box 12-1022, Narvarte, 03020, Mexico, D. F., Mexico, sberista@hotmail.com)

The Mexican Standard No. NOM-C-211-1977 on optimal reverberation times according to the enclosure use, published in December 1977, and without any substantial modification so far, establishes optimum re-

verberation times for different room applications, which by virtue of time reverberation measurements described in the Mexican Standard No. NOM-C-147-1977 on field measurements of reverberation time in enclosures, and the Eyring reverberation time equation, requires given acoustical conditions within rooms, and sets tolerance limits, so this is a method for testing and modification of already-built rooms' response. The full procedure is presented in this paper, so actually two Mexican standards are being discussed here, in order to be able to compare with other standards and methods employed elsewhere.

MONDAY MORNING, 17 OCTOBER 2005　　　　　　　　　SALON F, 9:45 A.M. TO 12:20 P.M.

Session 1aAAb

Architectural Acoustics, NOISE-CON and Noise: Indoor Noise Criteria

Erica E. Bowden, Chair

Univ. of Nebraska, Architectural Engineering, 245 Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182

Chair's Introduction—9:45

Invited Papers

9:50

1aAAb1. Relationship between noise criteria and room criteria ratings. Gregory C. Tocci (Cavanaugh Tocci Assoc., Inc., 327 F Boston Post Rd., Sudbury, MA 01776)

The September 2000 issue of Noise/News International featured a paper entitled "Room Noise Criteria—The State of the Art in the Year 2000" by Gregory C. Tocci. The paper in part investigated the relationship between NC (tangency), NCB, and RC ratings for a series of 268 spectra measured in building spaces during the course of Cavanaugh Tocci Associates, Inc. consulting work. These spectra ranged from uniform to widely varying in shape. This presentation extends this work to document the relationship between RC and NC tangency ratings for these 268 spectra, which was not included in the original 2000 NNI paper.

1aAAb2. Indoor sound criteria according to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)—An introduction. Ted N. Carnes (PMK, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

ASHRAE TC-2.6 Sound and Vibration Control technical committee has been activity involved with development, refinement, and promotion of Indoor Sound Criteria. ASHRAE, its members, and those in the HVAC industry are the largest users of Indoor Sound Criteria. As part of an effort to educate engineers and others particularly in the building design industry, an unofficial introduction to ASHRAE viewpoints and guidelines on Indoor Sound Criteria were developed into a presentation. Suggested use of dBA, NC, RC, NCB, and RC Mk II criteria descriptors have been included. The use of sound criteria in the Design, Commissioning, and Trouble-shooting or Diagnostics of background sound or noise due to HVAC systems were considered. Criteria descriptors that promote the awareness of Sound Quality have been an essential aspect of ASHRAE research. One aspect of the ASHRAE criteria descriptor has been to provide suggested means to determine and quantify Noise Induced Vibration (NIV) of building systems. This type of noise problem has been well identified by the use of RC Mk II, RC, and NCB sound criteria methodology. Actual examples of NIV will be described in the presentation.

1aAAb3. Comments on draft ANSI S12.2, American National standard on criteria for evaluating room noise. Leo L. Beranek (975 Memorial Dr., #804, Cambridge, MA 02138, beranekleo@ieee.org)

In 1995, American National Standard S12.2-1995 on room noise criteria was issued. Two methods for specifying and evaluating room noise were presented, the NCB noise-criterion curves [J. Acoust. Soc. Am. **86**, 650–664 (1989)] and the RC curves [Noise Control Eng. J. **45**, 243–250 (1997)]. The NCB curves were derived from the characteristics of hearing with emphasis on speech interference level and loudness level, while the RC curves were derived from experimental studies in offices where the HVAC noise was low enough that there were no complaints from room occupants. The NCB curves were applicable to rooms ranging from concert halls to factory spaces, while the RC curves were largely restricted to office, hospital, and residential spaces. The two sets were incompatible at low frequencies and low sound levels. The NCB curves assumed that the HVAC noise was free of surging and strong turbulence. The RC curves demanded lower levels at low frequencies, thus permitting some degree of surging and turbulence. The new draft standard specifies two methods for evaluating room noise, the NC [Noise Control **3**, 19–27 (1957)] and the RNC methods [Noise Control Eng. J. **48**, 85–96 (2000)]. The former (NC) have found widespread use and the latter (RNC) bridges the differences between the NC (and NCB) and the RC criteria curves. This paper comments on the promise of the new RNC method and compares measured data were available.

1aAAb4. Evaluations of indoor noise criteria systems based on human response. Erica E. Bowden and Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska–Lincoln, 1110 S. 67th St., Omaha, NE 68182)

The goal of this research is to examine human response to background noise, and relate results to indoor noise criteria. In previous work by the authors, subjects completed perception surveys, typing tasks, and proofreading tasks under typical heating, ventilating, and air-conditioning (HVAC) noise conditions. Results were correlated with commonly used indoor noise criteria systems including noise criteria (NC), room criteria (RC) and others. The findings suggested that the types of tasks used and the length of exposure can impact the results. To examine these two issues, the authors conducted a new study in which each test subject completed 38 total hours of testing over multiple days. Subjects were exposed to several background noise exposures over 20, 40, 80, and 240 minute trials. During the trials, subjects completed a variety of performance tasks and answered questions about their perception of the noise, the thermal environment, and various other factors. Findings from this study were used to determine optimum testing conditions for on-going research examining the effects of tonal or fluctuating background noise on performance, annoyance, and spectral perception. Results are being used to evaluate the effectiveness of commonly used indoor noise criteria systems. [Work supported by INCE and ASHRAE.]

1aAAb5. Sound quality in the assessment of noise annoyance in air-conditioned buildings. Ettore Cirillo and Francesco Martellotta (Dipartimento di Fisica Tecnica-Politecnico di Bari, via Orabona 4, 70125 Bari, Italy)

The results of the search for a new parameter to be used in describing the subjective sensation of noise annoyance in air-conditioned offices are presented. The typical analysis technique in this field is to correlate objective parameters with subjective ratings by means of parametric statistics. A common procedure to find better objective descriptors of the subjective sensation is to combine different parameters. Previous studies used to combine a parameter measuring the continuous noise level with one measuring the fluctuations of noise levels. However, this approach neglected the effect that the quality of the noise produces on subjects. As the RC Mark II procedure for rating the noise of HVAC systems in buildings introduced the "Quality Assessment Index" QAI, which provides a measure of the spectral imbalance of a noise, the combination of this parameter with the RC criterion number (which only measures noise level) was proposed. A suitable combination coefficient was proposed based on the results of a noise environment survey involving both objective and subjective measurements. A second survey confirmed that the combination of these indices may significantly contribute to better describe noise annoyance.

1aAAb6. When octave bands are not enough: Case histories of indoor room noise. Jerry G. Lilly (JGL Acoust., Inc., 5266 N.W. Village Park, Dr., Issaquah, WA 98027)

All of the published indoor noise criteria evaluation methods are defined on the basis of either the octave band or the overall A- or C-weighted sound pressure levels. While this may be sufficient for some acoustic environments, there are many situations where these sound level descriptors are totally inadequate. This paper uses case histories to show that 1/3-octave and narrow-band analysis is sometimes required to properly assess the acoustic environment. To take it one step further, a recommendation is made to avoid the use of octave bands altogether and to routinely use 1/3-octave band analysis for acoustical measurements and analysis relating to indoor noise criteria.

Contributed Papers

11:50

1aAAb7. Sound transmission class (STC) is not a good predictor of performance of insulated wood frame gypsum walls used as interior partitions in most North American homes. Richard D. Godfrey, Harry Alter, and Clarke Berdan II (Owens Corning, Sci. & Technol., 2790 Columbus Rd., Granville, OH 43023)

Home owners say that insulating interior walls improves the acoustic environment. Based on STC alone, no perceptible difference is expected. To define "improved," ethnographic and laboratory studies were conducted. Ethnographic studies in 33 homes, revealed owners want quieter, less reverberant environments, including rooms where added isolation is desired. Families lives are 24/7, leading to frustration that they cannot use their homes without disturbing others. Laboratory jury studies were conducted where 35 listeners rated the relative isolation of insulated and un-insulated walls. Noise sources included broadband and real home noises. Insulated walls were perceived to perform better than uninsulated walls in all cases. Noise control engineers know that STC is only a quick screening tool (actual sound transmission loss should be used to estimate noise reduction between rooms). This is what the jurors appeared to sense. Jury ratings and the midfrequency average SPL correlated reasonably well. The STC is determined by a structural resonance near 125 Hz. Above this band, insulation has a significant impact on transmission loss (perceptible,

6 dB average). A new rating system is needed that quantifies what actual listeners hear in quiet room environments. A model using some form of room criteria is suggested.

12:05

1aAAb8. Case study of restaurant successfully designed, constructed, and operated for excellent dining acoustics. Paul Bollard (Bollard Acoust. Consultants, Inc., 1293 Lincoln Way, Ste. A, Auburn, CA 95603) and Stephen Des Jardins (La Provence Restaurant, Roseville, CA 95747)

Prior to the construction of La Provence Restaurant in Roseville, California in 2004, the owner, Stephen Des Jardins, traveled with his cook, architect, and engineer to the Provence Region of France to study the cuisine, architecture, and acoustics of the local restaurants. This information was incorporated into the design, construction, and operation of his restaurant, with acoustical design assistance provided by the author, Paul Bollard. The result of the owner's painstaking attention to detail is a restaurant which has received very positive reviews for its architecture, quality of food, service, and acoustic ambience. This paper documents the measures included in the construction of the restaurant to ensure that the building acoustics enhance the dining experience, rather than detract from it. Photographs of acoustic treatments are included, as are reverberation time (RT60) test results and ambient noise level measurement results.

MONDAY MORNING, 17 OCTOBER 2005   LA SALLE ROOM, 8:55 TO 11:45 A.M.

Session 1aAO

Acoustical Oceanography: Inversion Using Ambient Noise Sources I

Peter Gerstoft, Cochair
Univ. of California San Diego, Scripps Inst. of Oceanography, Marine Physical Lab., 9500 Gillman Dr., La Jolla, CA 92093-0238

Martin Siderius, Cochair
Heat Light and Sound Research Inc., 12730 High Bluff Dr., San Diego, CA 92130

Chair's Introduction—8:55

Invited Papers

9:00

1aAO1. Noise modeling, noise experiment, and noise inversion. Chris H. Harrison (NATO Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy)

Recent developments in inverting ambient noise to obtain bottom reflection properties originated from an understanding of noise behavior and from noise modeling. Sea-surface noise can be modeled equivalently as a ray, wave, or mode phenomenon, but often the ray or flux approach provides the most insight (explaining, e.g., the noise-notch or low-frequency noise enhancement). Theoretically,

noise vertical directionality is surprisingly sensitive to the effective depth of the sheet of sound sources, and this is borne out in the spread of behaviors found experimentally. A number of noise inversion approaches have been proposed based on coherence or directionality of sea-surface noise, individual ships or shipping, and airplane doppler. Using a drifting vertical array as a means to measure directionality, one can infer reflection coefficient versus angle and frequency. One can then use this directly for propagation calculation, or alternatively invert to geoacoustic parameters with the help of a fast noise model. A further development is to recover the reflection phase using spectral factorization, then Fourier transform to get the impulse response, which becomes a sub-bottom profile for a drifting array. At 4-kHz design frequency, the layer structure thus determined compares well with a boomer down to 15 m.

10:00–10:20 Break

10:20

1aAO2. Inversion of underwater sound from a high-Doppler airborne source (a light aircraft) for the geoacoustic properties of the sea bed. Michael J. Buckingham and Eric M. Giddens (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0238, mjb@mpl.ucsd.edu)

An inversion technique has been developed that utilizes the sound from a light aircraft to obtain estimates of the low-frequency (100–500 Hz) sound speed and attenuation in shallow-water sediments. As the aircraft flies over the channel, each engine and propeller harmonic excites a set of normal modes in the water column. The modes are Doppler-upshifted ahead of the aircraft and downshifted behind. In addition to the mode number, these frequency shifts depend on the sound speed in the sediment. A high-resolution FFT of the subsurface sound from the aircraft returns the shifted modal frequencies, allowing the sound speed in the sediment to be determined. The sediment attenuation is obtained from the spatial gradient of the modal amplitudes along the aircraft track. Experiments on the inversion technique, using a variety of light aircraft, have been conducted in the Pacific Ocean about 1 km from shore between La Jolla and Del Mar, southern California, and as part of the ONR-sponsored SAX04 in the northern Gulf of Mexico, off Fort Walton Beach. For both locations, the aircraft data yield estimates of the sediment sound speed and attenuation that are consistent with the known sediment types. [Work supported by ONR.]

10:40

1aAO3. Time synchronization and geoacoustic inversion using baleen whale sounds. Aaron Thode, Peter Gerstoft, Dale Stokes (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 90238), Mike Noad (Univ. of Queensland, St. Lucia, Queensland, Australia), William Burgess (Greeneridge Sci., Felton, CA 95018), and Doug Cato (Defense Sci. and Technol. Organization, Sydney, NSW, Australia)

In 1996 matched-field processing (MFP) and geoacoustic inversion methods were used to invert for range, depth, and source levels of blue whale vocalizations. [A. M. Thode, G. L. D'Spain, and W. A. Kuperman, J. Acoust. Soc. Am. **107**, 1286–1300 (2000)]. Humpback whales also produce broadband sequences of sounds that contain significant energy between 50 Hz to over 1 kHz. In Oct. 2003 and 2004 samples of humpback whale song were collected on vertical and titled arrays in 24-m-deep water in conjunction with the Humpback Acoustic Research Collaboration (HARC). The arrays consisted of autonomous recorders attached to a rope, and were time synchronized by extending standard geoacoustic inversion methods to invert for clock offset as well as whale location. The diffuse ambient noise background field was then used to correct for subsequent clock drift. Independent measurements of the local bathymetry and transmission loss were also obtained in the area. Preliminary results are presented for geoacoustic inversions of the ocean floor composition and humpback whale locations and source levels. [Work supported by ONR Ocean Acoustic Entry Level Faculty Award and Marine Mammals Program.]

Contributed Papers

11:00

1aAO4. Passive acoustic and seismic tomography with ocean ambient noise. Peter Gerstoft, Karim Sabra, Phillippe Roux, W. A. Kuperman, and William S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr 9500 Gilman Dr., La Jolla, CA 92093)

We have demonstrated with ocean acoustic data and associated theory and land seismic data that the Green's functions between pairs of receiving sensors can be obtained from a long-time correlation process. Noise data from fixed acoustic and seismic sensors can be used to study the ocean structure, the ocean bottom structure below the sensor field, and the earth structure between ocean sensors and land sensors hundreds to thousands of kilometers away. We have already performed extensive seismic noise tomography using land sensors in Southern California in which the major noise component originated in the ocean. We present the previous data analysis that demonstrates the feasibility of this concept.

11:15

1aAO5. Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function. Karim Sabra, Philippe Roux, Peter Gerstoft, and W. A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238, ksabra@mpl.ucsd.edu)

It has been demonstrated experimentally and theoretically that an estimate of the Green's function between two receivers can be obtained from the time derivative of the long-time average cross correlation of ambient noise between these two receivers. The emergence rate of these deterministic coherent arrival times of the cross-correlation function (i.e., the Green's function estimate) from the recordings of an isotropic distribution of random noise sources is derived by evaluating the amplitude of the variance of the cross-correlation function. The knowledge of the emergence rate of NCF is essential for practical applications. To the first order, the variance is equal to the ratio of the product of the recorded energy by

both receivers and the time–bandwidth product of the recordings. The variance of the time derivative of the correlation function is shown to have a similar dependency. These simple analytic formulas show a good agreement with the variance determined experimentally for the correlation of ocean ambient noise recorded in shallow water at a depth of 21 m, in the frequency band [300–530 Hz] for receiver separation up to 28 m and averaging time from 1 to 33 min.

11:30

1aAO6. A passive fathometer and sub-bottom profiler using ambient noise. Martin Siderius and Michael B. Porter (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130)

Recent work has shown a rather remarkable property of the ocean noise field: by correlating the noise field measured by two sensors one can recover the two-point Greens function representing the impulse response between the two sensors. Obviously there are some restrictions about where, when, why, and how this works. However, it does work and we are just beginning to understand its restrictions. Another dimension to this subject is how can the two-point Greens function be exploited? Here we develop this technique to produce a passive fathometer and sub-bottom profiler. In principle the technique can be done using just two phones; however, we show how to combine the cross-correlations of all phone pairs in an array to obtain a stronger signature. The resulting algorithm yields both a map of the bottom depth (passive fathometer) and the locations of significant reflectors in the ocean bottom. In this paper, the technique will be described and illustrated using simulations and measurements from recent experiments.

MONDAY MORNING, 17 OCTOBER 2005

SALONS B AND C, 8:00 TO 9:15 A.M.

NOISE-CON 2005 Opening Ceremony and Plenary Session

Daniel J. Kato, Chair
Cummins Power Generation, 1400 73rd Ave., N.E., Minneapolis, MN 55432

Session 1aNCa

NOISE-CON: Plenary: Perspectives on Noise in the Menu of Environmental Issues, and the Role of Technical Solutions Relative to Policy Approaches

Robert J. Bernhard, Chair
Purdue Univ., R. W. Herrick Labs., School of Mechanical Engineering, West Lafayette, IN 47907-2031

Chair's Introduction—8:15

8:20

1aNCa1. Perspectives on noise in the menu of environmental issues, and the role of technical solutions relative to policy approaches Carl Burleson (Office of Environment and Energy, Federal Aviation Admin., AEE-1, 800 Independence Ave., SW, Washington, DC 20591)

Over the last thirty years, the number of people affected by aircraft noise has decreased while air travel has grown dramatically. Despite this success, environmental issues, both noise and emissions, may become the key constraint on future air transportation expansion. As part of the Next Generation Air Transportation System Plan, the goal is for environmental protection that allows for sustained aviation growth. This includes reductions, in absolute terms, of community noise and local air quality emissions from aviation that significantly impact human health and welfare. Achieving these goals will require a multi-pronged approach that combines source reduction technology, operational abatement procedures, intelligent land use planning, and airport and aircraft operating restrictions. Given the interdependencies of impacts, there is a need to move from stand-alone models to an integrated tool set that allows understanding of the interactions between noise, air quality, economic costs and benefits. Such models would enable more effective evaluation of proposed air transportation system changes and allow communities and decision-makers better information to reach decisions on the proposed changes. FAA has initiated a multi-year, multi-million dollar effort to develop these models and improved capabilities and decision-making that will aid in tackling the enviornnmental challenges of aviation growth.

Session 1aNCb

NOISE-CON and Noise: Numerical Methods in Acoustics I

Ron Gerdes, Cochair
3M, 3M Center Bldg., St. Paul, MN 55144

Ravi Raveendra, Cochair
3830 Packard, Ann Arbor, MI 48108

9:45

1aNCb1. Models of the acoustic radiation and transmission properties of complex trimmed structures. Vincent Cotoni, Phil Shorter, Arnaud Charpentier, and Bryce Gardner (ESI US R&D, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130)

A number of advances have been made recently in the development of a hybrid method for rigorously coupling finite element and statistical energy analysis descriptions of the dynamics of a vibro-acoustic system. The method provides an efficient way to analyze the acoustic radiation and transmission properties of a complex structure across a broad frequency range. In this paper, two case studies of engine components are used to validate the "hybrid area junction" formulation for coupling FE structures with trimmed SEA fluids, and to demonstrate the use of the method in a design process.

10:05

1aNCb2. Modeling vibration isolator performance with hybrid finite element/statistical energy (FE-SEA) analysis. Bryce Gardner, Phil Shorter, and Vincent Cotoni (ESI US R&D, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130)

Vibration isolators are often modeled as simple single degree of freedom systems. Such an approach is often adequate for characterizing the low frequency performance of a vibration isolator (assuming that the effective spring stiffness and damping loss factor of the isolator can be obtained). However, at mid to high frequencies, two problems are often encountered. The first is that the impedance of the structures connected to the isolator become important (assumptions of rigid body behavior of the components of interest are generally not valid). The second is that the internal dynamic behavior of the isolator becomes important (internal resonances of the isolator degrade the isolation performance). This paper describes the development of a hybrid FE-SEA model of an *in situ* vibration isolator that addresses these problems. The key factors affecting the mid and high frequency performance of an isolator are investigated.

10:25

1aNCb3. Finite element characterization of complex automotive panels for statistical energy analysis (SEA). Phil Shorter, Francois de Boussiers, and Noel Frederick (ESI US R&D, 12555 High Bluff Dr., Ste. 250, San Diego, CA 92130)

Complex panels typical of automotive structures are often difficult to describe using statistical energy analysis (SEA) for vibro-acoustic prediction at high frequencies. The panels are usually curved, ribbed, with non-homogenous thickness, and some of these features may be crucial to correctly capture the panels vibro-acoustic behavior. It is therefore sometimes desirable to include some details in the SEA description of the panel by characterizing the statistical panel properties based on a finite element simulation. In this paper, the key SEA parameters are identified and a general process for obtaining SEA properties from a finite element model is demonstrated on an engine component.

10:45

1aNCb4. Prediction of airplane aft-cabin noise using statistical energy analysis. Andrew K. Fung and Evan B. Davis (Boeing Commercial Airplanes, M.S. 67-ML, P.O. Box 3707 Seattle, WA 98124)

Statistical energy analysis (SEA) predictions of turbulent boundary layer and engine exhaust noise in the aft cabin of an airplane have been made and compared to flight data. Measurements of engine shock-cell sound pressure levels, characterized by relatively long correlation lengths and circumferential and axial variation along the fuselage surface, were extrapolated and used as source inputs to an SEA model of a widebody airplane fuselage. Turbulent boundary layer pressure fluctuations, characterized by relatively short circumferential correlation lengths and uniformity over the aft fuselage, were represented using Efimtsov empirical correlation functions. Model variance was predicted using the Langley method and combined with estimates of measurement uncertainty to verify the prediction process.

11:05

1aNCb5. Development of new vibration energy flow analysis software and its applications to vehicle systems. D.-J. Kim, S.-Y. Hong and Y.-H. Park (Dept. of Naval Architecture and Ocean Eng., Seoul Natl. Univ., Seoul 151-742, Korea)

The Energy flow analysis (EFA) offers very promising results in predicting the noise and vibration responses of system structures in medium-to-high frequency ranges. We have developed the Energy flow finite element method (EFFEM) based software, EFADSC++ R4, for the vibration analysis. The software can analyze the system structures composed of beam, plate, spring-damper, rigid body elements and many other components developed, and has many useful functions in analysis. For convenient use of the software, the main functions of the whole software are modularized into translator, model-converter, and solver. The translator module makes it possible to use finite element (FE) model for the vibration analysis. The model-converter module changes FE model into energy flow finite element (EFFE) model, and generates joint elements to cover the vibrational attenuation in the complex structures composed of various elements and can solve the joint element equations by using the wave tra! nsmission approach very quickly. The solver module supports the various direct and iterative solvers for multi-DOF structures. The predictions of vibration for real vehicles by using the developed software were performed successfully.

11:25

1aNCb6. Transmission loss assessments of sandwich structures by using a combination of finite element and boundary element methods. Shan Li, Tongan Wang, and Steven R. Nutt (Dept. of Material Sci., Univ. of Southern California, 3651 Watt Way, VHE-602, Los Angeles, CA 90089)

In this work, formulation of a 2D fully-coupled finite element method (FEM)/boundary element method (BEM) to simulate the measurements of sound transmission loss of sandwich panels is presented. Specifically, the structural behavior of the sandwich panels, based on a consistent higher-

order theory, is implemented using finite element method (FEM), and the reverberant/anechoic chambers are accessed by boundary element method (BEM). The coupling between the structure and the acoustic medium is achieved by assuming the continuity of the normal velocities at the interface. The absorption of the receiving anechoic chamber is calibrated by comparing the numerically-predicted sound pressure level difference between the two chambers with the Sewell's expression for the forced transmission. The obtained correction factors are then used without any modification to predict transmission loss of other sandwich panels with different dimensions and material properties. Numerical examples are presented to validate the numerical procedure. Compared with traditional finite element approach, the proposed hybrid method provides more computation efficiency, and, hence, can be used to study acoustic behavior of sandwich panels at higher frequencies.

MONDAY MORNING, 17 OCTOBER 2005   SALON A, 9:45 TO 11:45 A.M.

Session 1aNCc

NOISE-CON and Noise: Measurement of Information Technology Product Noise Emissions I

Terry W. Baird, Cochair
Hewlett-Packard Co., 11311 Chinden Blvd., Boise, ID 83714

Marco Beltman, Cochair
Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124

9:45

1aNCc1. Acoustic emissions of digital data video projectors—Investigating noise sources and their change during product aging. Michael Shane White (Environ. Test Labs, InFocus Corp., M.S. White, 27700B SW. Parkway Ave., Wilsonville, OR 97070)

Acoustic emission testing continues to be a growing part of IT and telecommunication product design, as product noise is increasingly becoming a differentiator in the marketplace. This is especially true for digital/video display companies, such as InFocus Corporation, considering the market shift of these products to the home entertainment consumer as retail prices drop and performance factors increase. Projectors and displays using Digital Light Processing(tm) [DLP(tm)] technology incorporate a device known as a ColorWheel(tm) to generate the colors displayed at each pixel in the image. These ColorWheel(tm) devices spin at very high speeds and can generate high-frequency tones not typically heard in liquid crystal displays and other display technologies. Also, acoustic emission testing typically occurs at the beginning of product life and is a measure of acoustic energy emitted at this point in the lifecycle. Since the product is designed to be used over a long period of time, there is concern as to whether the acoustic emissions change over the lifecycle of the product, whether these changes will result in a level of nuisance to the average customer, and does this nuisance begin to develop prior to the intended lifetime of the product.

10:05

1aNCc2. Identification of annoying sound emitted by consumer electronic devices. Menachem Rafaelof (Personal Storage Design Ctr., Seagate Technol., 389 Disc Dr., Longmont, CO 80503)

There is a growing interest to integrate hard disc drives in many consumer electronics (CEs) for storage of digital data. Many of these new applications dictate low overall noise operation by the whole system in extremely low noise background environments. In these quiet surroundings, simple analysis of the amplitude of sound may not be sufficient to assess the impression it might have on an observer in terms of its annoyance. This work attempts to develop a new approach to gauge the annoyance of sound emitted by all types of equipment slated for operation in very quiet environments. The proposed process is intuitive and may be directly related to the perceive sensation of sound by humans. As such, it provides a means by which the designer of either a subcomponent or the system may be able to assess the acoustic emission and hence its potential annoyance in advance during development.

10:25

1aNCc3. Measurement and analysis of noise from a ventilation fan. L. Huang, H. Z. Lu, J. Wang, and R. M. C. So (Dept. of Mech. Eng., The Hong Kong Polytechnic Univ., Kowloon, Hong Kong)

A window-type ventilation fan of 150 mm in diameter with a flow rate of 170 cfm is studied. The fan has four rotor blades and the downstream motor is held by two metal struts. Acoustic measurement in an anechoic chamber reveals an asymmetric pattern of sound intensity at the blade passing frequency with a major beam in the front and a minor beam at the back of the fan. The beams are basically aligned with the rotational axis, implying a thrust-noise mechanism, but there is irregularity in the distribution. Analysis shows that such irregularity may be caused by the interference between the leading order thrust noise and the higher order noise mechanisms. Modifications of the front and back geometries reveal that the dominant tone noise mechanism lies in the vortex shedding from the four rims of the short circular exit passage which is housed in a square outer frame. Correction of the exit duct geometry reduces the tone noise power by about 5 dB, and the pattern of the residual sound intensity is changed drastically. Comparison is also made between the directivity of the ventilation fan with that of a computer cooling fan with similar features.

10:45–11:05 Break

11:05

1aNCc4. Impulsive noise of printers: measurement metrics and their subjective correlation. Terrence Baird, Norman Otto (Sound Quality Consultant, Plymouth, MI 48170) (Hewlett-Packard, 11311 Chinden Blvd., Boise, ID 83714), Wade Bray, and Mike Stephan (HEAD Acoust., Brighton, MI 48116)

In the office and home computing environments, printer impulsive noise has become a significant contributor to user perceived quality or lack thereof, and can affect the user's comfort level and ability to concentrate. Understanding and quantifying meaningful metrics for printer impulsivity is becoming an increasingly important goal for printer manufacturers. Several methods exist in international standards for measuring the impulsivity of noise. For information technology equipment (ITE), the method for detection of impulsive noise is provided in ECMA-74 and ISO 7779. However, there is a general acknowledgement that the current standard method of determining impulsivity by simply measuring A-weighted sound pressure level (SPL) with the impulsive time weighting, I, applied is inadequate to characterize impulsive noise and ultimately to predict user satisfaction and acceptance. In recent years, there has been a variety of

new measurement methods evaluated for impulsive noise for both environmental and machinery noise. This paper reviews several of the available metrics, applies the metrics to several printer impulsive noise sources, and makes an initial assessment of their correlation to the subjective impressions of users. It is a review and continuation of the work presented at InterNoise 2005 (Baird, Bray, and Otto).

11:25

1aNCc5. The effect of faint, high-pitched tones on the perceived sound quality of electronic devices. John N. Olsen (Hewlett-Packard, 18110 SE 34th St., Vancouver, WA 98683)

Faint, high-pitched squeaks repeated during printing are a common acoustic flaw in inkjet printing and in other electromechanical devices. Such noises are added to or filtered out of actual inkjet printer acoustic time histories and then played for a number of acoustic quality juries. The effect of the squeaks on acoustic quality jury scores is determined, and the data is examined by age and gender to determine whether or not there is a age or gender sensitivity to high-pitched faint squeaks. Various metrics are examined to determine their applicability and correlation to this problem.

MONDAY MORNING, 17 OCTOBER 2005　　　　　　　　　　　　　　　SALON C, 9:45 A.M. TO 12:05 P.M.

Session 1aNCd

NOISE-CON and Noise: State and Local Noise Policies and Noise Ordinances

Paul L. Burge, Cochair
Acentech, Inc., 33 Moulton St., Cambridge, MA 02138

George A. Luz, Cochair
4910 Crowson Ave., Baltimore, MD 21212

9:45

1aNCd1. The state of state noise regulations in New England. Leslie D. Blomberg (Noise Pollution Clearinghouse, P.O. Box 1137, Montpelier, VT 05601) and Paul L. Burge (Acentech, Cambridge, MA 02138)

Together, the six New England states provide a unique study and comparison of state noise regulations. The six states combined are about the same size as the State of Washington. The 14 million people who live in New England reside in a diverse mix of urban, suburban, and rural areas. The regulatory approach to noise of the six states is also very diverse. This paper will describe the statewide noise rules as they exist, compare them to one another, and highlight what might make them unique or different. It will explore issues of enforceability, consistency, and effectiveness. It will also look at the relationship between state regulation and local regulation of noise, and provide an historical context for the state regulations.

10:05

1aNCd2. Regulating environmental noise in Minnesota. Timothy Casey (HDR Eng., Inc., 6190 Golden Hills Dr., Minneapolis, MN 55416)

In Minnesota, environmental noise is regulated by the Minnesota Pollution Control Agency (MPCA) noise control rules (MN Rules 7030). Features of the MPCA noise rule include use of L10 and L50 to regulate maximum allowable daytime and nighttime noise (on a dBA basis), and use of the state environmental noise rule to limit maximum allowable levels of highway traffic noise. The MPCA noise rule is guided by provisions in Minnesota Statutes (MS 116.07). Among those provisions is the requirement that MPCA adapt noise rules that give due consideration to such factors "that could interfere unreasonably with the enjoyment of life and property." This ambitious goal could be interpreted as having applicability to impulsive or ground-borne noise. Yet MPCA noise rules clearly state that they "do not identify limiting levels of impulsive noise needed

for the preservation of public health and welfare." Budget cuts eliminated the MPCA noise unit, and ended their enforcement of the state noise rule. This paper discusses the MPCA rule, program, and current noise issues in Minnesota.

10:25

1aNCd3. Regulatory scheme for noise enforcement in New Jersey. Eric M. Zwerling (Noise Tech. Assistance Ctr., Rutgers Univ., 14 College Farm Rd., New Brunswick, NJ 08822)

In the state of New Jersey, noise control programs exist at the state, county, and municipal levels. The Noise Control Act of 1971 (N.J.S.A.13:1G-1, et seq.) is the enabling legislation for the current statewide scheme of noise regulation, which empowered the New Jersey Department of Environmental Protection (DEP) to "promulgate codes, rules and regulations" to control and abate noise. Chapter 29, Noise Control (N.J.A.C. 7:29, et seq.) was adopted in 1974, regulating commercial, industrial, community, and public service facilities. Initially, the DEP conducted field enforcement of the code, but has since delegated the authority primarily to county enforcement programs, pursuant to the County Environmental Health Act (N.J.S.A. 20:3A2-21, et seq.). Local agencies may enforce the state code by agreement with their county or alternatively may adopt the Model Ordinance promulgated by the DEP. The Model Ordinance regulates additional sound sources, including residential, multidwelling and multiuse buildings, and enumerates a list of Restricted Uses and Activities. The state code and Model Ordinance are performance codes which establish specific sound level limits, measured per established protocol. Local jurisdictions lacking the desire or capability of enforcing a performance code may choose to adopt and enforce a more subjective nuisance code.

1aNCd4. Lessons learned from the Maryland environmental noise program. George Luz (Maryland Environ. Noise Advisory Council, 4910 Crowson Ave., Baltimore, MD 21212) and David Jarinko (Former Noise Control Officer, State of Maryland)

With more than 5 million people on about 6 million acres, Maryland is the fifth most-densely populated state in the U.S. By 2050, the population is expected to be denser than for all but two European nations: Belgium and The Netherlands. Thus, it comes as a surprise that funding for the noise control program was cancelled in April 2005. This paper discusses possible reasons that there is not an active constituency for the program. These include a scattered constituency, a ''one size fits all'' standard, exemption of some major noise sources, and the lack of noise planning at the county level. Potential solutions include encouraging communication among the scattered constituency, encouraging county noise planning and adopting the Illinois model of facilitating noise litigation in State courts.

1aNCd5. Progress in the development of a national noise action plan for the Commonwealth of Puerto Rico. Jose Alicea-Pou and Milagros Navon-Rivera (Environ. Quality Board of Puerto Rico, P.O. Box 11488, San Juan, PR 00910, josealicea@jca.gobierno.pr)

This presentation describes the strategies taken for the development of a noise plan for Puerto Rico. One of the most important initiatives was the establishment of an interagency committee to work on the plan. This committee identified as first priorities the need of information from studies on environmental noise levels and data from noise attitude surveys. The Environmental Quality Board is currently investigating the levels of environmental noise in different locations in the city of San Juan. The preliminary results reinforce the need of actions to control emitters of community noise, like recreational sounds (radios, TV, night clubs) and noise from vehicles with modified mufflers and boom cars. Additionally, to achieve the implementation of the plan and the development of local noise research, we developed an interagency agreement to establish a common fund to support those projects. This year we celebrated the second annual Puerto Rico Noise Awareness Day. The activities included the second symposium on noise pollution, the first art posters contest on noise in our environment for elementary school kids, and the award ceremony, ''Amigos del Menos Ruido.'' All these activities were developed to promote public awareness of this important social and environmental issue.

1aNCd6. New seismic codes and their impact on the acoustician. Patrick J. Lama (Mason Industries, 350 Rabro Dr., Hauppauge, NY 11788)

New seismic building codes for HVAC and electrical equipment, pipe ducts and conduits are being adopted nationwide. These codes affect the way acousticians practice their profession. Recently published model codes (such as IBC, NFPA, ASCE and NBC T1809-4) specify systems that require documented seismic protection. Specific performance and prescriptive code provisions that affect acoustical system applications and how they can be made to comply is included. Key terms in these codes (life safety, essential, seismic use group, category and importance factor) are explained and illustrated. A table listing major code seismic demand formulas (horizontal static seismic force, acting at the center of gravity of the equipment, pipe duct or conduit), is a useful reference. A table that defines which HVAC systems require static or dynamic analysis based on seismic use group, design category and importance factor is provided. A discussion of code-mandated Certificates of Compliance for both mountings and equipment is included and may impact acoustical decisions. New codes may require that engineers, architects and acousticians use seismic restraints with acoustical ceilings, floating floors, resilient pipe duct supports, HVAC equipment and architectural items. ''How To'' for all of this is presented with tables, details and graphs.

1aNCd7. Impact of low-frequency sound on historic structures. Louis C. Sutherland (Consultant in Acoust., 27803 Longhill Dr., Rancho Palos Verdes, CA 90275) and Richard D. Horonjeff (Consultant in Acoust. and Noise Control, Boxborough, MA 01719)

In common usage, the term soundscape usually refers to portions of the sound spectrum audible to human observers, and perhaps more broadly other members of the animal kingdom. There is, however, a soundscape regime at the low end of the frequency spectrum (e.g., 10–25 Hz), which is inaudible to humans, where nonindigenous sound energy may cause noise-induced vibrations in structures. Such low frequency components may be of sufficient magnitude to pose damage risk potential to historic structures and cultural resources. Examples include Anasazi cliff and cave dwellings, and pueblo structures of vega type roof construction. Both are susceptible to noise induced vibration from low-frequency sound pressures that excite resonant frequencies in these structures. The initial damage mechanism is usually fatigue cracking. Many mechanisms are subtle, temporally multiphased, and not initially evident to the naked eye. This paper reviews the types of sources posing the greatest potential threat, their low-frequency spectral characteristics, typical structural responses, and the damage risk mechanisms involved. Measured sound and vibration levels, case history studies, and conditions favorable to damage risk are presented. The paper concludes with recommendations for increasing the damage risk knowledge base to better protect these resources.

Session 1aNCe

NOISE-CON and Noise: Progress in Aircraft Noise Research

Robert J. Bernhard, Chair

Purdue Univ., R. W. Herrick Labs., School of Mechanical Engineering, West Lafayette, IN 47907-2031

9:45

1aNCe1. Noise impact study of a new 2004 noise abatement procedure at the Louisville airport. Natalia V. Sizov, John-Paul B. Clarke, Liling Ren (Dept. of Aeronautics and Astronautics, MIT, 77 Massachusetts Ave., Cambridge, MA 02439), Kevin R. Elmer and Belur N. Shivashankara (Boeing, Huntington Beach, CA 92647)

A flight demonstration test in September 2004 at Louisville was a continuation of research conducted in 2002 by a team sponsored by the Federal Aviation Administration's Center of Excellence for Air Traffic Systems. A continuous descent procedure was designed primarily to minimize environmental impacts such as community noise and aircraft emissions, and to maximize savings in fuel and flight time. The test was designed to show the operational suitability of the new area navigation arrival procedure that begins at cruise altitude and which may be used in daily operation on two opposite facing runways. Flyover noise measurements were taken during a two-week testing period, and a three-week baseline period. The latest research focused on detailed analysis of aircraft performance, collecting noise data and noise prediction. The noise measurements confirm increased repeatability and predictability of noise levels that resulted from the well-designed procedure. The Integrated Noise Model was used to compare noise levels of test and baseline flights. And noise predictions using precise flight data confirm that a continuous descent approach reduces noise levels by 4 to 6 decibels which in turn reduces contour area by as much as 30 percent.

10:05

1aNCe2. Airport noise complaint patterns and interviews of frequent complainers at two major air carrier airports. Nicholas Jaggers and Gary Eiff (Dept. of Aviation Technol., Purdue Univ., 1401 Aviation Dr., West Lafayette, IN 47907)

The complex and highly sensitive topic of aircraft noise and population annoyance continues to be a major inhibitor to airport development plans. The projected growth of air travel necessitates expanded capacity at many existing airports and the development and construction of new airports in order to accommodate burgeoning traveler needs. Concerns by citizens near major airports about their economic, health, and social welfare continue to generate community and individual declarations of annoyance and concern which threaten timely solutions to airport expansion plans. A deeper understanding of the nature of these concerns is important to more effectively cope with airport expansion concerns among adjacent communities and surrounding neighbors. This study analyzed existing noise complaints registered at Denver International Airport (DEN) and Fort Lauderdale/Hollywood International Airport (FLL) in an attempt to gain greater understanding of noise complaint drivers and public annoyance. Interviews of frequent complainers were utilized in order to gain richer data concerning individual annoyance issues.

10:25

1aNCe3. PARTNER Project 8: Sonic boom mitigation. Victor W. Sparrow (Grad. Prog. in Acoust., Penn State Univ., 316B Leonhard Bldg., University Park, PA 16802)

Current U.S. and international laws prohibit commercial supersonic flight over land due to the impact of conventional sonic boom noise. Aircraft manufacturers, however, now have modern computational fluid dynamics and optimization tools, unavailable when those laws were enacted, that will allow them to design and build aircraft with boom signatures that are substantially smoothed compared with traditional N-waves. One purpose of the FAA/NASA/Transport Canada PARTNER Center of Excellence Project 8 is to determine exactly which waveforms would be heard by the public if low-boom supersonic aircraft are put into service. Another purpose is to ascertain the acceptability of those waveforms. The project involves the following universities, government, and industry partners: Penn State, Purdue, Stanford, the National Aeronautics and Space Administration, the Federal Aviation Administration, Boeing, Cessna, Gulfstream, Lockheed-Martin, and Wyle Laboratories. Some of the initial project work includes studies on the propagation of sonic booms through atmospheric turbulence, on the mutual reproducibility of three sonic boom simulators, and on the realism of those simulators as determined by expert listeners. The results of all the studies are intended to provide the FAA with new data to reassess current regulations. [Work supported by NASA, the FAA, and the PARTNER industrial partners.]

10:45–11:05   Break

11:05

1aNCe4. Vibration and rattle impact due to low frequency noise investigated at Dulles airport. Erin M. Horan, Kathleen Hodgdon and Anthony Atchley (Grad. Prog. in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., P.O. Box 30, University Park, PA 16802)

A noise study was conducted at Washington Dulles International Airport to further investigate the noise and vibration impact of thrust reversers and start of takeoff roll on communities near runways. Two different houses located on the airport property were instrumented with outdoor and indoor microphones and accelerometers mounted inside on walls, windows and floor. This presentation will discuss both the experimental design and data obtained from the two different structures. Correlations between plane types, outdoor signatures and acceleration levels are considered. The outdoor signatures and acceleration levels are used to investigate the conditions under which rattle of a loosely coupled window is observed. Additionally, the design of a subsequent subjective study to examine the effects of rattle on human perception will be discussed. The combination of the objective measurements with the results of the subjective study is intended to give a better understanding of community impact due to airport noise. [Work is supported by the Federal Aviation Administration and the Exploratory and Foundational Program of the Applied Research Laboratory, Penn State University.]

11:25

1aNCe5. Characterization of aircraft noise during thrust reverser engagement. Remy M. Gutierrez, Anthony A. Atchley (Grad. Prog. in Acoust., The Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16801), and Kathleen K. Hodgdon (The Penn State Univ., University Park, PA 16801)

Airport noise impact on communities has been an area of considerable study. However, it has been determined that thrust reverser engagement is an area requiring further research. This paper presents findings on thrust reverser from a noise study done at Washington-Dulles International Airport (IAD) in October of 2004. Previous studies have found that high levels of acoustic energy in commercial aircraft during takeoff are contained below 300 Hz [Sharp, Ben H., Guovich, Yuri A., and Albee, William, W., "Status of Low-Frequency Aircraft Noise Research and Mitigation," Wyle Report WR 01-21, San Francisco, September 2001]. Preliminary analysis of thrust reverser signatures indicates similar findings. A categorization of aircraft noise during thrust reverser engagement is given and looks at factors that may affect the noise characteristics. Some of these factors include: plane type, engine type, and thrust ratings. In addition, a brief analysis of frequency weightings of the Equivalent Sound Level (Leq) and Sound Exposure Level (SEL) metrics, and their application to thrust reverser noise is discussed. [Work supported by the Federal Aviation Administration.]

11:45

1aNCe6. Subjective response to low-frequency aircraft noise. Peter J. Shapiro, Kathleen K. Hodgdon, and Anthony A. Atchley (Grad. Prog. in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804)

The standard metric currently in use by the U.S. government for community noise from aircraft is the day-night level (DNL), which is based on the A-weighting network. However, residents of communities near airports have reported annoyance due to aircraft noise in greater proportion than what the DNL contour associated with their address predicts. Complaints of "distant rumbling" appear to be related to the low-frequency noise that aircraft produce for which the DNL metric does not fully account. This presentation discusses measurements made in October 2004 at two residences near runways at Washington-Dulles International Airport. Noise data are recorded both inside and outside the structures, allowing for a measure of the events' impact both inside and outside the structure, as well as the filtering effect of the structure on the signal. Laboratory subjective evaluation studies of the recorded aircraft signatures are conducted. Objective metrics, those that are designed for low-frequency noise and those that are not, are calculated and correlated with subjective rankings of the signatures. This gauges the metrics' effectiveness in predicting subjective perception of aircraft noise signatures. [Work supported by the Federal Aviation Administration and the Exploratory and Foundational Program of the Applied Research Laboratory, Penn State University.]

MONDAY MORNING, 17 OCTOBER 2005   ROCHESTER ROOM, 9:45 TO 11:35 A.M.

Session 1aNS

Noise, NOISE-CON and ASA Committee on Standards: Special Session in Honor of William W. Lang

Paul D. Schomer, Cochair
Schomer & Associates Inc., 2117 Robert Dr., Champaign, IL 61821

George C. Maling, Jr., Cochair
60 High Head Rd., Harpswell, ME 04079

Chair's Introduction—9:45

Invited Papers

9:50

1aNS1. William W. Lang's days at the Massachusetts Institute of Technology, Bolt, Beranek and Newman, founding of the Institute of Noise Control Engineering and his efforts at creating a national noise policy. Leo L. Beranek (975 Memorial Dr., Cambridge, MA 02138)

William Lang entered acoustics through his master's degree program at MIT and through becoming the second full-time employee at Bolt Beranek and Newman. Bill stimulated the start of the Institute of Noise Control Engineering and this paper covers the early days of that organization and his contribution to its success. Finally, covered are his efforts to promote a national noise policy.

10:05

1aNS2. Bill Lang's contributions to acoustics at International Business Machines Corporation (IBM), signal processing, international standards, and professionalism in noise control engineering. George C. Maling, Jr. (60 High Head Rd., Harpswell, ME 04079)

Bill Lang joined IBM in the late 1950s with a mandate from Thomas Watson Jr. himself to establish an acoustics program at IBM. Bill created the facilities in Poughkeepsie, developed the local program, and was the leader in having other IBM locations with development and manufacturing responsibilities construct facilities and hire staff under the Interdivisional Liaison Program. He also directed IBMs acoustics technology program. In the mid-1960s, he led an IEEE standards group in Audio and Electroacoustics, and, with the help of James Cooley, Peter Welch, and others, introduced the fast Fourier transform to the acoustics community. He was the convenor of ISO TC 43 SC1 WG6 that began writing the 3740 series of standards in the 1970s. It was his suggestion to promote professionalism in noise control engineering, and, through meetings with Leo Beranek and others, led the founding of INCE/USA in 1971. He was also a leader of the team that founded International INCE in 1974, and he served as president from 1988 until 1999.

1aNS3. William W. Lang's contributions to the development of the International Institute of Noise Control Engineering. Tor Kihlman (Dept. of Appl. Acoust., Chalmers Univ. of Technol., S-412 96 Gothenburg, Sweden)

Bill Lang has been a member of the Board of I-INCE since its first meeting in 1975 in Sendai, Japan. From 1988 to 1999 he served as the President, and to this day remains a very active member of the Board. The INTER-NOISE congress series has been the core activity of I-INCE, and in every case Bill has contributed very actively to secure an excellent congress. He has also taken a number of initiatives both to improve the organization itself and to develop new activities, especially the Technical Study Groups that have prepared consensus reports on important issues such as noise barrier effectiveness and industrial noise hearing conservation policy. At present, he is very active in attempting to develop a global noise control policy.

10:35

1aNS4. Bill Lang's contributions to acoustics at the International Business Machines Corp. (IBM) and to IBM in general. Matthew A. Nobile (IBM Hudson Valley Acoust. Lab., Bldg. 704, Boardman Rd. Site, 2455 South Rd., Poughkeepsie, NY 12601) and Richard C. Chu (IBM Adv. Thermal Lab., Poughkeepsie, NY 12601)

Although Bill Lang's accomplishments and key roles in national and international standards and in the formation of INCE are widely recognized, sometimes it has to be remembered that for nearly 35 years he also had a ''day job'' at the IBM Corporation. His achievements at IBM were no less significant and enduring than those in external standards and professional societies. This paper will highlight some of the accomplishments and activities of Bill Lang as an IBM noise control engineer, the creator of the IBM Acoustics Lab in Poughkeepsie, the founder of the global Acoustics program at IBM, and his many other IBM leadership roles. Bill was also a long-serving IBM manager, with the full set of personnel issues to deal with, so his people-management skills were often called into play. Bill ended his long and fruitful IBM career at a high point. In 1988, he took an original idea of his to the top of IBM executive management, which led directly to the formation of the IBM Academy of Technology, today the preeminent body of IBM top technical leaders from around the world.

10:50

1aNS5. Bill Lang's contributions to the Acoustical Society of America. Jiri Tichy (5552 Citation Rd. North, Toledo, OH 43615)

This paper deals with the contribution of his activities as a member of the Executive Council of the Acoustical Society of America. His most important contribution was the restructuring of the financial system. As a treasurer of the Society, he converted the book keeping into a system that allows making clear decisions on the spending, keep the evidence on the balances and, most importantly charge all activities to the correct budget chapters. The budgets were of course computerized. The new system contributed substantially to balance the operating budget that was important due to the falling budget income from the JASA. Bill is the past chairman of the Technical Committee on Noise and a worldwide contributor to the acoustical standards in many functions particularly from the ASA bases. The paper will also deal with many other contributions linked to the Executive council decisions.

11:05

1aNS6. Bill Lang's many contributions to the Acoustical Society of America's domestic standard's program and to International Electrotechnical Commission (IEC) TC-29. Henning von Gierke (USAF Aerosp. Medical Res. Lab.–Retired, 1325 Meadow Ln., Yellow Springs, OH 45387)

Over the last 50 years, Bill Lang has been a major force in both National and International Standards. He has been a leader and major contributor to both, which together constitute our ASA Standards program that ASA conducts under the American National Standards Institute (ANSI) umbrella. This paper concentrates on Bill's contributions to the domestic portion of the Standards Program, contributions to our interactions with the International Electro-technical Commission (IEC) TC-29 (electro-acoustics), and Bill's unique contributions to IEC TC-29.

11:20

1aNS7. Bill Lang's many contributions to international standardization in acoustics. Roger Higginson (Bracknell, Berkshire, RG12 9JL, UK)

Over the past 40 years, Bill Lang has been a major force in Intl. Standards, not only as ongoing leader of the U.S. delegation, but in a personal capacity as project leader for the development of the pivotal group of standards on methods of determining sound power level. As Convenor of the previous Working Group 6 of Sub-Committee 1, Noise, of ISO Tech. Committee 43, Acoust., he oversaw the production of the ISO 3740 series of standards during the 1970s and early 1980s. A revision of this same series was made starting in the late 1980s, once more under Bill's supervision in the current Working Group 28. At the same time, this Group compiled the new series of ISO 11200 standards on determination of emission sound pressure levels, on declaration and verification of noise emission levels and on guidelines for writing noise test codes. As if this was not enough, the same group is now in the throes of another major short-term revision of all the documents and has plans for further more radical revisions in the future. This paper outlines Bill's contributions to this major international combined effort, which even after such sustained activity goes from strength to strength, gathering participation from a new generation of acousticians around the world.

Session 1pAAa

Architectural Acoustics: Speech Privacy in Buildings

Kenneth P. Roy, Chair

Armstrong World Industries, Innovation Ctr., 2500 Columbia Ave., Lancaster, PA 17604

Chair's Introduction—1:00

Invited Papers

1:05

1pAAa1. Forty-five years of open plan office experience: An unusual case study. Dennis A Paoletti (Shen Milsom & Wilke, Inc., 33 New Montgomery St., San Francisco, CA 94105, dpaoletti@sf.smwinc.com)

The open plan office design concept has been around for at least 45 years. Through research, design, and practical experience, the basic concepts and parameters that impact acoustics and speech privacy have been fairly well developed and documented. The psychological aspects of human response and expectation to open plan environments, and the conceptual designs that emerge from the architect and interior designers are frontiers that still allow for ongoing study. A large open plan project completed during the dot-com bubble allowed an opportunity to design and experience an unusual open plan environment that utilized exposed sound masking system loudspeakers and minimal sound absorption.

1:25

1pAAa2. Speech and noise levels for predicting the degree of speech security. John S. Bradley and Bradford N. Gover (Inst. for Res. in Construction, Natl. Res. Council, 1200 Montreal Rd., Ottawa ON, K1A 0R6, Canada, john.bradley@nrc-cnrc.gc.ca)

A meeting room is speech secure when it is difficult or impossible for an eavesdropper to overhear speech from within. The degree of security could range from less stringent conditions of being barely able to understand a few words from the meeting room, to higher levels, where transmitted speech would be completely inaudible. This paper reports on measurements to determine the statistical distribution of speech levels in meeting rooms and the distribution of ambient noise levels just outside meeting rooms. To select the required transmission characteristics for a meeting room wall, one would first decide on an acceptable level of risk, in terms of the probability of a speech security lapse occurring. This leads to the selection of a combination of a speech level in the meeting room and a noise level nearby that would occur together with this probability. The combination of appropriate estimates of meeting room speech levels and nearby ambient noise levels, together with the sound transmission characteristics of the intervening partition, makes it possible to calculate signal/noise ratio indices related to speech security [J. Acoust. Soc. Am. **116**(6), 3480–3490 (2004)]. The value of these indices indicates if adequate speech security will be achieved.

1:45

1pAAa3. Measuring speech privacy−open plan source speaker evaluation: Part 2. Kenneth Good and Kenneth Roy (Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17601)

Questions regarding differing results from equipment used to measure speech privacy in open plan areas prompted a study to evaluate the importance of the source speaker and its characteristics used in the measurement process. This paper builds on last year's evaluation of speakers of various performance characteristics. We will look more closely at variations within the spaces measured and reproduction of results.

2:05

1pAAa4. Validating a new approach for predicting the degree of speech security. Bradford N. Gover and John S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, 1200 Montreal Rd., Ottawa ON, K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

The new method of estimating the speech security of meeting rooms first predicts transmitted speech levels at spot positions 0.25 m from the outside boundaries of a meeting room. The degree of speech security at each spot measurement position is related to this transmitted speech level and to the ambient noise level at the same location. This information can be used to calculate new signal/noise ratio measures that indicate the degree of speech security [J. Acoust. Soc. Am. **116**(6), 3480–3490 (2004)]. The values of these indices will indicate whether speech will be audible or intelligible, and to what degree. This paper reports validation studies of the prediction of transmitted speech levels at points 0.25 m outside the meeting room, and the effects of varied room absorption on these predicted levels. It also reports on subjective evaluations of the transmitted speech sounds, in terms of the audibility and intelligibility of the transmitted speech, to validate expected judgments in realistic acoustical conditions.

2:25

1pAAa5. The role of sound masking treatment in the value engineering of office designs. Thomas R. Horrall and Carl Rosenberg (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

In the design and planning of office spaces, sound masking systems are often considered optional because of an owner/occupier requirement due to an understanding that can create a more pleasing and private acoustic environment or are required to provide adequate speech privacy. Installations are rarely considered on the basis of project value engineering or return on investment. In this paper, the authors argue that sound masking systems should in fact be considered as another important cost component of the office design, and evaluated along with ceiling tile, private office wall construction and workstation design for their mutual impact on the project budget, as well as on their contribution to acoustical performance.

The authors provide typical project data to demonstrate masking systems typically result in overall project cost savings.

2:40

1pAAa6. Speech privacy in customer service call centers. Timothy Foulkes (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

Customer service call centers are a special type of open plan office with different design considerations and different acoustical goals. Workers want a high degree of visual and verbal contact with adjacent members of their team, but customers do not want to hear crosstalk from other conversations. Achieving the correct balance between these conflicting goals is a challenge for the architect and acoustic consultant. This paper will present several case studies of different call centers with acoustic measurements, user comments, and other subjective assessment methods.

MONDAY AFTERNOON, 17 OCTOBER 2005   SALON F, 3:15 TO 5:10 P.M.

Session 1pAAb

Architectural Acoustics, Noise and NOISE-CON: Plumbing Noise

James E. Phillips, Chair
Wilson Ihrig & Assoc., Inc., 5776 Broadway, Oakland, CA 94618-1531

Chair's Introduction—3:15

Invited Papers

3:20

1pAAb1. Forty years of plumbing noise control. Ewart Wetherill (Ewart A. Wetherill, 28 Cove Rd., Alameda, CA 94502)

The noise of fluid movement in piping systems is a frequent source of distraction and annoyance in buildings ranging from remote mining camps to private homes and luxurious resorts. However, in most cases this problem is easy to avoid if simple precautions are taken in design and construction. This paper reviews specific examples illustrating inexpensive methods of plumbing noise control.

3:40

1pAAb2. Noise control of waste water pipes. Jerry Lilly (JGL Acoustic Inc., 5266 NW Village Park Dr., Issaquah, WA 98027)

Noise radiated by waste water pipes is a major concern in multifamily housing projects. While the most common solution to this problem is to use cast-iron pipes in lieu of plastic pipes, this may not be sufficient in high-end applications. It should also be noted that many (if not most) multifamily housing projects in the U.S.A. are constructed with plastic waste piping. This paper discusses some of the measures that developers are currently using to control noise from both plastic and cast-iron waste pipes. In addition, results of limited noise measurements of transient water flow in plastic and cast-iron waste pipes will be presented.

4:00

1pAAb3. Plumbing noise: Pressure levels and perception in a luxury condominium. Derek Watry (Wilson, Ihrig & Assoc., 577 Broadway, Oakland, CA 94618)

A consulting project has recently been completed that addressed a number of noise concerns in a 9-unit, 20-year-old luxury condominium building in the San Francisco Bay Area. Among other tasks, the noise levels of four bathroom plumbing functions (flushing, showering, bath filling, bath draining) were measured in adjoining units and an inventory of noise concerns was collected. This paper reports the measured noise levels (nearly a 20-dBA range for every function!) and looks at the corresponding resident assessments not always clearly correlated with sound-pressure level.

4:20

1pAAb4. Suggested noise criteria for plumbing systems. Jerry Lilly (JGL Acoustics Inc., 5266 NW Village Park Dr., Issaquah, WA 98027)

The issue of noise that is generated by plumbing systems has been addressed in several articles and texts in the acoustic literature, but most of this information deals with a description of the various noise generation mechanisms and recommended methods of controlling noise from plumbing fixtures and piping. As with any noise source that has the potential for generating annoyance, the question of how much noise is too much noise eventually arises. Chapter 47 of the 2003 ASHRAE Applications Handbook contains newly published guidelines for plumbing noise criteria as it impacts building occupants. This paper discusses the ASHRAE guidelines, and it also suggests additional noise criteria for other plumbing-related sources of noise in multitenant buildings.

Contributed Papers

4:40

1pAAb5. Plumbing noise in multifamily dwellings: 50 years and counting. John J. LoVerde and David W. Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404, jloverde@veneklasen.com)

Historically, plumbing sound was only of concern in luxury buildings and plumbing noise was rarely addressed explicitly in building codes or design guidelines. The last couple of decades have seen marked increases in population density, occupant expectations and sensitivity to noise, and the number of noise-related lawsuits. These demographic factors have considerably increased the importance of adequately addressing plumbing noise in multifamily projects, as shown by the increasing number of published guidelines (such as in the latest ASHRAE Handbook). Over the course of our firm's history, we have attempted to determine the relationship between the sound level from plumbing systems and the acceptability to the occupants. Our current criteria for plumbing noise and the reasoning behind them is discussed, including test cases and lawsuits, as well as the materials and methods available to meet those criteria, and their impact on the design of multifamily projects.

4:55

1pAAb6. Comparison of models for piping transmission loss estimations. Fred W. Catron (Fisher Controls Intl., 301 South 1st St., Marshalltown, IA 50158) and J. Adin Mann III (Iowa State Univ., Ames, IA 50011)

A frequency dependent model for the transmission loss of piping is important for accurate estimates of the external radiation from pipes and the vibration level of the pipe walls. A statistical energy analysis model is used to predict the transmission loss of piping. Key terms in the model are the modal density and the radiation efficiency of the piping wall. Several available models for each are compared in reference to measured data. In low frequency octave bands, the modal density is low. The model of the transmission loss in these octave bands is augmented with a mass law model in the low frequency regime where the number of modes is small. The different models and a comparison of the models will be presented.

MONDAY AFTERNOON, 17 OCTOBER 2005 LA SALLE ROOM, 1:00 TO 3:00 P.M.

Session 1pAO

Acoustical Oceanography: Inversion Using Ambient Noise Sources II

Peter Gerstoft, Cochair

Univ. of California San Diego, Scripps Inst. of Oceanography, Marine Physical Lab., 9500 Gilman Dr., La Jolla, CA 92093-0238

Martin Siderius, Cochair

Heat, Light and Sound Research Inc., 12730 High Bluff Dr., San Diego, CA 92130

Invited Papers

1:00

1pAO1. Inversion of bottom sound speed and dispersion using boat noise. Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

An experiment was conducted near shore in water depth between 2 and 10 m. The sediment consists of uniform sand. A lone hydrophone was moored 2 m above the bottom at 6 m depth. A small boat traveling at constant speed was used as the sound source, and ran both parallel and perpendicular to shore. Thus, both the range-independent as well as range-dependent cases can be investigated. Environmental parameters in both the water column and the sediment are independently measured. We first will study the propagation of broadband noise from bubbles emitted from a small boat in this special environment, especially the interferences of modes in the wedge-shape waveguide. Both analytical and numerical approaches are used to simulate the field experiment and to obtain general conclusions concerning mode interference in range-dependent environments. Then we will use the data to invert for sediment sound speed as a function of frequency. This is achieved by investigating mode cut-off for different frequencies at different water depths. In the present work, sediment sound speed is estimated over the frequency range of 500 C–4500 Hz. [Work supported by the U.S. ONR.]

1pAO2. Geoacoustic inversion using ship noise received on the ship-towed line array. T. C. Yang, Kwang Yoo, and Laurie Fialkowski (Naval Res. Lab., Washington, DC 20375)

There has been increased interest in geoacoustic inversion using a ship-towed line array including recent work by M. Siderius *et al.*, J. Acoust. Soc. Am. **11**, 1523 (2002), D. J. Battle *et al.*, IEEE J. Ocean Eng. **28**, 454 (2003), and M. R. Fallat *et al.*, IEEE J. Ocean Eng. **29**, 78 (2004). One notes that bottom properties in the littoral oceans can vary substantially over a small spatial scale (a few kilometers). Using either manned noise or own ship noise received on the towed array, one can invert for the geoacoustic properties of the bottom under the ship. With the ship's mobility, one can potentially survey a large area in a short time. The problem is that geoacoustic inversion can have ambiguous solutions, namely, that often more than one solution can fit the data. For building a database, this method becomes useless if inconsistent solutions are obtained between measurements. In this paper, recent advances that lead to a reliable, consistent, and often unique solution using a signal with known waveform are reviewed. This is demonstrated with at-sea data and generalized to geoacoustic inversion using own ship noise. [Work supported by the U.S. ONR.]

1pAO3. Geoacoustic inversion using broadband noise from surface ships. Robert A. Koch and D. P. Knobles (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8025, koch@arlut.utexas.edu)

Spatiotemporal techniques developed for geoacoustic inversion using broadband noise from surface ships will be discussed. Inversion processing will be demonstrated with data obtained from a bottom-mounted horizontal line array at a depth of 110 m in the Gulf of Mexico near Corpus Christi, TX. The inversion solutions for the geoacoustic solutions are validated by comparing modeled transmission loss at several frequencies with loss from a towed source that was measured in another portion of the experiment. The first step in developing the inversion algorithm was to obtain consistent results for geoacoustic parameters when the source track reconstruction is available. That such robustness requires temporal samples in the cost function is established from the results of inversions from normalized cross spectra. The next step in the algorithm development was to perform simultaneous inversion for geoacoustic and uniform source motion parameters to examine complications that might arise when source track reconstruction is unavailable. Examples showing the effect on robustness due to violations of uniform motion are presented. Also, the advantages and applications of alternative cost functions that match the structure in received level data and modeled fields will be described. [Work supported by ONR.]

Contributed Papers

1pAO4. Application of adaptive beamforming to geoacoustic inversion and source localization. Steven Stotts and Brian Hawkins (Appl. Res. Labs., Univ. of Texas, 10000 Burnet Rd., Austin, TX 78729-8029)

Utilizing sources of opportunity to perform simultaneous geoacoustic inversion and localization has been established using conventional beamforming (CBF) applied to data from subapertures of a bottom-mounted horizontal line array. [R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," J. Acoust. Soc. Am. **117**, 626–637 (2005)] Often, multiple sources may coexist within a field, and isolating a single source is desirable prior to performing inversions. Spatial filtering has been shown to improve localizations by focusing on bearing regions containing the source of interest. [S. A. Stotts, "A robust spatial filtering technique for multisource localization and geoacoustic inversion," J. Acoust. Soc. Am. **117**, 139–162 (2005)]. Adaptive beamforming (ABF) is an alternative approach used to suppress noise and increase the signal-to-noise ratio for sources of interest. Comparisons between previous CBF inversion results and ABF inversions applied to a known data set will be presented. The inversion algorithm uses simulated annealing with a cost function correlating cross-spectral data from the subapertures with modeled cross spectra.

1pAO5. Estimating geoacoustic properties of marine sediments by matched field inversion using ship noise as a sound source. Michael G. Morley and N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W3P6, Canada)

The first stage of a seafloor observatory to monitor marine gas hydrates will be deployed in the Gulf of Mexico in the fall of 2005. One component of the monitoring station is a vertical line hydrophone array moored on the seafloor. This paper reports research to investigate using matched field inversion of sound generated by passing ships to detect changes in the geoacoustic parameters of the seabed that may indicate the occurrence of a gas hydrate dissociation event. First, synthetic multifre-

quency data were used to investigate the performance of the inversion method for estimating sediment sound speeds near the seafloor. The synthetic data were generated using the parabolic equation method for a geoacoustic environment that simulated the seabed at the site in Mississippi Canyon. The geoacoustic model consisted of a multilayered structure with relatively slow sound speeds to significant depths below the seafloor. The inversion based on normal mode replica fields indicated that performance strongly depends on using a realistic parametrization of the geoacoustic model. Experimental ship noise data were recorded along radial tracks in a preliminary deployment of the vertical array. Results are reported for estimating a geoacoustic model from the noise data.

1pAO6. Geoacoustic inversion by using broadband ship noise recorded on the New Jersey Shelf. Altan Turgut (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

During the New Jersey Shelf RAGS03 experiment, acoustic signals emitted by ships of opportunity (merchant ships) are simultaneously recorded on three vertical line arrays (VLAs) and a horizontal line array (HLA). Recorded broadband (50–750 Hz) acoustic data sets are used to demonstrate the source localization capability as well as geoacoustic inversion capability. Waveguide invariant theory applied to beamforming by two vertical arrays that provided a range ratio of the source to the receivers. Beamforming by a horizontal array provided time-evolving spectrum for a particular look direction (LOFARgram). The trajectories of the striations observed both in vertical array beamformer output and LOFARgrams are used to estimate the source speed, range, and azimuthal direction as well as bottom geoacoustic parameters. Trajectories of striations are identified by the Hough transform that converts a difficult global detection problem in the image domain into a simpler local peak detection problem. Analysis of both simulated and RAGS03 data indicate the feasibility of source localization and geoacoustic inversion by using broadband noise signals emitted by distant surface ships. [Work supported by ONR.]

2:45

1pAO7. Matched-field passive inversion using ships of opportunity.
Donald R. DelBalzo, James H. Wilson, Robert A. Prater, Peter D. Neumann (Planning Systems, Inc., 40201 Hwy. 190 East, Slidell, LA 70461, ddelbalzo@plansys.com)

Acoustic field estimation in littoral areas is severely limited by uncertainty in sediment property characteristics. Direct methods of obtaining geophysical data are expensive, so inverse methods have recently become popular. Most techniques rely on inverting reverberant signals from controlled active sources received on vertical apertures. This work is focused on using surface ships of opportunity in a range-staggered way to create a range-dependent mosaic of sediment properties and bathymetry with ever increasing confidence. Matched-field/matched-phase techniques and adaptive simulated annealing are applied in a complex simulated shallow-water environment with a single vertical array and high signal-to-noise ratios. These procedures can also be applied to horizontal arrays. The inversion process is evaluated as a function of frequency bandwidth and array length. The simulations indicate significant potential for passive geoacoustic characterizations out to ranges of tens of water depths in shallow water.

MONDAY AFTERNOON, 17 OCTOBER 2005

CONRAD C, 1:00 TO 5:15 P.M.

Session 1pEA

Engineering Acoustics and ASA Committee on Standards: An ANSI Standard for Measuring In-Situ Directivity of Hearing Aids in 3-Dimensions

David A. Preves, Cochair
Starkey Labs., 6600 Washington Ave., South, Eden Prairie, MN 55344

Daniel M. Warren, Cochair
Knowles Electronics Inc., 1151 Maplewood Dr., Itasca, IL 60143

Chair's Introduction—1:00

Invited Papers

1:05

1pEA1. A tour of ANSI S3.35-2005. Daniel M. Warren (Knowles Electron., 1151 Maplewood Dr., Itasca, IL 60134, daniel.warren@knowles.com)

ANSI Standard S3.35-2005, "Method of measurement of performance characteristics of hearing aids under simulated real-ear working conditions," is a revision of an earlier, 1985 standard for measuring insertion gain and directional figures-of-merit for hearing aids on a head-and-torso testing manikin. The primary feature of the new revision is the measurement of directionality based on a three-dimensional array of sound sources rather than sources arrayed in a single plane. The author of the new revision will give an overview, highlighting distinctions from the 1985 revision. The methods for measurement of directionality in three dimensions will be described in detail.

1:25

1pEA2. A simulated diffuse field for *in situ* microphone measurements. Christopher Struck (Dolby Labs., Inc., 100 Potrero Ave., San Francisco, CA 94103)

The asymmetric polar response of an *in situ* microphone system requires calculation of the Directivity Index (DI) from the random incidence response, as a calculation of the DI from polar response data in a single plane is inadequate. A technique is presented enabling diffuse field measurements to be performed, without a reverb chamber, in a moderately sized, untreated room over a wide bandwidth. The limitations and practicality of alternative methods are discussed. The use of four appropriately positioned, uncorrelated noise sources is shown to be simple, fast, and practical in comparison to roving source free-field tests. This also greatly simplifies the DI calculation. The importance of carefully positioning and calibrating the uncorrelated noise sources is shown. The diffusivity of the sound field is further improved by rotation of the test object during the measurement. Methods of verifying the diffusivity of the simulated sound field are discussed. Results are shown for several nonsimple microphone devices including a measurement manikin, a behind-the-ear hearing aid mounted on a manikin, and an in-the-ear hearing aid mounted on a manikin. The method is shown to be applicable to any asymmetrical polar response microphone system including, but not limited to headsets, lavalier microphones, and loudspeaking telephones.

1:45

1pEA3. Computer simulation of three-dimensional directional response measurements. Daniel M. Warren (Knowles Electron., 1151 Maplewood Dr., Itasca, IL 60134, daniel.warren@knowles.com)

Early in the development of methods for measuring the three-dimensional directional response of hearing aids, there were many questions on how to distribute sound sources over the sphere. The method had to be accurate, yet practical to implement. Decisions on the scheme for distributing sound sources, the density of sound sources, and the weighting of results from each source have impact on accuracy. A simulation of the directional response of an array of microphones in the direct and scattered acoustic field near a rigid sphere was created and used to evaluate various means of distributing sound sources.

2:05

1pEA4. Comparison of 48-point 3-D anechoic and diffuse sound field measurements of directivity index. Ron Scicluna, Mead Killion, Andy Haapapuro, and Stephen Julstrom (Etymotic Res., Inc., 61 Martin La., Elk Grove Village, IL 60007)

The directivity index (DI) as the single best descriptor of directional hearing aid performance at a given frequency has been agreed to in ANSI Standard S3.35-2005. The standard specifies that the *in situ* directivity index of a hearing aid is not to be measured in a diffuse field (as in a reverberation chamber), as implied by the traditional definition, but in an anechoic chamber, where the rms average of measurements taken with the loudspeaker oriented at 48 locations spread out over an imaginary sphere is used to approximate the diffuse-field directivity index. The corresponding directivity index estimates can be labeled diffuse field and 3D, respectively. These two measurements are theoretically equivalent to within 0.2 dB for a first-order directional microphone. A round-robin series among several laboratories found three laboratories that obtained the same DI within an average of 0.13 dB between 500 and 4000 Hz. Nonetheless, no direct comparison between the 3-D and reverberation-chamber methods has been published. We will present data obtained using both methods, and speculate as to why no one likes the simpler reverberation method.

2:25

1pEA5. Diffuse field directivity index measurements of hearing aids operating with adaptive gain algorithms. Thomas Burns (Starkey Labs., 6600 Washington Ave. S., Eden Prairie, MN 55344)

In the summer and autumn of 2004, two hearing aids were used in a round robin study of 3-D directionality as per the test protocol in ANSI S3.35. This protocol required a simulated diffuse field in an anechoic environment by sequentially positioning a loudspeaker along the surface of a 1-m radius sphere circumscribing the device under test. The aids contained both an omni mic and a gradient directional mic, and operational switching between each mic was done with a pushbutton memory. Recently, these aids were brought into Riverbank Labs in Geneva, IL to investigate a different measurement protocol utilizing a simulated diffuse field in a reverberation chamber. The test protocol will be described, and measured results of the 3-D directivity index in this simulated diffuse field will be compared to results from the original round robin study. Finally, this new protocol will be used to measure the 3-D directivity index of hearing aids having an adaptive gain algorithm; these results will be compared to the results of the same adaptive aids as measured with the original ANSI S3.35 protocol even though this protocol specifically requires all adaptive settings to be disengaged.

2:45–3:00   Break

3:00

1pEA6. Data on directionality measures of adaptive systems. Michael J. Nilsson, Jim Wellington, Jay Kirsch, and Victor Bray (Sonic Innovations, 2795 East Cottonwood Pky., Ste. 660, Salt Lake City, UT 84121)

Directionality has become the dominant method to improve the signal-to-noise ratio in hearing aids. Quantification of differences in the amount of directionality between systems has become a source of scientific, as well as marketing, interest. Developments in directionality now involve signal processing methods that interact with directional microphones based upon adaptive algorithms in order to broaden the appeal and usefulness of these systems. The ANSI S3.35 revision for 2004 improves the methodology for relating measures of directionality to expected performance when using many systems, but fails to address measurements using these adaptive systems. These systems respond differently depending upon the type, volume, or location of a stimulus, and will generate different results using the single noise source technique included in the standard compared to diffuse noise conditions. This presentation presents data using various measurement techniques to quantify the differences between directional systems. Different types of adaptive systems will be identified that vary in their interaction with the ANSI methods and an understanding of how measurements change with changes in technique will be presented.

3:20

1pEA7. Verification of the directivity index and other measures of directivity in predicting directional benefit. Andrew Dittberner (GN Resound, 4201 W. Victoria St., Chicago, IL 60646, adittber@msn.com) and Ruth Bentler (Univ. of Iowa, Iowa City, IA 50214)

The relationship between various directivity measures and subject performance with directional microphone hearing aids was determined. Test devices included first- and second-order directional microphones. Recordings of sentences and noise (Hearing in Noise Test, HINT) were made through each test device in simple, complex, and anisotropic background noise conditions. Twenty-six subjects, with normal hearing, were administered the HINT test recordings, and directional benefit was computed. These measures were correlated to theoretical, free-field, and KEMAR DI values, as well as front-to-back ratios, *in situ* SNRs, and a newly proposed Db-SNR, wherein a predictive value of the SNR improvement is calculated as a function of the noise source incidence. The different predictive scores showed high correlation to the measured directional benefit scores in the complex (diffuse-like) background noise

condition ($r=0.89-0.97$, $p<0.05$) but not across all background noise conditions ($r=0.45-0.97$, $p<0.05$). The Db-SNR approach and the *in situ* SNR measures provided excellent prediction of subject performance in all background noise conditions ($0.85-0.97$, $p<0.05$) None of the predictive measures could account for the effects of reverberation on the speech signal ($r=0.35-0.40$, $p<0.05$).

3:40–4:50

Panel Discussion: Practical Implementations and Round Robin Results Thomas Burns, Moderater

A "round robin" measurement series was conducted using the same hearing aids tested at several different facilities. The panel papers discuss the equipment methods and results for measurement of hearing aid directivity according to the standard. The papers conclude with a comparison of the round robin results, and an open discussion of reasons for the differences among the measurements.

3:40

1pEA8. Method for calculating directivity index of a directional microphone in a hearing aid on a manikin. Andrew Dittberner (GN Resound, 4201 W. Victoria St., Chicago, IL 60646, adittber@msn.com) and Ruth Bentler (Univ. of Iowa, Iowa City, IA 50214)

A method for computing a directivity index (DI) on a manikin for directional microphones in hearing aids is proposed and investigated. Test devices included first- and second-order directional microphones in hearing aids. Signal presentation involved a single noise source rotated completely around the directional microphone, in free field and on a manikin, at a defined radius. The area covered was equivalent to the approximate surface area of a sphere. It was anticipated that an equal angular resolution of 10 deg (elevation and azimuth) would effectively estimate the DI of first-, second-, and higher-order directional microphone systems located in a hearing aid on a manikin. A total of 450 spatially varied presentation points was analyzed, each weighted in reference to direction of arrival on the directional microphone. The absolute difference between the Directivity Index derived from the modified method proposed in this investigation and the conventionally derived Directivity Index on a manikin were as large as 3.8 dB in the higher frequencies, depending on the device under test. The magnitude of these differences was dependent on microphone location. In other words, the further the microphone was placed into the ear of the manikin, the larger the absolute difference.

3:50

1pEA9. Methods for hearing aid directivity index measurements under simulated real-ear conditions. Randall Wagner (Natl. Inst. of Standards and Technol., Bldg. 233, Rm. B102, Gaithersburg, MD 20899-8220, randall.wagner@nist.gov)

One technique traditionally used to measure the directivity index of hearing aids involves acquiring data with sound source locations at multiple azimuth angles restricted to the horizontal plane. For measurement of the directivity index of a hearing aid under simulated real-ear conditions, this technique provides limited accuracy due to the inherent asymmetry of the directional response of a manikin-mounted aid. To address this limitation, Working Group 48 of the American National Standards Institute (ANSI) Accredited Standards Committee S3 on Bioacoustics worked on the development of standardized procedures with sound source locations out of the horizontal plane. One procedure given consideration specifies sound source locations on the surface of an icosahedron. The revision of ANSI Standard S3.35 eventually developed by this working group specifies 48 sound source locations on a spherical surface centered about the point on the manikin bisecting the line joining the centers of the ear canals. Details regarding the test setups and procedures used at the National Institute of Standards and Technology to support this development effort will be discussed. Data obtained with the different techniques will also be presented. [Work supported by the Department of Veterans Affairs.]

4:00

1pEA10. Practical aspects of *in situ* directional measurements. Oleg Saltykov (Siemens Hearing Instruments, Inc., 10 Constitution Ave., Piscataway, NJ 08855)

This paper describes practical aspects of measuring *in situ* directionality in simulated diffuse field. The method is based on exposing the object to a sequence of acoustic stimuli coming from a number of sound sources located on a surface of a test sphere. Empirical data are presented that describe how the number of elements in the spherical array affects the measurement errors. Construction of a spherical array is described that simplifies the test procedure and assures measurement errors less than 0.1 dB. A test setup for fast automated measurements of *in situ* directionality has been developed. The time delay spectrometry (TDS) method has been chosen for directional measurements because of its ability to improve signal-to-noise ratio. Another big advantage of the TDS method is its ability to filter out some undesirable acoustic reflections, therefore improving anechoic conditions in the test space. It should be possible to conduct *in situ* directional tests in a large, nonanechoic room with the TDS method.

4:10

1pEA11. Directivity benchmarks using an automated three-dimensional scanning system. Thomas Burns (Starkey Labs., 6600 Washington Ave. S., Eden Prairie, MN 55344)

In clinical studies relating a patient's perception in noise, it is interesting to note that the hearing industry has used exclusively the Directivity Index as an objective performance benchmark for the hearing aid. Considering, for example, that a dipole directional pattern has the same DI as a cardioid pattern, it is reasonable to require that additional directional performance benchmarks be reported in these clinical studies, along with the room acoustics parameters related to noise/source positions and the relationship between direct and reverberant fields. The purpose of this study is to describe an automated 3-D scanning system for benchmarking directional performance, and to review the traditional repertoire of directional benchmarking that has been used in the broader engineering acoustics community; namely, the null angle, maximum response angle, random energy efficiency, front to total random energy ratio, distance factor, and omni to directional array gain. Lastly, visualization of 3-D polar responses will be explored.

4:20

1pEA12. Measurement and visualization of three-dimensional directivity pattern. Georg-Erwin Arndt, Anton Gebert, Harald Klemenz, and Hartmut C. Ritter (SIEMENS Audiologische Technik, Gebbertstrasse 125, D-91058 Erlangen, Germany)

In order to optimize a new second-order multimicrophone technology for a KEMAR dummy head, a three-dimensional directivity measurement setup was developed. To minimize mechanical mass and to reduce total measurement time a C-Bow setup was used, containing 18 calibrated loudspeakers. Those small tweeters identical in construction are placed in every 10 deg of elevation in a semicircular arc of 2-m diameter. The only moving part of this setup is a full-circle rotating KEMAR. The ANSI Standard 3.35 for directional measurement is fully supported and the required 48 measuring points are completed in less than 3 min. Using this fast and simple setup, the various responses attained from different latitudes need to be weighted to calculate a three-dimensional directivity value. Utilizing an equally distributed number, for example 400 measuring points easily executable with this setup, weighting can be omitted and a three-dimensional plot with high resolution can be visualized. Additionally, two-dimensional cuts of different planes in horizontal, vertical, and sagittal direction can be displayed. Data of unaided KEMAR, as well as data from the hearing aid used during those measurements, are presented and discussed.

4:30

1pEA13. Measurement of directivity index by three-dimensional spatial sampling. Stephen D. Julstrom (Etymotic Res., 61 Martin Ln., Elk Grove Village, IL 60007, SJulstrom@comcast.net)

The traditional method for finding the directivity index (DI) of a microphone by extrapolation from its two-dimensional polar pattern may not be accurate when the microphone or its mounting surface is acoustically large. *In situ* hearing aid microphone DI determination requires knowledge of the full three-dimensional directional response. Microphone DI may be measured directly by comparison of the on-axis anechoic sensitivity to the sensitivity in a diffuse sound field, normalizing the comparison by reference to a perfect omnidirectional microphone. In practice, creating an accurately diffuse field is difficult, and is traditionally accomplished by averaging measurements taken at several locations in a reverberant room excited by multiple uncorrelated sound sources. Theoretically equivalent results can be obtained by three-dimensional spatial sampling in an anechoic environment, inferring the full spherical response from a finite number of directional measurements. The necessary density, positioning, and positioning accuracy of measurement directions is determined by sampling theory, tests employing mathematically defined polar patterns, and practical considerations, leading to a minimum specification of 48 measurement points arranged in five horizontal zones. Non-uniform spacing of these points necessitates employing a weighted average of the sensitivity measurements to obtain the best prediction of the actual diffuse field response.

4:40

1pEA14. Round robin measurements of 3-D directivity using the ANSI S3.35 protocol. Thomas Burns (Starkey Labs., 6600 Washington Ave. S., Eden Prairie, MN 55344)

In December 2004, the 1985 version of ANSI S3.35 entitled Method of Measurement of Performance Characteristics of Hearing Aids under Simulated Real-ear Working Conditions was revised. An important addition to this revision included a process to estimate the Directivity Index of a hearing aid, when mounted on a mannequin, using spatial scans of a loudspeaker in three dimensions. In an effort to acquire empirical data with the new protocol, the ANSI S3.35 working committee decided to conduct a set of round robin measurements during the summer and autumn of 2004. Two hearing aids were circulated to seven different facilities. Three-dimensional, *in situ*, directivity measurements were taken on a KEMAR mannequin for each of the custom aids, and also for the mannequin without aids. Results were exceptional, with a standard deviation much less than 1.0 dB for 1/3 octave bands from 500 Hz to 8 kHz. Additional details of the results will be discussed, as will sources of systematic error encountered during the round robin testing.

4:50–5:15

Open Discussion

Session 1pEDa

Education in Acoustics: Acoustics Demonstrations I

Andrew C. H. Morrison, Chair
Northern Illinois Univ., Physics Dept., 202 Faraday West, De Kalb, IL 60115

Contributed Papers

1:00

1pEDa1. Introducing phonetics students to spectral components of vowel-like sounds. Geoffrey Stewart Morrison (Dept. Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada, gsm2@ualberta.ca)

Undergraduate students in phonetics classes typically have difficulty understanding the concept behind Fourier analysis: that vowels can be decomposed into a series of spectral components. The concept can be introduced to students beginning with its inverse: that vowel-like sounds can be constructed from a series of sine waves. The demonstration consists of a PRAAT script which students are given to work with as a homework assignment before vowels are covered in class (PRAAT is free cross-platform software). The students are asked to input three frequencies, the script plots the time-domain representation of each sine wave and the sum of the three sine waves, and the frequency-domain representation of the sum of the three sine waves. It also plays the three sine waves and the sum of the three sine waves. Instructions include suggestions of frequencies for the students to try, and ask which vowels the results sound most like. Instructions also ask students to experiment with different frequencies to try to make sounds similar to other vowels. The students gain hands-on experience with vowel-like synthesis in order to give them an intuitive sense of the spectral components of vowel-like sounds before the theoretical concepts are introduced in the classroom.

1:15

1pEDa2. A speech-perception training tool to improve phonetic transcription. Noelle R. Padgitt, Benjamin Munson, and Edward J. Carney (Dept. Speech Lang. Hear. Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, Padg0010@umn.edu)

University instruction in phonetics requires students to associate a set of quasialphabetic symbols and diacritics with speech sounds. In the case of narrow phonetic transcription, students are required to associate symbols with sounds that do not function contrastively in the language. This learning task is challenging, given that students must discriminate among different variants of sounds that are not used to convey differences in lexical meaning. Consequently, many students fail to learn phonetic transcription to a level of proficiency needed for practical application (B. Munson and K. N. Brinkman, Am. J. Speech Lang. Path. [2004]). In an effort to improve students' phonetic transcription skills, a computerized training program was developed to trains students' discrimination and identification of selected phonetic contrasts. The design of the training tool was based on similar tools that have been used to train phonetic contrasts in second-language learners of English (e.g., A. Bradlow *et al.*, J. Acoust. Soc. Am. **102**, 3115 [1997]). It consists of multiple stages (bombardment, discrimination, identification) containing phonetic contrasts that students have identified as particularly difficult to perceive. This presentation will provide a demonstration of the training tool, and will present preliminary data on the efficacy of this tool in improving students' phonetic transcription abilities.

1:30

1pEDa3. Visualizing vowel-production mechanism using simple educational tools. Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ., 7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan, arai@sophia.ac.jp)

To develop intuitive and effective methods for educating Acoustics to students of different ages and from varied backgrounds, Arai [J. Phonetic Soc. Jpn. **5**, 31–38, (2001)] replicated Chiba and Kajiyama's physical models of the human vocal tract as educational tools and verified that the physical models and sound sources, such as an artificial larynx, yield a simple but powerful demonstration of vowel production in the classroom. We have also started exhibiting our models at the Science Museum "Ru-Ku-Ru" in Shizuoka City, Japan. We further extended our model to a lung model as well as several head-shaped models with visible vocal tract to demonstrate the total vowel-production mechanism from phonation to articulation. The lung model imitates the human respiratory system with a diaphragm. In the head-shaped model, the midsaggital cross section is visible from the outside. To adjust the degree of nasopharyngeal coupling, the velum may be rotated. Another head-shaped model with the manipulable tongue position was also developed. Two test results were compared before and after using these physical models, and the educational effectiveness of the models was confirmed. The homepage of the vocal-tract models is available at http://www.splab.ee.sophia.ac.jp/Vocal_Tract_Model/index-e.htm. [Work supported by KAKENHI (17500603).]

1:45

1pEDa4. Demonstrations of simple and complex auditory psychophysics for multiple platforms and environments. Seth S. Horowitz, Andrea M. Simmons (Dept. of Psych., Brown Univ., Box 1853, Providence, RI 02912, shorowitz@neuropop.com), and China Blue (Brooklyn, NY 11211)

Sound is arguably the most widely perceived and pervasive form of energy in our world, and among the least understood, in part due to the complexity of its underlying principles. A series of interactive displays has been developed which demonstrates that the nature of sound involves the propagation of energy through space, and illustrates the definition of psychoacoustics, which is how listeners map the physical aspects of sound and vibration onto their brains. These displays use auditory illusions and commonly experienced music and sound in novel presentations (using interactive computer algorithms) to show that what you hear is not always what you get. The areas covered in these demonstrations range from simple and complex auditory localization, which illustrate why humans are bad at echolocation but excellent at determining the contents of auditory space, to auditory illusions that manipulate fine phase information and make the listener think their head is changing size. Another demonstration shows how auditory and visual localization coincide and sound can be used to change visual tracking. These demonstrations are designed to run on a wide variety of student accessible platforms including web pages, stand-alone presentations, or even hardware-based systems for museum displays.

2:00

1pEDa5. On the use of waveform images to describe the initial response of finite-length waveguides. Jerry H. Ginsberg (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405)

The d'Alembert solution of the wave equation can be adapted to describe reflection from planar boundaries. One technique for doing so images the incident wave on the opposite side of the boundary. This concept has been introduced in a few texts, most extensively by Morse and Ingard [*Theoretical Acoustics*, McGraw-Hill, New York (1964), pp. 106–115], but only for nondissipative ends (infinite or zero impedance.) This paper formalizes the procedure for the case where the boundary has a resistive impedance that is independent of frequency, and then extends it to treat waveguides of finite length. It is shown that the field that results from arbitrary initial conditions can be represented by an infinite number of images. This leads to a representation of the acoustic field as oppositely propagating wave in an unbounded waveguide, with only a limited number of images overlapping at any instant. Both mathematical and graphical descriptions of these waves are derived. In addition to assisting the student to understand the evolution of the field, mathematical analysis of the image construction leads to a number of physical and mathematical insights to fundamental acoustic phenomena. These include the fact that the field in the dissipationless case can be represented as a modal series with associated natural frequencies, and a quantitative understanding of the manner in which the field decays when either end is dissipative. A corollary of the latter analysis is an expression for reverberation time that is remarkably similar to the Norris–Eyring formula. From an instructional viewpoint, the fact that all results are derived without recourse to solving differential equations makes the image waveform concept especially useful as a way of introducing new students to fundamental acoustic phenomena.

MONDAY AFTERNOON, 17 OCTOBER 2005   CONRAD A, 3:00 TO 4:30 P.M.

Session 1pEDb

Education in Acoustics: "The Physics Force"–Bigger is Better, Physics Demonstrations with Emphasis on Acoustics

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Chair's Introduction—3:00

1pEDb1. Demonstrations with Jon Barber, Dan Dahlberg, Jack Netland and Hank Ryan. (Univ. of Minnesota, Inst. of Technology, Minneapolis, MN 55455)

Physics is Fun. If that sounds like an oxymoron to you, accept an invitation to see these demonstrations by the University of Minnesota's The Physics Force. The Physics Force performs the Physics Circus a dozen times a year. The Circus can best be described as a "Mega-Circus" considering the size of the equipment, the amount of humor, and the rapid pace. This demonstration team of award winning physics teachers has performed widely to international acclaim. Although constraints will limit the "Mega" equipment for this performance, the humor and rapid pace will be present in a special performance with an emphasis on acoustics.

MONDAY AFTERNOON, 17 OCTOBER 2005   SALON E, 1:00 TO 3:20 P.M.

Session 1pNCa

NOISE-CON and Noise: Aircraft Source Noise Research

Daniel Sutliff, Chair
NASA Glenn Research Ctr., 21000 Brookpark Rd., Cleveland, OH 44135

1:00

1pNCa1. Acoustics of excited jets—a historical perspective. Clifford A. Brown (NASA Glenn Res. Ctr., 21000 Brookpark Rd., Cleveland, OH 44135)

The idea that a jet may be excited by an external forcing is not new. The first demonstration of a jet responding to external pressure waves occurred in the late 1800s. It was not, however, until the 1950s, with the advent of commercial jet aircraft, that interest in the subject greatly increased. Researchers first used excited jets to study the structure of the jet and attempt to determine the nature of the noise sources. The jet actuators of the time limited the range (size and Mach number) of jets could be excited. As the actuators improved, more realistic jets could be studied. This has led to a better understanding of how jet excitation may be used not only as a research tool, but also as a method to reduce jet noise.

1:20

1pNCa2. An overview of flow control for fan noise reduction. Matthew Langford, (Techsburg, Inc., Blacksburg, VA 24060) R. A. Burdisso and W. Ng (Dept. of Mech. Eng., Virginia Tech, Blacksburg, VA 24061)

The dominant tonal noise source from modern high-bypass-ratio turbofan aircraft engines is due to the impingement of viscous fan rotor wakes onto the downstream stator row. Prior research conducted at NASA

Glenn Research Center's Active Noise Control Fan (ANCF) rig has demonstrated that significant tonal noise attenuation can be achieved by injecting 1.2% to 1.8% of the fan throughflow along a slot in the trailing edge of the rotor. Efforts presented in this paper have focused on reducing the required blowing mass flow while maintaining equivalent noise attenuation levels. Two primary approaches were investigated: blowing in circumferentially non-uniform patterns (i.e., blowing on every other blade), and optimizing the injection scheme itself. The concept of blowing on alternate rotors was experimentally tested in the ANCF rig using NASA's current slot-blown rotors, and improved attenuation efficiencies were found (although the overall attenuation levels were less). Cascade wind tunnel tests of several different injection schemes were conducted, and the best-performing configuration was incorporated into a new rotor design for experimental validation in the ANCF rig. These rotors achieved similar tonal noise attenuation levels as the slot-blown configuration, but they required injecting less than 0.75% of the fan throughflow.

1:40

1pNCa3. Deconvolution approach for the mapping of acoustic sources (DAMAS) applied to enhance turbofan engine duct mode measurements. Robert P. Dougherty (OptiNav, Inc., 10914 NE 18th St., Bellevue, WA 98004)

Inlet fan noise in turbofan engines is frequently characterized in terms of modal amplitude distribution. The modes can be measured nonintrusively using an array of wall mounted microphones. At high frequency, the number of cuton modes can be much larger than any reasonable number of microphones. This suggests a beamforming analysis technique, rather than a least squares fit, as an appropriate method to extract the mode amplitude from the data. The quality of beamforming output is limited by array resolution and sidelobes. The CLEAN deconvolution algorithm has been used in the past to attempt to improve mode maps from beamforming processing. The recently introduced DAMAS algorithm is a much more comprehensive deconvolution technique, and is well suited to processing mode maps. The benefit of DAMAS is illustrated using simulated mode data. The engine inlet is represented as a cylinder with uniform flow. Two array designs are considered. In each case, several mode distributions are postulated and used to create synthetic array data. Beamforming is applied to create mode maps, and DAMAS is used to deconvolve the array response functions from the mode maps. The results are evaluated by comparing the true mode distributions, the beamform maps, and the deconvolved maps.

2:00

1pNCa4. Rotating rake turbofan duct mode measurement system. Daniel. L. Sutliff (NASA Glenn Res. Ctr., Cleveland, OH 44135)

An experimental measurement system was developed and implemented by the NASA Glenn Research Center in the 1990s to measure turbofan duct acoustic modes. The system is a continuously rotating radial microphone rake that is inserted into the duct. This rotating rake provides a complete map of the acoustic duct modes present in a ducted fan and has been used on a variety of test articles: from a low-speed, concept test rig, to a full-scale production turbofan engine. The rotating rake has been critical in developing and evaluating a number of noise reduction concepts as well as providing experimental databases for verification of several aero-acoustic codes.

2:20

1pNCa5. Status of duct liner technology for application to aircraft engine nacelles. Tony L. Parrott, Michael G. Jones, and Willie R. Watson (NASA Langley Res. Ctr., M.S. 463, 2 N. Dryden St., Hampton, VA 23681)

The peculiar technical requirements that constrain the design of acoustic liners used to suppress turbo-machinery noise emission from commercial aircraft engine nacelles continue to make the design of such liners a challenge. The need for increased liner efficiency and suppression band-

width is particularly critical in the face of ever decreasing nacelle wall area available for liner treatments in modern high bypass ratio engines. This paper reviews some of the key advances in the enabling measurement technologies that have been critical to creating and validating impedance models for liner structures suitable for the harsh aeroacoustic environments encountered on the inside of nacelle inlets and exhaust ducts. Specifically, the paper describes recent advances and outstanding issues related to determining liner impedance beneath high speed grazing flows combined with high sound levels. Evolution of this technology has resulted from the need for more reliable/accurate impedance determinations, and from the impetus provided by modern data acquisition systems and computational methodologies. Outstanding issues include discrepancies between impedances as educed in different laboratories under similar aeroacoustic conditions. For single layer, perforate over honeycomb type structures, these discrepancies are mainly confined to the normalized resistance component and may be as large as 0.50.

2:40

1pNCa6. Source localization for active control of turbofan rotor-stator broadband noise. Bruce E. Walker (Hersh Walker Acoust., 780 Lakefield Rd., Unit G, Westlake Village, CA 91361)

In order to identify a reference signal source for an active noise cancellation system, cross-correlation techniques were used to localize broadband noise source regions on exit guide vanes of the NASA Glenn Research Center Advance Noise Control Fan (ANCF). Arrays of surface pressure sensors were imbedded in one guide vane and in the wall of the fan. Synchronous sampling was used with a multichannel data acquisition system to allow removal of periodic components from the signals. The signals were then cross-correlated to assess radiation directivity and the relationship between vane surface pressure and in-duct acoustic noise. The results of these measurements indicated that broadband unsteady pressures near the leading edge tip of the guide vane were well enough correlated with acoustic radiation that 2–3 dB active noise cancellation could be achieved using a simple gain-delay control algorithm and actuator array. After successful simulation in a wind tunnel environment the concept was incorporated on 15 guide vanes and tested in ANCF. Cross-correlation measurements were further used to evaluate system performance and to identify competing noises from rotating and stationary sources within the fan.

3:00

1pNCa7. Modal decomposition method for high-frequency acoustic impedance testing. Todd Schultz, Louis Cattafesta, and Mark Sheplak (Univ. of Florida, Gainesville, FL 32611-6250, cattafes@ufl.edu)

Accurate duct acoustic propagation models are required to characterize and reduce aircraft engine noise. These models ultimately rely on measurements of acoustic impedance for candidate materials used in engine nacelle liners. This paper seeks to increase the frequency range of acoustic impedance testing by extending the standard two-microphone method (TMM), which is limited in bandwidth to ranges where only plane waves propagate, to include higher-order modes. The modal decomposition method (MDM) presented includes the first four normal modes in the model of the sound field and thus increases the frequency range from 6.7 to 13.5 kHz for a 2.54- by 2.54-cm waveguide. To resolve both the incident and reflected amplitudes of the first four modes, eight microphone measurements are required. This paper formulates and simulates the MDM and applies it to measurement of the reflection coefficient for various specimens over the tested frequency range of 0.2 to 13 kHz. Mode scattering, predominately from the plane-wave mode into other modes, is revealed. The experimental uncertainty of the method is discussed.

Session 1pNCb

NOISE-CON and Noise: Numerical Methods in Acoustics II

Ron Gerdes, Cochair
3M, 3M Ctr. Bldg. St. Paul, MN 55144

Ravi Raveendra, Cochair
3830 Packard, Ann Arbor, MI 48108

1:00

1pNCb1. Analytic, numerical and experimental techniques to formulate four-pole matrices for hybrid system modeling. Prasad Kadam and Jay Kim (Mech., Industrial and Nuclear Eng., Univ. of Cincinnati, 589 Rhodes Hall Cincinnati, OH 45221-0072)

Four-pole matrix is a very convenient concept in modeling composite acoustic systems; however, its application has been largely limited to the systems composed of one-dimensional and lumped parameter components as four poles are defined in terms of the acoustic quantities at points. In this work, four-pole parameters of a three-dimensional rectangular cavity are formulated by analytical, numerical and experimental methods according to a general procedure that utilizes the pressure response functions. Results from three methods are compared and confirmed to be very close to one another, which validates the procedure. Therefore, the technique allows formulating the equation of a complex acoustics system by assembling sub-system equations obtained by applying any method, experimental, analytical or numerical, most suitable to modeling the sub system. Some practical issues that arise in the modeling process such as source modeling, convergence problems and measurement techniques are explained. The procedure developed in this work enables modeling of practically any acoustic system, regardless of its size, shape, and complexity, as long as the system has definable input and output ports.

1:20

1pNCb2. A case study comparing 1-D and 3-D analytical modeling methods for vehicle intake system design. Allan C. Aubert, Ed Green, Jennifer Bastiaan and Thomas Leclercq (Roush Industries, Inc., 12011 Market St., Livonia, MI 48150)

There is intense competition among automakers to create ever-quieter vehicles and powertrains. Exterior and interior noise of many vehicles is significantly influenced by noise coming from the engine intake system. In order to address this source, significant effort needs to be expended on the noise design of an engine's intake system. Cost and time constraints tend to make intake designers ignore many complexities in their modeling efforts, specifically in assuming rigid walls, and using 1D model methods. This paper, presents a case study comparing results for a particular intake system between a 1D model, the 3-D boundary element analysis (BEA) method, both coupled and uncoupled to the structure and the new ACT-RAN code by fft/MSC of analyzing coupled structure/fluid systems. The structure accuracy of the result is compared to measurements of a physical prototype. Costs and time required are compared, as well as some of the modeling issues encountered and the relative accuracy of each method.

1:40

1pNCb3. Determination of shell-radiated noise of an automotive engine air intake system using numerical and experimental techniques. D. J. Moenssen (Visteon Corp., One Village Center Dr., Van Buren Township, MI 48111)

Automotive engine air intake noise affects the quality of the total vehicle sound. The degree to which engine air intake noise contributes to the total vehicle sound is controlled by the relative contributions of the airborne, structure-borne, and shell-radiated noise components. Proper tuning of intake noise requires an understanding of how each of these noise sources is generated. In addition, accurate methods for modeling and measuring each noise source are required for efficient development of an acoustically tuned engine air intake system. This paper describes a numerical method that simulates shell-radiated noise due to the coupling between the internal pressure pulsations of an engine air intake system and the structural properties of the intake system. Techniques for tuning the acoustic and structural properties of the numerical models are illustrated. In addition, a hybrid-experimental technique for accurate measurement of shell-radiated noise is proposed. The hybrid measurement technique is also used for further validation of the numerical method. The numerical and experimental techniques are designed to determine the shell noise of vibrating components in the presence of background noise, as is typically the case for engine air intake systems. A duct and sidebranch Helmholtz resonator is used to validate both methods.

2:00

1pNCb4. Vibration analysis of beams treated with passive constrained layer damping using spectral element method. J.-H. Song and S.-Y. Hong (Dept. of Naval Arch. and Ocean Eng., Seoul Natl. Univ., Seoul 151-742, Korea)

This paper introduces a spectrally formulated finite element method (SFEM) for the beams treated with passive constrained layer damping (PCLD). The viscoelastic core of the beams has a complex modulus that varies with frequency. The SFEM is formulated in the frequency domain using dynamic shape functions based on the exact displacement solutions from progressive wave methods, which implicitly account for the frequency-dependent complex modulus of the viscoelastic core. The dynamic responses obtained by the spectral element method and the conventional finite element method (CFEM) are compared to evaluate the validity and accuracy of the present spectral PCLD beam element model. The spectral PCLD beam element model is found to provide very reliable results when compared with the conventional finite element model.

2:20

1pNCb5. Application of acoustic radiosity methods to noise propagation within buildings. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., 3201 S. Dearborn St., Chicago, IL 60616) and C. Walter BeamerIV (Univ. of Colorado, Boulder, CO 80309)

The prediction of sound pressure levels in rooms from transmitted sound is a difficult problem. The sound energy in the source room incident on the common wall must be accurately predicted. In the receiving room, the propagation of sound from the planar wall source must also be accurately predicted. The radiosity method naturally computes the spatial distribution of sound energy incident on a wall and also naturally predicts the propagation of sound from a planar area source. In this paper, the application of the radiosity method to sound transmission problems is introduced and explained.

1865   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1865**

2:40

1pNCb6. Acoustic scatter from a submerged target in a waveguide.
Joseph M. Cuschieri (Lockheed Martin, M.S. 2, Perry Technologies, Riviera Beach, FL 33404), Emmanuel Charmes, and Pierre-Phillip Beaujean (Florida Atlantic Univ., Boca Raton, FL 33431)

In very shallow water or bounded acoustic medium (waveguide), the acoustic scattering from a target located in the waveguide is influenced by the presence of the boundaries. This is due to the multiple reflections occurring between the target and the waveguide boundaries. In this paper, the influence of boundaries on the acoustic scattering from a rigid sphere is investigated. A superposition method with equivalent point sources located inside the surface of the target and calibrated by satisfying the boundary conditions on the surface of the target is used to model the problem. The target is insonified by an equivalent plane acoustic wave traveling parallel to the waveguide boundaries. The presence of the boundaries breaks the symmetry in the scattering patterns and influences the scattering strength of the target. At particular frequencies or scatterer depths, a significant change in the magnitude and spatial distribution of the scattered field occur due to strong coupling between the target and the boundary. It is concluded that the scattering in VSW can be significantly different from that of the same target in an unbounded medium.

3:00

1pNCb7. Time-domain calculations of attenuative and dispersive sound propagation in porous media. D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03755-1290, d.keith.wilson@erdc.usace.army.mil), Vladimir E. Ostashev (NOAA/Environ. Technol. Lab., Boulder, CO 80305), Sandra L. Collier, David H. Marlin (U.S. Army Res. Lab., Adelphi, MD 20783), David F. Aldridge, and Neill P. Symons (Sandia Natl. Lab., Albuquerque, NM 87185)

Previously, a set of time-domain equations for broadband sound propagation in rigid-framed porous materials [Wilson *et al.*, J. Acoust. Soc. Am. **116**, 1889–1892 (2004)] was derived by modeling viscous and thermal diffusion in the pores as relaxational processes. In this paper, the time-domain relaxational model is shown to be analogous to a phenomenological model previously proposed by Zwikker and Kosten, when the resistance constant, structure constant, and compressional modulus are allowed to be slowly varying functions of frequency. An algebraic inversion of the structure constant and compressional modulus is described that leads to a state-equation system amenable to numerical implementation. Attenuation and dispersion are described by convolution integrals between the acoustic fields and the relaxation functions. Numerical techniques are formulated that discretize the convolution integrals and reduce to known results in low- and high-frequency limits. The techniques are implemented in finite-integro-difference, time-domain calculations of sound interactions with absorbing porous panels and outdoor propagation scenarios involving hills and barriers.

MONDAY AFTERNOON, 17 OCTOBER 2005

SALON A, 1:20 TO 4:40 P.M.

Session 1pNCc

NOISE-CON and Noise: Measurement of Information Technology Product Noise Emissions II

Marco Beltman, Cochair
Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124

Terry W. Baird, Cochair
Hewlett-Packard Co., 11311 Chinden Blvd., Boise, ID 83714

1:20

1pNCc1. Personal computer, printer, and portable equipment noise in classrooms. Robert D. Hellweg, Jr. (Hewlett-Packard Co., 200 Forest St., Marlborough, MA 01752), Egons K. Dunens (Hewlett-Packard Co., Houston, TX 77070-2698), Terrance Baird (Hewlett-Packard Co., Boise, ID 83714-0123), and John Olsen (Hewlett-Packard Co., Vancouver, WA 98683)

The American National Standard on classroom acoustics (ANSI S12.60-2002) presents criteria for background noise in classrooms. The criteria in ANSI S12.60 do not apply to the noise from portable or permanent built-in equipment used during the course of instruction, such as computers, audiovisual equipment, and printers. ANSI S12.60 provides general guidance on the selection and installation of such equipment; however, the standard does not present specific recommendations for schools to purchase these products to ensure they meet acoustics needs. This paper updates the recommendations made in a NOISE-CON 2003 paper by the authors on an approach for ensuring that computers and instructional material are acceptable acoustically in classroom environments. Schools should use unambiguous purchase specifications based on product A-weighted sound power level criteria. Schools should then install and operate the equipment in a manner to minimize noise emissions—

considering both location of products and students and special noise reduction techniques. This paper recommends a purchase specification format; product sound power criteria based on experience and existing environmental criteria; and procedures for installing and locating such equipment in the classroom to minimize noise.

1:40

1pNCc2. Structural and acoustic noise radiated by compact disk drives. M. J. J. Nijhof, Y. H. Wijnant, A. de Boer (Dept. of Mech. Eng., Univ. of Twente, P.O. Box 217, 7500AE Enschede, The Netherlands), and W. M. Beltman (Corporate Technol. Group, Intel Corp., Hillsboro, OR 97124)

Optical drives inside PCs typically operate at high speed, which may result in significant noise. These drives function both as airborne and structural vibration sources. Three main paths can be distinguished through which noise is emitted to the surroundings: (1) the vibrations of the front of the optical drive emit noise directly into the far field, (2) the sound waves induced by the vibrations of the faces of the optical drives inside the enclosure excites the enclosure, and (3) the structural path from the mounting points of the optical drive to the enclosure. The techniques

1866   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1866**

used to determine the contribution of each path are described and the results of an experimental setup are presented in this paper. The contribution of the structural path is determined by comparing the results of a normal setup with the result of a setup for which the structural path is eliminated by mounting the CD drive on a support that is structurally uncoupled from the PC. Direct measurements of the pressure with a scanning microphone were made. Also, a reconstruction of the pressure field using the boundary element method based on the measured surface velocities of the main radiating surfaces was made.

2:00

1pNCc3. Audibility threshold spectrum for prominent discrete tone analysis. Ikuo Kimizuka (Shimotsuruma 1623-14, Kanagawa-ken, Yamato-shi, ZIP 242-8502, Japan)

To evaluate the annoyance of tonal components in noise emissions, ANSI S1.13 (for general purposes) and/or ISO 7779/ECMA-74 (dedicatedfor IT equipment) state two similar metrics: tone-to-noise ratio (TNR) and prominence ratio(PR). By these or either of these two parameters, noise of question with a sharp spectral peak is analyzed by high resolution FFF and classified as prominent when it exceeds some criterion curve. According to present procedures, however this designation is dependent on only the spectral shape. To resolve this problem, the author proposes a threshold spectrum of human ear audibility. The spectrum is based on the reference threshold of hearing which is defined in ISO 389-7 and/or ISO 226. With this spectrum, one can objectively define whether the noise peak of question is audible or not, by simple comparison of the peak amplitude of noise emission and the corresponding value of threshold. Applying the threshold, one can avoid overkilling or unnecessary action for noise. Such a peak with absolutely low amplitude is not audible.

2:20

1pNCc4. Development of noise emission measurement specifications for color printing multifunctional devices. Ikuo Kimizuka (Shimotsuruma 1623-14, Kanagawa-ken, Yamato-shi, 242-8502, Japan)

Color printing (including copying) is becoming more popular application in home, as well as in offices. Existing *de jule* and/or industrial standards (such as ISO 7779, ECMA-74, ANSI S12.10 series, etc.), however, state only monochrome patterns, which are mainly intended for acoustic noise testing of mechanical impact type printers. This paper discusses the key issues and corresponding resolutions for development of color printing patterns for acoustic noise measurements. The results of these technical works will be published by JBMS-74 (new industry standard of JBMIA within 2005), and hopefully be the technical basis of updating other standards mentioned above. This paper also shows the development processes and key features of proposed patterns.

2:40

1pNCc5. Background noise levels in PC home environments. Rina Doherty, Eric Salskov, Philip J. Corriveau, Paul Sorenson, Doug Gabel, and Willem M. Beltman (Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124)

A study was designed and conducted to determine the background noise levels in the home environment. This is an important factor in determining the acoustic performance of the computing devices that go into these environments. A specialized methodology was developed and measurements were carried out in homes in the United States, Sweden, Germany, and China. The sound levels in three rooms in 15–18 homes in each country were collected over 24-h periods. The results indicated that the background noise levels ranged between 30 and 40 dBA across the four countries. Variations in the background noise levels between the different rooms in the homes were minimal. No significant variations were found between home types (detached, semi-detached, and apartment) and community types (urban, suburban). However, European homes were quieter than United States and Chinese homes. The variations between countries

were statistically significant. In addition to the background noise levels, the acoustical characteristics of the rooms were measured. The results indicated that the reverberation radius was typically below 1 m and was fairly consistent between geographies.

3:00–3:20   Break

3:20

1pNCc6. Radiated noise from structural vibrations of a PC. Willem M. Beltman (Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124)

Computer systems typically contain several noise sources like fans, hard drives, and optical drives. The noise emission of these devices can be characterized on the component level, for example with sound power measurements. For fans, the backpressure effect due to chassis airflow impedance can be characterized to capture an installation effect. However, the induced vibration levels and resulting radiated structural noise when sources are placed in a system are less well understood. This paper focuses on a quantification of the radiated noise from enclosure vibrations due to vibration of a source. The surface acceleration distribution on the enclosure panels was measured. This measured vibration pattern was then used to calculate the radiated noise with several models: the Rayleigh integral model, the radiation mode model, and finally a boundary element model. The panel radiation noise was compared to the measured total noise level. This allowed a quantification of the enclosure vibration noise contribution. The results indicated significant contributions at the rotational frequency and its harmonics. The chassis enclosure resonances were apparent.

3:40

1pNCc7. Sound quality parameters of an axial fan. Mark A. Trautman and Willem M. Beltman (Intel Corp., 2111 NE 25th Ave., Hillsboro, OR 97124)

In addition to meeting sound level criteria, an increasing emphasis is placed on the sound quality of systems and fans. In previous studies, numerical techniques were developed to predict noise from small cooling fans and a specialized experimental setup was presented for flow visualization. This paper outlines the next step: an experimental setup to quantify the sensitivity of several sound quality metrics to fan design parameters. The key metrics included in this study were sound pressure, loudness, tonality, sharpness, and roughness. In addition, the spatial variation of these metrics was characterized and empirical correlations between metrics were extracted. As an example, the experimental setup was used to characterize the design parameters that affect the sound quality of a small axial fan.

4:00

1pNCc8. Application of noise source identification techniques to information technology equipment. Roger Upton and Gijs Dirks (Bruel & Kjaer Sound and Vib. Measurement A/S, Skodsborgvej 307, DK-2850 Naerum, Denmark)

Conventional noise source identification techniques are usually associated with larger mechanical sources such as engines, gearboxes, and the like, having medium to high noise emissions. The IT industry presents a different problem-set, though, in that the noise sources are usually smaller and are relatively low level. However, this does not prevent use of these techniques, in that they can be adapted to measurements on typical IT products. Tools for doing this include the use of non-contacting transducers such as laser vibrometers as references, making array measurements with a small-geometry transducer such as a probe microphone, and using conventional microphone arrays but with reduced grid spacing. This paper examines the application of these tools to noise source identification techniques such as coherent power, intensity mapping, and stationary and non-stationary near-field acoustical holography, such that they can be used on IT equipment. The different methods are illustrated with noise source identification exercises on a laptop PC and a 3 1/2-in. hard-disk-drive. Finally, the paper discusses the application of beam-forming techniques for noise source identification on IT products in general.

4:20

1pNCc9. Study on sound quality evaluation of fan noise. Gaku Minorikawa, Shoji Suzuki, Akio Nagamatsu (Dept. of Mech. Eng., Hosei Univ., 3-7-2, Kajinocho, Koganei-shi, Tokyo, 184-8584, Japan), Hiroshi Irikado (NEC-Toshiba Space System Co., Ltd., Tokyo, Japan), and Takahiro Ito (Oriental Motor Co., Ltd., Tokyo, Japan)

This study is to establish new evaluation scales of sound quality which is more suitable for the human sense of hearing than the conventional evaluation scales such as A-weighted sound-pressure level. The authors focus on fan noise and have investigated the meaning space of its sound quality by hearing tests and factor analyses. The hearing tests were per-formed using semantic differential (SD) method and their test results were analyzed to extract dominant factors of the fan noise sound quality. The factor analyses, influence of noise level and 1/3 octave band level on sound quality of fan noise were also investigated. It was found that the sound quality of fan noise had two major factors; ''offensive'' and ''booming-powerful'' factors. The evaluation of hearing test using a paired comparison method was performed in order to investigate the correlation of these two factors with physical quantities that represent noise level and quality such as A-weighted sound pressure level and loudness level. Analysis of hearing tests results revealed that ''offensive'' factor has posi-tive correlation with the physical quantities such as A-weighted sound pressure, but ''booming'' factor has negative correlation with them.

MONDAY AFTERNOON, 17 OCTOBER 2005

SALON C, 1:00 TO 3:00 P.M.

Session 1pNCd

NOISE-CON and Noise: Sound Absorption, Materials, Transmission

J. Stuart Bolton, Chair

Purdue Univ., Ray W. Herrick Labs., 140 S. Intramural Dr., West Lafayette, IN 47907-2031

1:00

1pNCd1. Evaluating the bonding condition of NASA spray on foam insulation (SOFI) using audio frequency sound absorption measurements. Todd A. Thompson and J. Adin Mann III (Mech. Eng., Iowa State Univ., Ames, IA 50011)

The bonding condition of the spray on foam insulation (SOFI) used to insulate the external tank of the NASA space shuttle can be found by using the audio frequency sound absorption coefficient. The ASTM E1050 stan-dard method for sound absorption measurements was used with an open-ended 1-in-diam cast acrylic impedance tube sealed to the SOFI with closed cell PVC foam. Two artificially disbonded locations, measuring 1.0 in. by 5.5 in. by 0.0625 in. and 2.0 in. by 8.0 in. by 0.0625 in., were detected by peaks in the sound absorption coefficient spectrum. The peaks in the sound absorption spectrum between 1000 and 4000 Hz were 25% to 50% higher over disbonded areas when compared to bonded locations. The maximum and minimum sound absorption levels for the foam ranged between approximately 0.1 and 0.3. The entire sample was scanned using the sound absorption peaks as indicators. Samples of 2-in.-thick polysty-rene foam were used with different sized defects at different locations in the foam to relate defect size and location to peaks in absorption coeffi-cient spectrum. [Work supported by NASA under Award No. NAG102098.]

1:20

1pNCd2. Effect of lining anisotropy on sound attenuation in lined ducts. Jeong-Woo Kim, J. Stuart Bolton (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031), and Jonathan H. Alexander (3M Corp., 3M Ctr., St. Paul, MN 55144-1000)

In this paper, the effect of duct-lining anisotropy, with respect to flow resistivity, in particular, is considered. The duct lining is modeled by using a fully poroelastic theory derived from the Biot theory. The lining is con-sidered to occupy only a part of the duct cross section, and is imagined to be infinite in length. When the appropriate solutions for the sound fields within the airway and the anisotropic porous material are substituted into the boundary conditions of the problem, a homogeneous system of equa-tions results. The propagation characteristics of the duct modes can then be found by identifying the conditions under which the determinant of that system of equations is zero. Corresponding experiments have been con-ducted using samples of aviation-grade glass fiber partially lining a square-section, four-microphone standing wave tube. Excellent agreement between measured attenuations and predictions was found. It will be shown that the modal attenuation is primarily sensitive to the flow resis-tivity of the lining in the axial direction.

1:40

1pNCd3. Application of the heterogeneous (HG) blankets for controlling the base structure vibration levels. Ashwini Gautam, C.R. Fuller (Vib. and Acoust. Labs., Virginia Tech, 146 Durham Hall, Blacksburg, VA 24060), and James Carneal (Virginia Tech, Blacksburg, VA 24060)

This work presents an extensive analysis of the properties of the het-erogeneous blankets (HGs) and their effectiveness in controlling the vibra-tion of the base structures. The HG blankets act as a distributed vibration absorbers consisting of mass inhomogeneities inside a layer of porous media (acoustic foam). To asses the effectiveness of these HG blankets in controlling the vibration of the base structure (plate), detailed finite ele-ment (FE) models of the foam, the HG blanket, and the plate have been developed. The foam has been dicretized using the eight node hexahedral elements. The HG blanket model consists of the foam model with point masses attached to the nodes of the elements. The structural (plate) do-main is discretized using four node rectangular plate elements. Each of the FE models has been individually validated by comparing the numerical results with their respective analytical and experimental results. The struc-tural and the HG blanket FE models were then combined into a larger FE model comprised of a base plate with the HG treatment on its surface. The results from this numerical model have shown that there is a significant reduction in the vibration levels of the base plate due to the HG treatment on it.

2:00

1pNCd4. Wave based optimization of distributed vibration absorbers. Marty Johnson and Brad Batton (Vib. and Acous. Labs, Virginia Tech, 143 Durham Hall, VA 24061-0238)

The concept of distributed vibration absorbers or DVAs has been in-vestigated in recent years as a method of vibration control and sound radiation control for large flexible structures. These devices are comprised of a distributed compliant layer with a distributed mass layer. When such a device is placed onto a structure it forms a sandwich panel configuration with a very soft core. With this configuration the main effect of the DVA is to create forces normal to the surface of the structure and can be used at low frequencies to either add damping, where constrain layer damper

treatments are not very effective, or to pin the structure over a narrow frequency bandwidth (i.e., large input impedance/vibration absorber approach). This paper analyses the behavior of these devices using a wave based approach and finds an optimal damping level for the control of broadband disturbances in panels. The optimal design is calculated by solving the differential equations for waves propagating in coupled plates. It is shown that the optimal damping calculated using the infinite case acts as a good "rule of thumb" for designing DVAs to control the vibration of finite panels. This is bourn out in both numerical simulations and experiments.

2:20

1pNCd5. An environmental and economical solution to sound absorption using straw. Courtney McGinnes, Mendel Kleiner, and Ning Xiang (Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, mcginc@rpi.edu)

The growing attentiveness to using environmentally friendly materials in the building construction industry as a whole has led many architects to research the use of natural materials. Natural fibers, such as straw, are advantageous due to their accessibility, ease of processing, and their ability to be discarded with a lesser environmental impact through biodegradability. While the material focus may have shifted, the need for quality acoustic environments has remained the same. In a set of preliminary studies, the absorption coefficients of sound absorbers using natural fibers have been proven to have comparable values relative to other highly ab-

sorptive, nonenvironmentally friendly materials such as mineral wool and fiberglass. The absorption coefficients were evaluated for octave and third-octave frequencies (125 to 4000 Hz) using the impedance tube method with one microphone. While these absorbers may be acoustically effective as well as environmentally friendly, there may be potential concerns in using natural fibers such as their life span, fire rating, and potential health risks, i.e., disposal techniques, allergenic reactions, and insect and fungus infestation.

2:40

1pNCd6. Rating of sound insulation associated to different climates in Spain. Dwellings located in Mediterranean subtropical climate. Leonardo Meza Marín and Manuel Recuero López (Div. I+D Acustica, INSIA. E. T. Superior de Ingenieros Industriales, Universidad Politecnica de Madrid, Madrid, Spain, lmeza@insia.upm.es)

The purpose of this paper is to correlate the sound insulation of dwellings with its location in different climatic zones existent in Spain. Dwellings situated in a region with Mediterranean subtropical climate were analyzed in this stage. Fifty dwellings situated in Malaga, south of Spain (36 43N, 4 25W) were selected. These dwellings are representative of the type of construction used in this area. To characterize the airborne conduction of noise, 150 measurements of airborne sound insulation of facade were carried out, following ISO 140-5 standard. In living rooms and bedrooms, R45w was obtained, following ISO 717-1 standard. Finally, a correlation was found between the apparent sound insulation and the different type of construction of the different type of facade.

MONDAY AFTERNOON, 17 OCTOBER 2005

DULUTH ROOM, 1:00 TO 4:30 P.M.

Session 1pNCe

NOISE-CON and Noise: Critical Assessment of Technology Available for Power Plant Noise Prediction and Control

Frank H. Brittain, Chair
Bechtel Corp., 3011 Triumph Dr., Alameda, CA 94501

1:00

1pNCe1. Identification and prediction of piping system noise. Allen Fagerlund, Fred Catron (Fisher Controls Intl., 205 South Ctr. St., Marshalltown, IA, 50158), Denis Karzcub (SVT-Eng. Consultants, Leederville, Western Australia, 6902, Australia), and Tucker Martin (Fisher Controls Intl., Jacksonville, FL 32218)

In a power plant environment, the piping systems form a network that extends throughout the facility. Various components can be major sources of in-plant and community noise, both the pipe and the contained fluid can be propagation paths, and radiation can occur from the external surface. Though the basics of these phenomena may be understood, the translation into workable predictive tools has been slow. Too often, source, propagation, and radiation effects are treated as separate entities without allowing for the interactions that exist. The spectral characteristics of the sources will govern the type of response by the system. Pipe transmission loss models will change in pipe to duct transition areas. The status of these concepts will be reviewed with a discussion of current and possible future efforts to improve existing models.

1:30

1pNCe2. Assessment of technologies for predicting insertion loss and directivity of power plant exhaust systems. Jim R. Cummins (Cummins Consulting Services, Inc., P.O. Box 645, Panacea, FL 32346) and Marshall Loewenstein (ESI Group-North America, Bloomfield Hills, MI 48304)

Much technology exists for predicting the insertion loss, directivity and radiation of various components in a power plant. For example, the ASHRAE noise guide gives the IL and/or directivity for many duct configurations. Unfortunately, there are also many components or geometries for which there is currently no practical prediction methodology. The state of the technology for predicting insertion loss, directivity and sound radiation by power plant components, especially ducted sources, such as intake and exhaust, is reviewed with emphasis on modeling techniques and verification. Several cases where the normal prediction methods are both adequate and inadequate are presented. Suggestions are given as to methods and/or future development that could provide more accurate, reliable, or useable results.

2:00

1pNCe3. Critical assessment of thermal/acoustical insulation for pipe and duct noise. Marlund E. Hale (Adv. Eng. Acoust., 663 Bristol Ave., Simi Valley, CA 93065)

Piping and duct networks are among the major sources and distributors of noise in power plants. This noise is generated primarily by large rotating equipment and devices, primarily valves, with a large pressure drop. Piping and duct noise controls include silencers, thermal/acoustical insulation (lagging), and special treatments (e.g., low noise valve trims and diffusers). The effectiveness of acoustical lagging for pipe and duct noise depends on the type of source, coupling of the piping to sources or support structures, and the frequency spectra radiated. The acoustical performance of pipe lagging systems has been reported by a number of researchers over the past several decades. Today both ASTM and ISO have issued testing standards for the evaluation of pipe lagging systems, and the ISO standard also provides pipe lagging design guidance. Current technology for measuring the TL of lagging, predicting pipe and duct noise, and predicting the IL of lagging is critically assessed. This paper reviews the methods and results of the two standards and compares several laboratory test results with field measurements. Some challenges are presented along with suggestions for improving prediction of piping and ductwork noise abatement within power plants and in the community.

2:30–3:00   Break

3:00

1pNCe4. Critical assessment of indoor noise propagation and prediction in power plants. Frank H. Brittain (Bechtel Corp., 50 Beale St., San Francisco, CA 94105)

Accurate prediction of indoor noise propagation in power plants is important to help estimate occupational noise exposures, and to help predict community noise radiated by plant walls—from levels predicted just inside of each wall. Unfortunately, the basic theories of room acoustics are not applicable. Most power plant rooms are both too large, and too odd shaped for basic room theory, including the Sabine and Norris-Erying theories, to be applicable. Even more important, basic room theory requires empty rooms, and power plant spaces are densely packed with equipment, piping, cable trays, etc. (called fittings). This paper reviews basic room theory, and outlines deficiencies for use in predicting noise propagation inside power plant buildings. Examples are given of walk-away measurements showing that there is no reverberant field, and that reverberation measurements do not correlate well with walk-away test data. Using measurements as an alternative to levels predicted just inside of plant walls to help predict community noise radiated by each wall are discussed. Software for predicting noise in industrial spaces is identified, and their suitability for power plants, which have unusually high fitting densities, is also discussed.

3:30

1pNCe5. Acoustical performance of power plant enclosure walls—a critical assessment of needs. Robert A. Putnam (Siemens Westinghouse Power Corp., 4400 Alafaya Trail, Orlando, FL 32826) and Noel Frederick (ESI North America, Bloomfield Hills, MI 48304)

The acoustical design of power generating facilities invariably requires knowledge of the transmission loss characteristics of acoustical enclosures, from component specific housings to generation buildings. Many of the composite wall configurations used by vendors are either proprietary or of such recent design that laboratory test data (TL per ASTM E90) are not available. Several factors contribute to uncertainty in the application of test data. At frequencies of 125 Hz and below, all such laboratory based test data should be considered suspect, especially since for plants in which sound emissions are heavily attenuated, the low frequencies tend to become progressively more dominant at far field positions. Thus the far field A-weighted sound levels from power plants, especially those of low noise design, are governed by sound energy in precisely the frequency range for which the TL's for walls and enclosures are least reliable. Further, the size and mounting of power plant walls differs significantly from the same walls mounted in an ASTM E90 test facility. This paper gives a critical assessment of the sources of the uncertainties, suggestions for methodologies to obtain more reliable TL data and discusses the effects to be expected, illustrated by way of selected examples.

4:00–4:30
Panel Discussion

MONDAY AFTERNOON, 17 OCTOBER 2005

SALON C, 3:20 TO 5:40 P.M.

Session 1pNCf

NOISE-CON, Noise and Architectural Acoustics: Speech Issues in Building Environments

Amanda L. Kachur, Cochair
Acoustics By Design, 124 E. Fulton, Grand Rapids, MI 49503

Kenric Van Wyk, Cochair
Acoustics By Design, 124 E. Fulton, Grand Rapids, MI 49503

3:20

1pNCf1. New security and privacy laws require basic changes in professional practice. David M. Sykes (The Remington Group LP, 23 Buckingham St., Cambridge, MA 02138 and Cambridge Sound Management/Acentech, 33 Moulton St., Cambridge, MA 02138)

Everybody knows about HIPAA—but what about GLBA? FIPA? The Patriot Act? Homeland Security? NCLB? FCRA? CASB1? PIPEDA? All of these are recent laws that impact acoustical design. Throw in the American Hospital Association/ASHE and AIA's about-to-be-released ''Guidelines for the Design of Healthcare Facilities'' as well as the redrafting of DCID 6/9 and it looks like time for careful examination of some professional practices relating to security and privacy. Should INCE members join with and endorse the ASA's recently formed Joint TCAA/TCN Subcommittee which aims to fill a policy vacuum in Washington and Ottawa relating to the fundamental protection of citizens' rights to privacy? This group will formulate consistent guidelines to enable federal and state agencies in the US and Canada to enforce and monitor their laws—will their guidelines affect INCE members? Those who advise or give expert testimony to government agencies, defense/security organizations, courts, and large institutions in financial services, healthcare or education likely find themselves in a rapidly shifting landscape and recognize the need to respond with new research and professional practices.

3:40

1pNCf2. The development of specifications and discussion of business models for ensuring speech privacy in the healthcare industry. Timothy Lavallee (LPES, Inc., 14053 Lawnes Creek Rd., Smithfield, VA 23430), Kenneth Good (Acoust. Privacy Enterprises, LLC, Mount Joy, PA 17552), and David Sykes (Acentech/CSM Technol., Cambridge, MA 02138)

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) was developed, among other reasons, to hold healthcare providers accountable for the privacy of patient's personal and medical information. It includes language addressing the need for "reasonable safeguards" for speech privacy and oral communication in a healthcare setting. After 50 years of development, speech privacy science and mechanisms are well understood. However, current specifications cannot be directly applied and are not specifically written to address the application of the current acoustical knowledgebase to the health care industry's need for compliance. This is a discussion of the state of existing privacy technology and specifications; the ability and availability of mechanisms currently in the health care industry as a possible route for implementation of the regulation; the state of development of specification to address specifically the industry's needs; and a potential business model for implementation.

4:00

1pNCf3. Acoustical and noise redesign considerations when trying to increase patient privacy while ensuring comfort. Eric Klavetter (Mayo Clinic, 200 First St. SW, Rochester, MN 55905)

An internal assessment was undertaken to understand the flow of patients to ensure comfort and privacy during their health care experience at Mayo Clinic. A number of different prototypes, work flows, and methodologies were utilized and assessed to determine the "best experience for our patients." A number of prototypes ranging from self-check in to personal pagers were assessed along with creating environments that introduced "passive distractions" for acoustical and noise management, which can range from fireplaces, to coffee shops to playgrounds to "tech corridors." While a number of these designs are currently being piloted, the over-reaching goal is to make the patient experience "like no other" when receiving their care at Mayo Clinic.

4:20

1pNCf4. Speech privacy: Beyond architectural solutions. Susan Mazer (Healing HealthCare Systems, 100 W. Grove St., Ste. 245, Reno, NV 89509)

HIPAA regulations have brought unparalleled pressures on healthcare organizations to protect private and confidential information from reaching third parties. Yet, as this paper explains, often in the middle of noisy corridors and waiting rooms, this same information needs to be quickly transferred from physician to nurse to family member to others for the care of patients. Research and examples are presented that show that when families, patients, staff are participating together, although independently, in the same or adjacent spaces, the "caf effect" produces rising noise levels as each person competes to be heard. This threatens the very confidentiality demanded by HIPAA. Solutions to this problem are not easy or completely resolved by engineering or design specifications. This paper makes the case that it is ultimately the culture of a healthcare organization that determines the "sound" of a hospital, and any other organization that battles openness with privacy. It presents and discusses proven solutions to address culture in tandem with architectural and acoustic design interventions.

4:40

1pNCf5. Speech privacy performance of a new hospital and medical office building. Kenneth P. Roy, Kenneth W. Good, Jr., Anita M. Snader, and Sharon K. Hatzel (Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604)

Shortly after the occupation of a new hospital and medical office building, both objective and subjective evaluations of the acoustic performance of these facilities were made. The goals of this work were twofold: first, to survey the occupants' subjective perception of the acoustic environment relative to noise, distractions, speech privacy, etc; and second, to relate the subjective perception to objective measures of noise isolation rating (NIC), background noise (dBA), and speech privacy rating (PI). Knowing the construction details of the walls, ceiling, doors, etc. also allowed a comparison of the measured NIC to the expected STC for each type of construction. In this way it was possible to identify robust architectural systems versus weak systems with inherent flanking and leakage paths.

5:00

1pNCf6. Appropriate and inappropriate uses of classroom amplification. David Lubman (Acoust. Consultant, 14301 Middletown Ln., Westminster, CA 92683) and Louis C. Sutherland (Consultant in Acoust., Rancho Palos Verdes, CA 90275)

Currently, classroom amplifiers are being aggressively advocated as substitutes for good acoustics in small mainstream classrooms. Amplifiers are routinely installed without regard to unoccupied classroom noise levels and reverberation times. Amplifiers are being specified by some school districts as a money-saving alternative to mandating compliance with the ANSI standard on classroom acoustics, S12.60-2002. Manufacturers of portable classrooms and noisy wall mounted HVAC systems have joined in supporting the use of classroom amplifiers, claiming that low (35 dBA) classroom noise levels specified by the ANSI standard are unaffordable and unnecessary given amplifiers. The authors believe that the routine use of classroom amplification is appropriate in very large lecture rooms, in special education classrooms for hearing impaired students, for voice-impaired occupants, and perhaps in certain other limited circumstances. The authors explain why they believe the routine use of amplifiers in small mainstream classrooms is an inappropriate substitute for the good classroom acoustics specified in the ANSI standard.

5:20

1pNCf7. Acoustical renovation of temporary classrooms to meet American National Standards Institute (ANSI) standard. David Lubman (Acoust. Consultant, 14301 Middletown Ln., Westminster, CA 92683)

This exercise demonstrates that key requirements of the ANSI S12.60-2002 on classroom acoustics can be met through renovations at modest cost. Free-standing classrooms at two schools in Southern California were successfully renovated. Noise levels were greatly reduced, and now meet requirements of the ANSI standard. A 1950s style bungalow classroom at the Paradise Canyon School (PCS) in Los Angeles County and a double-trailer classroom at the Wildomar Elementary School (WES) in Riverside County were targeted. The aim was to make classrooms suitable for pupils with severe hearing disabilities. But the methods used apply equally to mainstream classrooms. At both schools the above-ceiling air distribution systems were modified to produce quiet airflow. At PCS, a noisy rooftop heat pump was replaced with a much quieter rooftop heat pump and installation. At WES, a noisy wall-mounted heat pump was replaced with a new unit mounted on an exterior concrete pad. At PCS, exterior noise intrusion was achieved by replacing a poorly performing accordion wall with a permanent wall. Further noise reduction can be achieved by replacing leaky windows. Exterior noise intrusion at WES was obtained by replacing leaky windows. The resulting acoustical improvements are documented. Costs and lessons learned are discussed.

Session 1pNCg

NOISE-CON and Noise: Compressor Noise

Jay Kim, Chair

*Univ. of Cincinnati, Applied Acoustics, Mechanics Lab., Mechanical, Industrial and Nuclear Engineering Dept.,
Cincinnati, OH 45221-0072*

3:40

1pNCg1. Analysis and control of air compressor noise. Wonjoon Song, Shrikant Pattnaik, Wancheng Zhao, Milind Bapat, and Hyunsoo Kim (Mech., Industrial and Nuclear Eng., Univ. of Cincinnati, 589 Rhodes Hall, Cincinnati, OH 45221)

Study of noise characteristics of a small air compressor used in typical construction sites was conducted to develop possible noise reduction strategies of the compressor as a student group project. Operating the compressor at the steady-state condition, the sound power of the tool was measured by using a 10-microphone measurement method in the semi-anechoic chamber. The SPL spectrum from each of the microphones was carefully compared and studied. A sound intensity measurement was conducted to identify major noise sources of the compressor, which identified the electric motor and inlet air pulsation as the major noise sources. Several possible measures were applied for noise reduction of the compressor, which included a newly designed intake muffler, an absorptive liner around the motor area and an improved mounting. The effect of each of these measures was measured one by one first, then in combinations. The resulting changes were measured and discussed in terms of the noise characteristics and noise levels.

4:00

1pNCg2. Study of noise transmission from an air compressor. Subhro Nathak, Anand Puranik, Jeffrey Schut, Lee Wells, and M.D. Rao (Mech. Eng. Dept., Michigan Technolog. Univ., Houghton, MI 49931)

The paper discusses the reduction of noise from a Porter Cable 4-gal, 135-psi air compressor. The objectives were to identify the major sources of noise, implement possible noise control measures, and evaluate their effectiveness. Sound measurements were taken according to a procedure that was developed and based on the standards for power tools. Broadband analysis (1/12 octave band) was done to determine the main sources of noise. Ranking of noise sources was done accordingly. The major source of noise was determined to be the piston cylinder assembly and efforts were taken to reduce this main source of noise. The noise control treatments included manufacturing of new parts, use of gasket made of cork between the housing and the cylinder sleeve to isolate the vibrations, implementing a silencer at the exhaust side of the piston and finally, an enclosure encircling the piston/cylinder assembly. The sound-pressure level measurements of the prototype were compared to the baseline measurements to test the effectiveness of the treatments. The overall sound pressure level was reduced from 101.5 to 94.3 dB with the implementation of enclosure around the piston/cylinder assembly and the silencer at the exhaust side of the piston.

4:20

1pNCg3. Statistical energy analyis of a scroll compressor using Auto SEA and SEAM software. Zheng Yu, William B. Rockwood (Trane Corp., A business of American Standard Companies, 3600 Pammel Creek Rd., Bldg. 12-1, La Crosse, WI 54601), Michael Sanderson, and Denis Blanchet (Esi Group, Bloomfield, MI 48304)

Vibration and internal pressure pulsation propagation in a scroll compressor were modeled using AutoSEA and SEAM software. It was found that the two programs predict very similar results in terms of total dBA levels for the six design alternatives evaluated under three different excitations. In general, however, AutoSEA tended to predict slightly lower responses than SEAM. The main reason for the discrepancy is believed to be differences in how each code treats systems with low modal density. Some appreciable differences were also found for the response level at the bottom of the compressor under certain excitations. Investigation into the cause(s) for that discrepancy in ongoing.

4:40

1pNCg4. Statistical energy analysis for a compact refrigeration compressor. Ji Min Lim, J. Stuart Bolton (Ray W. Herrick Labs, Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907), Sung-Un Park (LG Electro., Inc., 391-2, Gaeum jeong-dong, Changwon City, Gyeongnam, Korea), and Seon-Woong Hwang (LG Electron., Inc., Geumcheon-gu, Seoul, Korea)

Traditionally the prediction of the vibrational energy level of the components in a compressor is accomplished by using a deterministic model such as a finite element model. While a deterministic approach requires much detail and computational time for a complete dynamic analysis, statistical energy analysis (SEA) requires much less information and computing time. All of these benefits can be obtained by using data averaged over the frequency and spatial domains instead of the direct use of deterministic data. In this paper, SEA will be applied to a compact refrigeration compressor for the prediction of dynamic behavior of each subsystem. Since the compressor used in this application is compact and stiff, the modal densities of its various components are low, especially in the low frequency ranges, and most energy transfers in these ranges are achieved through the indirect coupling paths instead of via direct coupling. For this reason, experimental SEA (ESEA), a good tool for the consideration of the indirect coupling, was used to derive an SEA formulation. Direct comparison of SEA results and experimental data for an operating compressor will be introduced. The power transfer path analysis at certain frequencies made possible by using SEA will be also described to show the advantage of SEA in this application.

1872   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1872**

Session 1pNCh

NOISE-CON, Noise and Physical Acoustics: Advances in Military Jet Noise Modeling

Victor W. Sparrow, Cochair

Pennsylvania State Univ., Graduate Program in Acoustics, 316 B Leonhard Bldg., University Park, PA 16802

J. Micah Downing, Cochair

Wyle Laboratories, 2001 Jefferson Davis Hwy., Arlington, VA 22202

3:30

1pNCh1. Nonlinearity in outdoor propagation of high-power jet noise: measurement results. Michael M. James, J. Micah Downing Christopher M. Hobbs (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202), Kent L. Gee (Penn State Univ., State College, PA 16804), J. Micah Downing (Wyle Labs., Arlington, VA 22202), Sally A. McInerny (The Univ. of Alabama, Tuscaloosa, AL 35487), and Victor W. Sparrow (Penn State Univ., State College, PA 16804)

Static run-up and flyover measurements of a high-performance military aircraft were conducted at Edwards Air Force Base, CA, as part of a joint program among Wyle Laboratories, The Pennsylvania State University, and The University of Alabama to develop better noise models for military aircraft. This paper describes the measurement program and the resulting acoustic data. These measurements clearly demonstrate nonlinear source and propagation effects from military jet aircraft. Moreover, these results serve as a comparative data set for the evaluation of propagation models. [Work is supported by the Strategic Environmental Research and Develop Program.]

3:50

1pNCh2. Quantifying nonlinearity in the propagation of noise from military jet aircraft. Kent L. Gee and Victor W. Sparrow (The Penn. State Univ., 202 Appl. Sci. Bldg., University Park, PA 16802)

Because of the high noise levels radiated by military jet aircraft, it has been hypothesized that nonlinearity influences the propagation of the noise. A numerical model, which accounts for second-order cumulative nonlinearity, atmospheric absorption and dispersion, and geometrical spreading, has been developed to propagate jet noise waveforms. Numerical propagation of recorded waveforms from recent static engine run-up measurements demonstrates significant waveform steepening and an accompanying transfer of spectral energy to high frequencies that agrees well with measured spectra. Furthermore, the measured and nonlinearly predicted waveforms are perceived to be significantly "louder" or "more annoying" than linearly predicted waveforms, despite the fact that standard metrics such as overall sound pressure level (with flat-, A-, and C-weighting) and Stevens Mark-VII perceived loudness show little difference between linearly and nonlinearly predicted spectra. The results of this study demonstrate the need for additional investigations with alternate metrics that more closely relate to perceived annoyance or loudness of high-amplitude jet noise. [Work supported by the Strategic Environmental Research and Development Program.]

4:10

1pNCh3. Advanced simulation noise model for modern fighter aircraft. Bruce Ikelheimer (Wyle Labs., 2001 Jefferson Davis Hwy., Ste. 701, Arlington, VA 22202)

NoiseMap currently represents the state of the art for military airfield noise analysis. While this model is sufficient for the current fleet of aircraft, it has limits in its capability to model the new generation of fighter aircraft like the JSF and the F-22. These aircraft's high-powered engines produce noise with significant nonlinear content. Combining this with

their ability to vector the thrust means they have noise characteristics that are outside of the basic modeling assumptions of the currently available noise models. Wyle Laboratories, Penn State University, and University of Alabama are in the process of developing a new noise propagation model for the Strategic Environmental Research and Development Program. Source characterization will be through complete spheres (or hemispheres if there is not sufficient data) for each aircraft state (including thrust vector angles). Fixed and rotor wing aircraft will be included. Broadband, narrowband, and pure tone propagation will be included. The model will account for complex terrain and weather effects, as well as the effects of nonlinear propagation. It will be a complete model capable of handling a range of noise sources from small subsonic general aviation aircraft to the latest fighter aircraft like the JSF.

4:30

1pNCh4. Numerical simulation of jet noise from different jet nozzle geometries. Umesh Paliath (Penn State Univ., 233C Hammond Bldg., State College, PA 16802) and Philip J. Morris (Penn State Univ., State College, PA 16802)

This paper describes the numerical simulation of flow-induced noise from jets with different nozzle geometries. The nozzles considered include axisymmetric and nonaxisymmetric nozzles, such as circular and rectangular. Also the study is extended to examine the differences between noise radiated from nozzles with planar exists and those with nonplanar exist, such as beveled nozzles. The detached-eddy simulation (DES) approach is used to simulate both the jet nozzle internal and external flows as well as the jet plume. This methodology allows the turbulence model to transition from an unsteady Reynolds averaged Navier-Stokes (URANS) method for attached boundary layers to a large-eddy simulation (LES) in separated regions. Thus, it is ideally suited to jet flow simulations when the nozzle is included. Both cylindrical polar and Cartesian coordinate systems are used as the basis for grid generation. The one equation Spalart-Allmaras turbulence model is used to describe the evo! lution of the turbulent eddy viscosity. Dispersion relation preserving algorithms are used for spatial discretization and an explicit 4th order Runge-Kutta scheme is used for time marching. The far-field sound is evaluated using the Ffowcs Williams-Hawkings permeable surface acoustic analogy. This permits the noise to be predicted at large distances from the jet based on fluctuations in the jet's near field. This provides a good compromise between numerical accuracy and computational cost. The results are compared with experimental data for both unheated and heated jet cases.

4:50

1pNCh5. Nonlinear propagation modeling: Guidelines for supporting measurements. Thomas B. Gabrielson, Timothy M. Marston and Anthony A. Atchley (Grad. Prog. in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

In nonlinear propagation, evolution of waveforms is strongly dependent on the characteristics of the waveform. The commonly observed waveform steepening is a coherent process that produces a specific phase relationship between spectral components. High-amplitude military jet noise may show peak pressures five to ten times the root-mean-square

1873 J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005 **1873**

pressure and may have local rise times less than several microseconds. Measurement of these waveforms containing fast rise times requires large bandwidth and high fidelity in phase. If the bandwidth and phase are not preserved then the waveforms can lose precisely those features that distinguish them from linear propagation. Whether such waveforms are used as starting fields for nonlinear codes or as code validation, the impact of limitations in the measurement system must be understood—the simple frequency-response magnitude is insufficient. By generating freely propagating weak-shock waves with rise times comparable to those observed in full-scale jet measurements, the time-domain response (or, equivalently, the magnitude and phase response) was determined for a number of sensor configurations. Taken together with full-scale jet-noise data, these results suggest guidelines for making or using field measurements when those measurements are intended to support nonlinear modeling efforts.

MONDAY AFTERNOON, 17 OCTOBER 2005

ROCHESTER ROOM, 1:00 TO 5:00 P.M.

Session 1pNS

Noise and NOISE-CON: Workshop on Methods for Community Noise and Noise Policy

Brigitte Schulte-Fortkamp, Cochair

Technical Univ. Berlin, Inst. of Technical Acoustics, Secr TA 7, Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair

Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Chair's Introduction—1:00

Invited Papers

1:05

1pNS1. Prediction of psychoacoustic parameters. Klaus Genuit, and Andre Fiebig (HEAD acoustics GmbH, Ebertstrae 30a, 52134 Herzogenrath, Germany, klaus.genuit@head-acoustics.de)

Noise is defined as an audible sound which either disturbs the silence, or an intentional sound that listening to leads to annoyance. Thus, it is clearly defined that the assignment of noise cannot be reduced to simple determining objective parameters like the A-weighted SPL. The question whether a sound is judged as noise can only be answered after the transformation from the sound event into an hearing event has been accomplished. The evaluation of noise depends on the physical characteristics of the sound event, on the psychoacoustical features of the human ear as well as on the psychological aspects of men. The subjectively felt noise quality depends not only on the A-weighted sound-pressure level, but also on other psychoacoustical parameters such as loudness, roughness, sharpness, etc. The known methods for the prediction of the spatial A-weighted SPL distribution in dependence on the propagation are not suitable to predict psychoacoustic parameters in an adequate way. Especially, the roughness provoked by modulation or the sharpness generated by an accumulation of high, frequent sound energy cannot offhandedly be predicted as distance dependent.

1:25

1pNS2. Evaluation of urban scape—a combined approach of acoustics and architecture. Juergen Bauer (Mierendorffstrasse 11, D-10589 Berlin, Germany) and Brigitte Schulte-Fortkamp (Tech. Univ. Berlin, Berlin, Germany)

Related to a concept of integrating acoustics into architectural education a new model of designing will be introduced. Combined noises as well as soundscapes with its specific view points on measurement procedures will be discussed against urban fabrics.

1:45

1pNS3. Model community noise ordinance—A new approach to community based policy. Bennett M. Brooks (Brooks Acoust. Corp., 27 Hartford Tpke., Vernon, CT 06066) and Lawrence S. Finegold (Finegold and So Consultants, Centerville, OH 45459)

This paper describes the progress of work ongoing to develop a model community noise ordinance as a U.S. American National Standards Institute (ANSI) national Standard. The Standard expected to result from the efforts of ANSI S12 (Noise) Working Group 41 will provide guidance to governmental officials, acoustical consultants, and other interested persons on how to develop a community noise ordinance which is appropriate for the existing local circumstances, including issues such as public and government priorities and values, available resources, etc. It will give local communities a technical basis for being able to manage their local sound environment and will be sufficiently flexible to be tailored to the needs of a variety of urban, suburban, and rural communities. The key to the usefulness and effectiveness of this new Standard is that it will provide a "menu" of options and will also provide brief discussions of the tradeoffs involved for decisions that must be made by government officials. It will also emphasize that enforcement of a community noise ordinance is crucial to its success and will provide a means to accomplish this important function.

1874   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1874**

2:05

1pNS4. Soundscape measurements: Moving towards a standard. Robert Kull (Parsons, 5800 Lake Dr., Ste. 101, Norfolk, VA 23502)

Soundscapes are multidimensional and require a more robust examination of the influences and effects of the environment. The spectrum of soundscapes spans the continuum from completely natural environments to entirely urban. The approach has to make assumptions and generalizations of the contribution of anthropogenic and non-anthropogenic sources and to determine which aspects of soundscapes are measurable. Step-by-step procedures on how to fill a model with data, including a basic discussion about the kind of measurements that have to be carried out, will be discussed. There will also be discussion as to whether the same measurement procedure is appropriate with respect to animals and humans and the requirements to develop a standard concerned with soundscape evaluation.

2:25

1pNS5. Noises from combined sources—are there any standards? Brigitte Schulte-Fortkamp (Tech. Acoust., TA 7, Tech. Univ. Berlin, Einsteinufer 25 D-10587 Berlin, Germany)

Combined noises have to be investigated with regard to their effects on the encumbrance of the inhabitants of residential areas. As far as noise encumbrance and its effects are concerend, investigating the level of loudness is not sufficient. Furthermore, it is of central importance to investigate the sounds from different sources in the context of everyday life activities, expectancies and different life situations. Therefore, certain methods of investigation and assessment have to be applied with whose help encumbrances caused by noises from different sources can be analyzed. Combining socioacoustical and psychoacoustical methods gives the opportunity to reveal direct connections between subective understanding and subjective perception. Moreover, with the help of this combination noise can be assessed and characterized; psychoacoustical parameters as well as noise emmission can be analyzed in detail. The question is whether there are any standards with respect to combined sources which meet this difficult assessment. As an example the first draft of a German standard will be discussed.

2:45

1pNS6. Overview of noise policy in California. James Phillips (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618)

This paper provides a general review of noise policy in California including general plan noise elements, municipal code noise ordinances, and the California Building Code of Regulations Title 24. The challenges that an acoustical consultant faces while trying to meet the requirements or, as is often the case, trying to achieve the intent of such policies, will be discussed. What works or doesn't work?

3:05

1pNS7. Biases introduced by the fitting of functions to attitudinal survey data. Paul D. Schomer (Schomer and Assoc., Inc., 2117 Robert Dr., Champaign, IL 61821)

For years researchers have toiled to understand community response to noise environments such as the environment around airports, near highways, or near railroads. Frequently, following the lead of Schultz and others, researchers have attempted to develop a ''dose-response'' relationship by fitting curves through aggregated data that report some measure of community response versus some measure of the noise environment. For example, in 1992 FICON, a U.S. Government interagency group, reviewed the then available totality of attitudinal survey data and developed a functional fit to those data. This paper examines the fit used and shows that the assumptions employed by FICON lead to a systematic under-estimation of the percent highly annoyed in the critical 55 to 75 DNL region. In fact, they underestimated the true percentage highly annoyed by a factor of more than 4 times. FICON predicts 6 percent ''highly annoyed'' at 60 DNL when the true value is 27 percent.

3:25–3:40 Break

3:40–5:00
Panel Discussion

Session 1pSP

Signal Processing in Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: Biomedical Acoustic Signal Processing

David H. Chambers, Chair
Lawrence Livermore National Lab., Electrical Engineering, P.O. Box 808, Livermore, CA 94551-5508

Chair's Introduction—1:55

Invited Papers

2:00

1pSP1. Signal processing to enhance ultrasonic tissue characterization in diagnostic medical imaging. James G. Miller, Kirk D. Wallace, and Mark R. Holland (Dept. of Phys., Washington Univ., St. Louis, MO 63130, james.g.miller@wustl.edu)

The goal of this presentation is to lay the background for identifying potential new contributions to medical diagnosis with ultrasonic imaging that might be provided by advanced signal-processing techniques. The focus of this talk is on ultrasonic tissue characterization, an approach that complements and extends presently available high-resolution imaging. Regional estimates of intrinsic material properties of normal and pathologically altered tissue are obtained from local measurements of attenuation and backscatter derived from the same signals that are used to form the ultrasonic images. Examples are drawn primarily from echocardiography, in part because imaging and characterizing the beating heart present significant challenges. Current echocardiographic images, frequently acquired with harmonic mode (nonlinear) imaging, exhibit substantial resolution of anatomical detail. Nevertheless, the estimates provided by tissue characterization of alterations in tissue properties arising from specific pathologies offer the promise of elevating echocardiographic imaging to a new plane. Among the challenges to be met in order to accomplish this goal are properly accounting for the intrinsic anisotropy of the heart and overcoming limitations that arise as a result of phase and amplitude aberrations associated with propagation through an intrinsically inhomogeneous medium. [Work supported by NIH R37HL40302 and R01 HL72761.]

2:30

1pSP2. Generation and processing of finite-amplitude ultrasonic signals in the context of nonlinear (harmonic) medical imaging. Kirk D. Wallace, Todd M. Krueger, Christopher W. Lloyd, Mark R. Holland (Dept. of Phys., Washington Univ., St. Louis, MO 63130, kirk.wallace@wustl.edu), and James G. Miller (Washington Univ., St. Louis, MO 63130)

The focus of this talk is on the application of finite-amplitude ultrasonic waves in the context of diagnostic imaging. Techniques based on nonlinearly generated harmonic signals (tissue harmonic imaging) have rapidly supplanted linear (fundamental) imaging methods as the standard in two-dimensional echocardiography. Improvements in image quality achieved through the use of nonlinear imaging modes have served as the prime motivating factor that continues to drive this revolution. For many patients, the resulting harmonic images are widely recognized as being of apparently higher quality, and less corrupted by artifacts, than the corresponding fundamental images. The benefits associated with nonlinear imaging have been attributed to a number of potential factors, including reduced sensitivity to phase and amplitude aberration, enhanced beam compactness (narrower mainlobe and reduced sidelobe levels), higher imaging frequencies while suffering only a portion of the corresponding attenuation penalty, and reduced interference from reverberations and off-axis echoes. Although much of the physics describing the nonlinear propagation and distortion of finite-amplitude ultrasonic waves is well understood, there remain many unanswered questions about the details of the physics and rank ordering of the factors underlying the improvements associated with nonlinear (harmonic) imaging.

3:00

1pSP3. Processing ultrasound backscatter to monitor high-intensity focused ultrasound (HIFU) therapy. Peter J. Kaczkowski, Ajay Anand, and Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab, Univ. of Washington, Seattle, WA 98105)

The development of new noninvasive surgical methods such as HIFU for the treatment of cancer and internal bleeding requires simultaneous development of new sensing approaches to guide, monitor, and assess the therapy. Ultrasound imaging using echo amplitude has long been used to map tissue morphology for diagnostic interpretation by the clinician. New quantitative ultrasonic methods that rely on amplitude and phase processing for tissue characterization are being developed for monitoring of ablative therapy. We have been developing the use of full wave ultrasound backscattering for real-time temperature estimation, and to image changes in tissue backscatter spectrum as therapy progresses. Both approaches rely on differential processing of the backscatter signal in time, and precise measurement of phase differences. Noise and artifacts from motion and nonstationary speckle statistics are addressed by constraining inversions for tissue parameters with physical models. We present results of HIFU experiments with static point and scanned HIFU exposures in which temperature rise can be accurately mapped using a new heat transfer equation (HTE) model-constrained inverse approach. We also present results of a recently developed spectral imaging method that elucidates microbubble-mediated nonlinearity not visible as a change in backscatter amplitude. [Work supported by Army MRMC.]

1pSP4. Ultrasonic characterization of soft tissue vibrations based on the two-dimensional Fourier transform. Siddhartha Sikdar, Yongmin Kim (Dept. of Bioengineering, Univ. of Washington, Seattle, WA 98195, ssikdar@u.washington.edu), and Kirk W. Beach (Univ. of Washington, Seattle, WA 98195)

It has recently been demonstrated that soft tissue vibrations in the body, traditionally associated with vascular bruits and cardiac murmurs, can potentially be used for the ultrasonic diagnosis of coronary artery disease and vascular trauma. In this paper, the ultrasonic spectrum of soft tissue vibrations is formulated using the two-dimensional Fourier transform, making full use of the information present in the backscattered ultrasound echoes from vibrating tissue. Parametric simulation studies show that vibrations with amplitude 1 μm may be detected even with tissue velocity of 20 cm/s and acceleration of 5 m/s^2, e.g., during peak cardiac motion. Vibrations with amplitude as low as 0.1 μm may be detected when the tissue acceleration is negligible, e.g., during mid-diastole. Also, it was found that tissue vibrations in a direction transverse to the ultrasound beam can be detected. *In vivo* examples of cardiac wall vibrations in patients with coronary artery disease are presented. Tissue vibrations can provide improved sensitivity over conventional duplex ultrasound since the scattering strength from tissue is significantly higher than that from blood. In addition, detection of tissue vibrations has reduced angle dependency and does not require visualization of the vessel lumen, making the exam less dependent on operator skill.

Note: Separate registration fee required. See page A28

MONDAY EVENING, 17 OCTOBER 2005   SALON E, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture on Diagnostic Imaging in Biomedical Ultrasound

Peter H. Rogers, Chair
Georgia Inst. of Technology, School of Mechanical Engineering, Atlanta, GA 30332-0405

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Diagnostic imaging in biomedical ultrasound. E. Carr Everbach (Dept. of Eng., Swarthmore College, 500 College Ave, Swarthmore, PA 19081-1397)

Since World War II, ultrasonic imaging has evolved from its sonar origins to include techniques borrowed from MRI, image processing, and computer modeling. Although diagnostic ultrasound techniques make increasing use of second-order physics such as tissue nonlinearity, much potentially useful information remains unexploited. In addition, safety and health concerns must be balanced against clinical usefulness. The tutorial will review the fundamental concepts and practices of diagnostic ultrasound imaging, and describe current research efforts underway to extend this modality into new areas.

Session 2aAA

Architectural Acoustics, NOISE-CON and Noise: Special Session in Honor of Cyril Harris

David Braslau, Chair

David Braslau Associates, Inc., 1313 5th St., SE, Minneapolis, MN 55414

Chair's Introduction—10:00

Invited Papers

10:05

2aAA1. Cyril Harris: A biographical sketch. Patricia Kuhl (Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98195, pkkuhl@u.washington.edu)

For over half a Century, Cyril Harris has authored numerous scholarly books and papers and designed concert halls that enhance our enjoyment of the arts. In recognition of these accomplishments, Cyril has been honored with medals and awards that pay tribute to his professional talents, including, but not limited to, the highest awards of the Acoustical Society of America. In this talk, I will attempt to shed some light on Cyril's personal history, the paths that led him to a distinguished career in acoustics, and the people and events that provided inspiration during Cyril's long and highly successful career.

10:25

2aAA2. Acoustical design in architecture. Steven Garrett (Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

Cyril Harris's name is best known to the general public for the outstanding concert halls he has designed. But his impact on reduction of noise and vibration through the application of established fundamental engineering design principles is far greater through the work of others who have benefited from the books (and particularly handbooks) he has written and edited. His first book, co-authored with Vern Knudsen, *Acoustical Design in Architecture*, was published over a half-century ago. It was the first milestone on a continuing quest to explain the physics and engineering of acoustics, as well as noise and vibration control, to those who are responsible for the design of habitable structures. [See the excellent review in J. Acoust. Soc. Am. **22**, 521 (1950).] To expand the utility of this approach, Professor Harris assembled and regularly revised two comprehensive compilations: the *Shock and Vibration Handbook* (now in its fourth edition) and the *Handbook of Acoustical Measurements and Noise Control* (now in its third edition and available from ASA). Those handbooks make the application and the evaluation of required techniques accessible to the entire community of noise control engineers and architects. The presentation will also briefly describe his books on architecture and architectural history.

10:45

2aAA3. Cyril Harris and shock and vibration engineering. Allan G. Piersol (Piersol Engineering Co., 23021 Brenford St., Woodland Hills, CA 91364)

Cyril Harris is well known as an internationally recognized authority in the field of architectural acoustics. However, some may not be aware of the fact that Cyril Harris has also been for many years a leading authority in shock and vibration engineering. He became particularly prominent in this area in 1961 with the publication by McGraw-Hill of the "Shock and Vibration Handbook," which was originally coedited by Cyril Harris and Charles Crede. That Handbook is now in its fifth edition with Harris' name in the title and Allan Piersol as the coeditor. The purpose of this presentation is to summarize the importance of that Handbook to shock and vibration engineering for the past 44 years, as well as the experience of working with Cyril Harris on the preparation of the fifth edition.

11:05

2aAA4. Working with Cyril M. Harris at Benaroya Hall. Robert M. Hoover (11900 Barryknoll, #2115, Houston, TX 77024) and Ashton Taylor (HFP Acoust. Consultants, Inc., Houston, TX 77036)

To contribute to a project such as Benaroya Hall, now the home of the Seattle Symphony Orchestra, as team members of acoustical giant Cyril M. Harris was a golden opportunity for Robert M. Hoover and Ashton Taylor. Their involvement in the project and association with Cyril began at the early stages of the project as consultants in noise control. During the schematic design phase, Cyril dictated to the LMN Architects design team that authority for all noise control decisions rested with Bob Hoover. This led to good working relationships, especially with the mechanical engineering firm and the air-conditioning contractor. Throughout the project, Cyril insisted upon the highest standards of materials and workmanship, and relied on the expertise of his acoustical design team, even if it went against his initial idea on some issue. At the beginning with the setting of noise criteria and in the finished building when measuring background sound, Cyril participated in the noise control process. Benaroya Hall has been critically acclaimed by the media and musicians, it was an honor for both Bob Hoover and Ashton Taylor to be part of that success.

2aAA5. Cyril M. Harris: Acoustician, educator, and mentor. Daniel R. Raichel (CUNY Grad. Ctr. and Eilar Assoc., 2727 Moore Ln., Fort Collins, CO 80526, draichel@comcast.net)

Cyril M. Harris belongs to that rare breed of technologists who excel in more than one aspect of acoustics. His achievements are legendary in the realms of architectural acoustics, physical acoustics, ultrasound, and speech processes. He coauthored with Vern O. Knudsen a classic text on architectural acoustics and served as editor of several major references. He also excelled as a teacher and mentor. When the author set up the Acoustics Research Center at the Cooper Union, Dr. Harris came over to see the new facility. At my invitation, he gave a lecture before an audience constituted of engineering and architectural students; one surely had to rate this event as one of the best seminars ever given at that institution. When I wrote *The Science and Applications of Acoustics*, Dr. Harris was generous in giving me permission to use a number of figures from his published texts. He also sent me students from Columbia to use the acoustic facilities at Cooper. One of his referees, a master's degree candidate, conducted a major study on Guastavino tiles for this thesis, which resulted in a prize-winning publication. The manifold accomplishments, influence, and legacy of Cyril Harris will surely resonate for generations to come.

11:45

2aAA6. On the diffusion of sound in an auditorium. Cyril M. Harris (Dept. of Elec. Eng. and Grad. School of Architecture, Columbia Univ., New York, NY 10027)

A condition of perfect diffusion of sound is said to exist in an auditorium if, at any point within it, the reverberant sound travels in all directions with equal probability, and if the level of the reflected sound is everywhere equal. In deriving the reverberation time formula, which predicts how long sound will bounce around an enclosed space after the source has stopped, W.C. Sabine assumed perfect diffusion within it. When this is not the case, his formula may predict inaccurate results. For example, the Sabine equation will not give correct results in an auditorium with poor diffusion, as when there is a large overhanging balcony, or if one of the dimensions of the enclosed space is very much greater than the other dimensions, or if the auditorium is divided into spaces having different acoustical properties. An auditorium with excellent diffusion beneficially affects the uniformity of decay of sound within the space and pleases the listener's ear. Among techniques that contribute to good diffusion are the surface irregularities found in the elaborate styles of architecture of the past. Illustrations will be presented showing some approaches within the modern architectural idiom that have yielded successful results.

TUESDAY MORNING, 18 OCTOBER 2005   HENNEPIN ROOM, 8:30 TO 10:55 A.M.

Session 2aAB

Animal Bioacoustics and Psychological and Physiological Acoustics: Cognition in the Acoustic Behavior of Animals I

Caroline M. DeLong, Chair
Brown Univ., Neuroscience, Box 1953, Providence, RI 02912

Chair's Introduction—8:30

Invited Papers

8:35

2aAB1. Vocally mediated social recognition in anurans. Mark A. Bee (Institut für Biologie und Umweltwissenschaften, Postfach 2503, Carl von Ossietzky Univ., D-26111 Oldenburg, Germany, mark.bee@uni-oldenburg.de)

Anuran amphibians (frogs and toads) are among the most vocal of vertebrates and have long served as model systems for investigating the mechanisms and evolution of acoustic communication. Compared to higher vertebrates, however, the role of cognition in anuran communication has received less attention, at least in part due to the lack of evidence that juvenile anurans learn to produce signals or associate them with particular social contexts. Recent studies of social recognition in two anuran families indicate that territorial male frogs in some species are able to learn about and recognize the individually distinctive properties of the calls of nearby neighbors. For example, male bullfrogs (*ranidae*) learn about the pitch of a neighbor's vocalizations (an individually distinct voice property) and associate a familiar pitch with the location of the neighbor's territory. As in songbirds, this form of vocally mediated social recognition allows territory holders to direct low levels of aggression toward well-established neighbors, while maintaining a readiness to respond aggressively to more threatening strangers that may attempt a territory takeover. A brief review of currently available data will be used to illustrate how anurans can serve as model systems for investigating the role of cognition in acoustic communication.

2aAB2. The psychology of songbird acoustic communication: Open-ended categorization and its role in songbird perception.
Christopher Sturdy (Dept. of Psych., Ctr. for Neurosci., P-217 Biological Sci. Bldg., Univ. of Alberta, Edmonton, AB T6G 2E9, Canada)

Songbirds are used as a model system for understanding vocal learning, production, and perception. Over the past several years one particular aspect of songbird communication has been examined, vocal perception that has parallels with human speech perception, namely, the perception of natural acoustic categories in songbird vocalizations. A series of studies will be presented highlighting the role of open-ended categorization in the perception of vocalizations in songbirds. Attempts will be made to make the case that open-ended categorization is a fundamental proximate mechanism for acoustic communication in both humans and songbirds, that it is used in a highly analogous manner in each, and future directions will attempt to integrate these two areas of research and, ultimately, lead to a more complete understanding of acoustic communication in animals.

2aAB3. Long-term memory of heterospecific vocalizations by African lions. Jon Grinnell (Dept. of Biol., Gustavus Adolphus College, 800 W. College Ave., St. Peter, MN 56082), Gus van Dyk (Tswalu Game Reserve, Kuruman, 8460, South Africa), and Rob Slotow (Univ. of Natal, Durban, 4041, South Africa)

Animals that use and evaluate long-distance signals have the potential to glean valuable information about others in their environment via eavesdropping. In those areas where they coexist, African lions (*Panthera leo*) are a significant eavesdropper on spotted hyenas (*Crocuta crocuta*), often using hyena vocalizations to locate and scavenge from hyena kills. This relationship was used to test African lions' long-term memory of the vocalizations of spotted hyenas via playback experiments. Hyena whoops and a control sound (*Canis lupus* howls) were played to three populations of lions in South Africa: (1) lions with past experience of spotted hyenas; (2) lions with current experience; and (3) lions with no experience. The results strongly suggest that lions have the cognitive ability to remember the vocalizations of spotted hyenas even after 10 years with no contact of any kind with them. Such long-term memory of heterospecific vocalizations may be widespread in species that gain fitness benefits from eavesdropping on others, but where such species are sympatric and often interact it may pass unrecognized as short-term memory instead.

2aAB4. Vocal behavior and risk assessment in wild chimpanzees. Michael L. Wilson (Gombe Stream Res. Ctr., P.O. Box 1182, Kigoma, Tanzania, and Dept. of Ecol., Evol. & Behav., Univ. of Minnesota, St. Paul, MN 55108 wilso198@tc.umn.edu), Marc D. Hauser, and Richard W. Wrangham (Harvard Univ., Cambridge, MA 02138)

If, as theory predicts, animal communication is designed to manipulate the behavior of others to personal advantage, then there will be certain contexts in which vocal behavior is profitable and other cases where silence is favored. Studies conducted in Kibale National Park, Uganda investigated whether chimpanzees modified their vocal behavior according to different levels of risk from intergroup aggression, including relative numerical strength and location in range. Playback experiments tested numerical assessment, and observations of chimpanzees throughout their range tested whether they called less frequently to avoid detection in border areas. Chimpanzees were more likely to call to playback of a stranger's call if they greatly outnumbered the stranger. Chimpanzees tended to reduce calling in border areas, but not in all locations. Chimpanzees most consistently remained silent when raiding crops: they almost never gave loud pant-hoot calls when raiding banana plantations outside the park, even though they normally give many pant-hoots on arrival at high-quality food resources. These findings indicate that chimpanzees have the capacity to reduce loud call production when appropriate, but that additional factors, such as advertising territory ownership, contribute to the costs and benefits of calling in border zones.

Contributed Papers

2aAB5. Classification of communication signals of the little brown bat. Karla V. Melendez, Douglas L. Jones, and Albert S. Feng (Univ. of Illinois Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801)

Little brown bats, *Myotis lucifugus*, are known for their ability to echolocate and utilize their echolocation system to navigate, locate, and identify prey. Their echolocation signals have been characterized in detail, but their communication signals are poorly understood despite their widespread use during the social interactions. The goal of this study was to characterize the communication signals of little brown bats. Sound recordings were made overnight on five individual bats (housed separately from a large group of captive bats) for 7 nights, using a Pettersson ultrasound detector D240x bat detector and Nagra ARES-BB digital recorder. The spectral and temporal characteristics of recorded sounds were first analyzed using BATSOUND software from Pettersson. Sounds were first classified by visual observation of calls' temporal pattern and spectral composition, and later using an automatic classification scheme based on multivariate statistical parameters in MATLAB. Human- and machine-based analysis revealed five discrete classes of bat's communication signals: downward frequency-modulated calls, constant frequency calls, broadband

noise bursts, broadband chirps, and broadband click trains. Future studies will focus on analysis of calls' spectrotemporal modulations to discriminate any subclasses that may exist. [Research supported by Grant R01-DC-04998 from the National Institute for Deafness and Communication Disorders.]

2aAB6. Communication calls of Japanese pipistrelle bats: Examination by acoustic characteristics. Yoshiki Osawa, Yoko Kondo, Tsuyoshi Nagato, Yoshiaki Watanabe, and Hiroshi Riquimaroux (Grad. School of Eng., Doshisha Univ., 1-3, Miyakodani, Tatara, Kyotanabe Kyoto 610-0321, Japan, dtf0745@mail4.doshisha.ac.jp)

We classified communication calls of Japanese pipistrelle bats (*Pipistrellus abramus*) by acoustic characteristics D. The Japanese pipistrelles emitted communication calls that were completely different from FM echolocation calls. Data showed that in general duration of communication calls was longer than that of echolocation calls (e.g., echolocation call, 1 ms; long CF communication call, 50 ms) and that frequency of communication calls were lower than that of echolocation calls (e.g., echolocation call, 80–40 kHz; long CF communication call, about 14 kHz). Typical

classified communication calls were as follows: slightly short CF call (20 ms, 14 kHz), short CF call (5 ms, 14 kHz), slightly long CF call (30 ms, 14 kHz), long CF call (50 ms, 14 kHz), slightly long FM call (15 ms, 30–15 kHz), long FM call (20 ms, 25–15 kHz), complicated FM call (10–50 ms, 25–15 kHz), short intermittent CF call (1 ms, 14 kHz) and noise call (20–100 ms, below 60 kHz). Details will be discussed more specifically. [Research supported by a grant to RCAST at Doshisha Univ. from MEXT and by the Innovative Cluster Creation Project promoted by MEXT.]

10:40

2aAB7. Modeling neural adaptation in the frog auditory system. Janine Wotton, Kimberly McArthur, Amit Bohara, Michael Ferragamo (Gustavus Adolphus College, St. Peter, MN 56802), and Andrea Megela Simmons (Brown Univ., Providence, RI 02912)

Extracellular recordings from the auditory midbrain, *Torus semicircularis*, of the leopard frog reveal a wide diversity of tuning patterns. Some cells seem to be well suited for time-based coding of signal envelope, and others for rate-based coding of signal frequency. Adaptation for ongoing stimuli plays a significant role in shaping the frequency-dependent response rate at different levels of the frog auditory system. Anuran auditory-nerve fibers are unusual in that they reveal frequency-dependent adaptation [A. L. Megela, J. Acoust. Soc. Am. **75**, 1155–1162 (1984)], and therefore provide rate-based input. In order to examine the influence of these peripheral inputs on central responses, three layers of auditory neurons were modeled to examine short-term neural adaptation to pure tones and complex signals. The response of each neuron was simulated with a leaky integrate and fire model, and adaptation was implemented by means of an increasing threshold. Auditory-nerve fibers, dorsal medullary nucleus neurons, and toral cells were simulated and connected in three ascending layers. Modifying the adaptation properties of the peripheral fibers dramatically alters the response at the midbrain. [Work supported by NOHR to M.J.F.; Gustavus Presidential Scholarship to K.McA.; NIH DC05257 to A.M.S.]

TUESDAY MORNING, 18 OCTOBER 2005
CARVER ROOM, 8:15 A.M. TO 12:00 NOON

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Topical Meeting on Imaging and Control of HIFU-Induced Lesions

Emad S. Ebbini, Chair
Univ. of Minnesota, Dept. of Electrical and Computer Engineering, 200 Union St., SE, Minneapolis, MN 55455

Chair's Introduction—8:15

Invited Papers

8:20

2aBB1. Real-time monitoring of ultrasound imaging of clinical high intensity focused ultrasound (HIFU) exposures. Gail ter Haar (Phys. Dept., Inst. of Cancer Res., Royal Marsden Hospital, Sutton, Surrey, UK), James Kennedy, Tom Leslie (Churchill Hospital, Oxford, UK), and Feng Wu (Chongqing Univ. of Medical Sci., Chongqing, China)

Currently, many clinical devices use the change in gray scale seen on a real-time ultrasound image for the assessment of the success of HIFU treatment. It has been shown previously that, for a single HIFU lesion, the presence of gray-scale change was indicative of successful ablation in 100% of cases for 1.6-MHz beams, and in 90% of cases for 0.8-MHz exposures. The absence of gray-scale change was a reliable indicator of lack of ablative damage only for 0.8-MHz exposures (80%); in 80% of exposures using 1.6-MHz beams there was a lesion even in the absence of gray-scale change. This study has been extended to more realistic clinical treatment protocols. The image appearance has been studied for the different volume ablation techniques that are used in the treatment of liver and kidney cancer. The results will be presented.

8:45

2aBB2. Hyperechogenicity during high intensity focused ultrasound (HIFU). Lawrence Crum, Michael Bailey, Brian Rabkin (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), Vera Khokhlova (Moscow State Univ., Moscow Russia), and Shahram Vaezy (Univ. of Washington, Seattle, WA 98105)

Ultrasound guidance of HIFU therapy is attractive because of its portability, low cost, real-time image processing, simple integration with HIFU instruments, and the extensive availability of diagnostic ultrasound; however, the use of ultrasound visualization for the guidance and monitoring of HIFU therapy often relies on the appearance of a hyperechoic region in the ultrasound image. It is often assumed that the formation of a hyperechoic region at the HIFU treatment site results from bubble activity generated during HIFU exposure. However, it has been determined that this region can be generated with relatively short bursts of HIFU (on the order of 30 ms), bursts so short that negligible temperature elevations are expected to occur. In examining the histology associated with these hyperechoes, there is little evidence of traditional cavitation damage; rather, it appears as if there are many bubbles generated within the individuals cells, suggesting a thermal mechanism. Thermocouple measurements of the temperature elevation were inaccurate due to the short insonation period, but showed only a few-degree temperature rise. These anomalous results will be presented, along with additional data on HIFU hyperechogenicity, and a hypothesis given for the phenomenological origins of this effect. [Work supported in part by the NSBRI, U.S. Army, and the NIH.]

2aBB3. Reproducibility of hemorrhage near fetal rat bone: Preliminary results. Timothy A. Bigelow[a)], Rita J. Miller, James P. Blue, Jr., and William D. O'Brien, Jr.   (Bioacoustics Res. Lab., Dept. of Elec. and Computer Eng., Univ. of Illinois, Urbana, IL 61801)

The reproducibility of ultrasound-induced hemorrhage near the fetal rat skull is important in therapy and related bioeffect studies. A 0.92-MHz $f/1$ spherically focused transducer (5.1-cm focal length) was used to expose the skull of 18- to 19-day-gestation exteriorized rat fetuses ($9.6 - \mu$s pulse duration). The fetuses were exposed to one of four exposure conditions (denoted A, B, C, and D) in addition to a sham exposure. Three of the exposures consisted of a peak compressional pressure of 10 MPa, a peak rarefactional pressure of 6.7 MPa, and PRFs of 100 Hz (A), 250 Hz (B), and 500 Hz (C), corresponding to time-average intensities of 1.9, 4.7, and 9.4 W/cm^2, respectively. Exposure D consisted of a peak compressional pressure of 6.7 MPa, a peak rarefactional pressure of 5.0 MPa, and a PRF of 500 Hz corresponding to a time-average intensity of 4.6 W/cm^2. The hemorrhage sizes were 1.4 ± 0.7 mm^2 for A (occurrence of 15). [a)] Currently at Electrical Engineering Department, University of North Dakota, Grand Forks, ND 58202.

2aBB4. Monitoring high-intensity focused ultrasound (HIFU) therapy using radio frequency ultrasound backscatter to quantify heating. Peter J. Kaczkowski and Ajay Anand   (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

The spatial distribution and temporal history of tissue temperature is an essential indicator of thermal therapy progress, and treatment safety and efficacy. Magnetic resonance methods provide the gold standard noninvasive measurement of temperature but are costly and cumbersome compared to the therapy itself. We have been developing the use of ultrasound backscattering for real-time temperature estimation; ultrasonic methods have been limited to relatively low temperature rise, primarily due to lack of sensitivity at protein denaturation temperatures ($50 - 70$ °C). Through validation experiments on gel phantoms and *ex vivo* tissue we show that temperature rise can be accurately mapped throughout the therapeutic temperature range using a new BioHeat Transfer Equation (BHTE) model-constrained inverse approach. Speckle-free temperature and thermal dose maps are generated using the ultrasound calibrated model over the imaged region throughout therapy delivery and post-treatment cooling periods. Results of turkey breast tissue experiments are presented for static HIFU exposures, in which the ultrasound calibrated BHTE temperature maps are shown to be very accurate (within a degree) using independent thermocouple measurements. This new temperature monitoring method may speed clinical adoption of ultrasound-guided HIFU therapy. [Work supported by Army MRMC.]

10:00−10:20   Break

2aBB5. Detection of tissue coagulation by ultrasonic displacement map. Shin-ichiro Umemura, Takashi Azuma, Kazuaki Sasaki, and Ken-ichi Kawabata   (Hitachi Central Res. Lab., Kokubunji, Tokyo 185-8601, Japan)

Real-time detection of tissue coagulation before boiling will significantly reduce the current limitation and the potential problems of non-invasive treatment with high-intensity focused ultrasound (HIFU). First, the throughput of the treatment will be doubled because the ultrasonic energy required to heat tissue to its coagulation temperature is about half of that to its boiling temperature. Second, the safety of the treatment will be enhanced because the explosive boiling of tissue can be avoided thereby. Coagulation of tissue can be detected through ultrasonically mapping its displacement. Thermal denaturation of protein is normally accompanied by a slight increase in its volume caused by breakage of hydrogen bonds. This volume change induces small displacement, which can be detected with ultrasound. This volume change as well as change in speed of sound may produce change in acoustic impedance of tissue, which can increase the reflectivity of the coagulated tissue and the acoustic radiation force on it. *In vivo* as well as *ex vivo* experiments were performed in combination with HIFU exposures with normal single spot focusing and split focusing. The former effect was dominant with split focusing while the latter was with single spot focusing. These will be discussed in the presentation.

2aBB6. MRI guidance for focused ultrasound surgery. Nathan McDannold and Kullervo Hynynen   (Brigham and Women's Hospital, 221 Longwood Ave., Rm. 521, Boston, MA 02115)

Magnetic resonance imaging (MRI) based monitoring has been shown in recent years to enhance the effectiveness of minimally or noninvasive thermal therapy techniques, such as focused ultrasound surgery. MR imaging's unique soft tissue contrast and ability to image in three dimensions and in any orientation make it extremely useful for treatment planning and for imaging the tissue response to the therapy. The temperature sensitivity of several intrinsic parameters enables MRI to visualize and quantify the progress an ongoing thermal treatment. The most useful temperature-sensitive parameter appears to be the proton resonant frequency, which allows for precise and accurate temperature measurements in water-based tissues. By acquiring a time series of quantitative temperature images, it is possible to monitor the accumulated thermal dose delivered to the target tissue and accurately predict the areas that are thermally ablated, while at the same time ensuring nearby critical structures are not heated. The method is currently used in an FDA approved focused ultrasound device for the treatment of uterine fibroids. Our research and clinical experience with these techniques will be reviewed.

2aBB7. MRI-guided therapeutic ultrasound: Temperature feedback control for extracorporeal and endoluminal applicators.
Rares Salomir (U556 Inserm, 151 Cours Albert Thomas, 69424 Lyon Cedex 03, France)

Therapeutic ultrasound is a mini-invasive and promising tool for *in situ* ablation of non-resectable tumors in uterus, breast, esophagus, kidney, liver, etc. Extracorporeal, endoluminal, and interstitial applicators have been successfully tested to date. Magnetic resonance imaging (MRI) is the only available technique providing non-invasive temperature mapping, together with excellent contrast of soft tissue. Coupling of these two technologies offers the advantage of both: (1) on line spatial guidance to the target region, and (2) thermal dose control during the treatment. This talk will provide an overview of the author's experience with automatic, active feedback control of the temperature evolution in tissues, which has been demonstrated with MRI compatible extracorporeal transducers (focused beam) or endoluminal applicators (plane waves). The feedback loop is based on fast switching capabilities of the driving electronics and real time data transfer out of the MR scanner. Precision of temperature control was typically better than 1°C. This approach is expected to improve the efficacy of the treatment (complete tumor ablation) and the thermal security of the critical regions crossed by the acoustic beam. It also permits one to reach an under-lethal heating regime for local drug delivery using thermosensitive liposomes or gene expression control based on hsp promoters.

2aBB8. Active ultrasonic imaging methods for lesion visualization. Ajay Shrestha, Brett Otteson, Hui Yao, and Emad Ebbini (Univ. of Minnesota Twin Cities, Minneapolis, MN 55455)

Several transient and irreversible changes in thermal and mechanical tissue properties are known to occur during lesion formation using HIFU. Increase in tissue absorption, decrease in tissue perfusion, and increase in tissue hardness are all associated with tissue necrosis associated with the application of HIFU at sufficiently high dosage. We have apparatus and algorithms for estimation of these tissue properties before, during, and after lesion formation using HIFU. Imaging data are obtained synchronously with the application of HIFU and processed for tissue property measurement. Absorption rate, shear elastic modulus, and perfusion are measured by application of 1 HIFU pulse of duration 0.1 to 0.2 s and sub-therapeutic intensity levels. The HIFU beam is interrupted for collection of imaging lines to obtain a spatio-temporal displacement map with sufficient temporal resolution for viscoelastic tissue property measurements. A description of the imaging system and synchronization between the HIFU and imaging transducers is given. Data from lesion formation experiments *in vitro* and *in vivo* are given and discussed.

TUESDAY MORNING, 18 OCTOBER 2005　　　　　　　　　CONRAD B, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: "Acts of Sound"—Hands-on Workshop for High School Students

Murray S. Korman, Cochair
U.S. Naval Academy, Physics Dept., Annapolis, MD 21402

Robert M. Keolian, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Chair's Introduction—10:00

2aED1. Acts of Sound: Hands-On Workshop for High School Students. Robert Keolian, Alexandra Loubeau, Matthew E. Poese, Linde J. Clark, (Applied Research Lab., The Pennsylvania State Univ., State College, PA 16804), Uwe J. Hansen, (Indiana Univ., Terre Haute, IN 47809), Andrew Morrison, (Northern Illinois Univ., DeKalb, IL 60115), Murray S. Korman, and Edward Tucholski (U.S.Naval Academy, Annapolis, MD 21402)

The Committee on Education in Acoustics is proud to present a workshop for high school students from the Minneapolis public schools. The theme of mentorship will be developed throughout the program. The eight mentors have individually developed their own workstations to inspire, challenge and delight a student's imagination about sound. The students will split into eight groups of three and work for 30 minutes at their own stations. After initial guidance, the students will do hands-on experiments, take data and try to come up with their own discoveries. Next, students have 15 minutes to prepare impromptu drawings and transparencies for a "team" presentation. Each team will be given 5 minutes to present at their workstation. A mini-cam and projector will be used to zoom in. During refreshments, there will be time for feedback and a musical skit. Then students may walk around to all the other stations to meet the other mentors and do more hands-on experiments at their own pace, as time permits. [Special thanks go to James Bickel, the science curriculum coordinator for the Minneapolis public schools, all the students and their teachers.]

PVC Filters Demo: **The spectrum of sound generated at one end of a PVC pipe can be made to change dramatically when simple PVC "filter sections" are placed on the other end. A low-pass filter (bass) is achieved with an expansion chamber, a high-pass filter (treble) requires a T branch, and a band-stop filter uses a Helmholtz resonator (jug with a thin neck). Measurements are made using a microphone and an oscilloscope. - Alexandra Loubeau**

Thermoacoustic Laser: **An ordinary test tube houses a small honeycombed ceramic "stack" material. Now, using a 6 volt battery and some fine wire wrapped around six grooves cut on one side of the cylindrical stack material, the gas on that side of the material heats up because you made a small toaster. But wait. Why does the tube then emit sound? Experiments are performed to show how the frequency changes with tube length, stack position, and sound level changes with electrical power. - Linde J. Clark**

Standing Wave Patterns on a Drum Head: **A thin elastic sheet (latex) is stretched on a circular frame and excited into motion by a loudspeaker. The sheet is painted with a grid of concentric circles and radial lines. Beautiful standing wave shapes are stroboscopically illuminated when the frequency of the sound coincides with a certain resonance wave pattern (mode). Modal patterns are measured using a laser Doppler vibrometer. - Edward Tucholski**

Solitons and Faraday Waves in a Water Trough: **Some unusual waves can be seen on the surface of water contained in a vertically oscillating trough, including a localized non-propagating soliton. - Robert Keolian**

Acoustic Levitation and Standing Waves in Tubes: **A loudspeaker placed at the end of a transparent tube is used to levitate packing peanuts. In a second experiment, an ultrasonic horn is used to levitate small styrofoam balls. The "standing wave shape pattern" can be measured by moving a small microphone along the axis of a tube and displaying the response on an oscilloscope. - Uwe Hansen**

Turn it Up! Loudspeakers in Boxes: **When sound is generated from a "naked" loudspeaker that is not placed in a box or enclosure, its sound intensity radiation pattern is remarkably different than when the identical speaker is snuggly mounted in a hole in the enclosure. Measurements using a microphone and an oscilloscope for both cases reveal that the latter case (termed monopole radiation) is drastically more desirable than in the former case (termed dipole radiation). - Matthew Poese**

Musical Acoustics - A Study of Woodwind Instruments: **The tube length and hole placement are important parameters in determining the allowed wavelengths (in this resonant system) and therefore the natural frequencies of each "standing wave pattern." Using instrumentation to measuring the ratio of acoustic pressure to the air's vibration velocity in the tube (called impedance) one can determine all the resonant frequencies of the system. - Andrew Morrison**

Acoustic Landmine Detection: **Sound waves generated in the air are used to detect vibrating objects buried in sand by comparing the vibration patterns of the sand's surface (1) over a mine simulant or (2) when the object is absent. Our simulant is a miniature acrylic drum. - Murray S. Korman**

Special Notice: Regular ASA meeting participants are welcome to the session, as long as their participation does not interfere with student hands-on activities.

Session 2aNCa

NOISE-CON: Plenary: Reducing Traffic Noise with Quieter Pavements

J. Stuart Bolton, Chair

Purdue Univ., Ray W. Herrick Labs., 140 S. Intramural Dr., West Lafayette, IN 47907-2031

Chair's Introduction—8:15

8:20

2aNCa1. Reducing traffic noise with quieter pavements. Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952)

In recent years, interest has increased in the use of pavement type to reduce traffic noise. This has been driven by public concern over noise from freeways and state transportation agencies' interest in using pavement instead of sound walls to mitigate traffic noise. Beginnings of the recent interest go back to 1998 with the formation of the Institute for Safe, Quiet & Durable Highways at Purdue University and the initiation long-term research by the California Department of Transportation (Caltrans) on the effectiveness of quieter pavements. In 2002, the State of Arizona announced plans to overlay 115 miles of concrete freeway in the greater Phoenix area with a quieter asphalt rubber surface. This turned into the first Quiet Pavement Pilot Program in partnership between Federal Highway Administration (FHWA) and the Arizona Department of Transportation. Since that time, the FHWA in cooperation with the American Association of State Highway Transportation Agencies conducted a fact finding ''Scan'' tour in Europe to evaluate their quiet pavement technology and policy. This was followed by the first comparative tire/pavement noise testing in the US and Europe using the same procedures and test tires. The results, issues, and future directions surrounding these activities will be discussed.

TUESDAY MORNING, 18 OCTOBER 2005

SALON B, 9:45 A.M. TO 12:05 P.M.

Session 2aNCb

NOISE-CON, Noise and Signal Processing in Acoustics: Array Methods for Noise Source Visualization I

Courtney B. Burroughs, Cochair

Noise Control Engineering, 1241 Smithfield St., State College, PA 16801

J. Stuart Bolton, Cochair

Purdue Univ., Ray W. Herrick Labs., 140 South Intramural Dr., West Lafayette, IN 47907-2031

9:45

2aNCb1. Determination of transmission loss of materials from array measurements using nearfield acoustical holography/inverse boundary element methods in a transmission loss facility. Earl G. Williams, Nicolas Valdivia, Peter Herdic (Naval Res. Lab., Washington, DC 20375), Jacob Klos, Daniel Palumbo, Richard Silcox (NASA Langley Res. Ctr., Hampton, VA 23681), and Bernard Sklanka (The Boeing Co., Seattle, WA 98124)

A new approach for the measurement of transmission loss based on NAH is discussed in this paper. Two array configurations were tested to determine the transmission loss of sound isolation materials applied to an aircraft fuselage panel in Boeing's interior noise test facility (INTF). The first configuration consisted of a 120 element planar array positioned conformal to the fuselage panel surface in the anechoic side of the TL facility. A second configuration used 50 microphones in a spherical array of diameter 0.4 meters placed in front of the panel. The normal intensity on the surface of the fuselage panel isolation layer was reconstructed using NAH/IBEM (nearfield acoustical holography/inverse boundary element methods) using partial fields generated from a set of reference accelerometers. The panel excitation was random plane wave fields generated in the reverberation room of the TL facility. This new approach measures the TL in narrow bands or 1/3 octave bands from 50 to 2 kHz, and is shown to be very accurate when compared with conventional TL measurements. This approach also provides the normal panel vibration for each partial field so that panel modes can be uncovered and studied. Furthermore, the spherical array provides a reconstruction of the vector acoustic intensity in the volume in front of the panel so that the magnitude and direction of the transmitted energy fields can be mapped.

10:05

2aNCb2. Comparison of different measurement technologies for the in-flight assessment of radiated acoustic intensity. Jacob Klos, Daniel L. Palumbo, Ralph D. Buehrle (Structural Acoust. Branch, NASA Langley Res. Ctr., M.S. 463 Hampton, VA 23681), Earl G. Williams, Nicolas Valdivia, Peter C. Herdic (Naval Res. Lab., Washington DC 20375), and Bernard Sklanka (The Boeing Co., Seattle, WA 98124)

Near-field acoustical holography (NAH) microphone array measurement technologies are currently being investigated for use in aircraft for determining the sound intensity that is radiated through a trimmed aircraft fuselage. Two types of microphone array geometries are studied in this effort, spherical and conformal to the sidewall. Due to the complexity of the interior sound field and in-flight excitation, questions have arisen concerning the capability of array techniques to measure these desired acoustic quantities. The main questions relate to the agreement between array measurements and traditional two-microphone intensity probe measurements and the sensitivity of array measurements to backside sources present in the aircraft cabin. Thus, an investigation was performed in which the NAH array results were compared to intensity probe results measured in the NASA Langley Structural Acoustic Loads and Transmission facility, the Boeing Interior Noise Test Facility, and in flight on the NASA Aries 757 aircraft. An overview of this test sequence and a comparison of NAH, IBEM, and two-microphone intensity probe measurements are made. These comparisons show the relative behavior of the different measurement techniques in both idealized laboratory and realistic in-flight environments. Conclusions are drawn concerning the use of array measurements to assess the sound intensity transmitted through an aircraft sidewall in-flight.

10:25

2aNCb3. The application of nearfield acoustical holography in the visualization of sound fields in enclosed spaces associated with patches on the surface of the space. Courtney B. Burroughs (Noise Control Eng. Inc., 799 Middlesex Tpke., Billerica, MA 01821)

A description of the application of nearfield acoustical holography in the visualization of the acoustic fields produced inside an enclosure by sections or patches of the vibrating surface of the enclosure are described. An array conformal to the surface that covers the patch of interest is used to measure the nearfield unsteady complex pressures. These pressures are used with nearfield freefield Greens functions to reconstruct the velocities on the surface under the array. Farfield Greens functions for the enclosed space are the applied to the reconstructed surface velocities to visualize the acoustic field inside the enclosure produced by the surface under the array. A mathematical simulation of the backward propagation to the surface and forward propagation to points inside the enclosure is used to explore the sensitivity of the process to (1) simulated extension of the array to improve resolution, (2) noise, and (3) measurement errors.

10:45

2aNCb4. Acoustic sources' localization in presence of reverberation. E. Julliard, S. Pauzin, F. Simon, and D. Biron (ONERA, 2 Ave. Edouard Belin, BP 4025, Toulouse Cedex 4, 31 055, France)

For several years, aeronautical industries have wished to improve internal acoustical comfort. In order to make it, they need metrological tools which are able to help them to spot acoustical sources and the associated path in a specific frequency range (i.e., for helicopters' internal noise: 1000–5000 Hz). Two major source' localization' tools exist: holography and beamforming, but these two techniques are based on a free field's hypothesis. So, problems appear when these techniques are used in a reverberant medium. This paper deals with the study and the comparison of holography and beamforming results in an enclosed area. To complete the study, intensimetry is also implemented to have information on the energy propagation. In order to test the performances of each method, two reflecting panels are put at right angles to create a reverberant environment, in an anechoic chamber. We seek to locate loudspeakers clamped in one panel, in the presence of parasite loudspeakers located on the other one. Then, a parametrical study is led: localization and number of sources, coherent or noncoherent sources. Thus, using limitations, precautions to take, and a base of comparison three methods are put forward. Finally, some envisaged solutions to limit problems of reflections (signal processing, over-turning, etc.) are presented.

11:05

2aNCb5. Acoustical imaging by the use of local basis functions. Lei He and J. Stuart Bolton (Ray W. Herrick Labs., Purdue Univ., 140 Intramural Dr., West Lafayette, IN 47907-2031)

Experimental acoustical source imaging has been studied extensively, and many methods have been proposed to address this issue. Generally, the acoustical imaging process is modeled as an estimation problem: i.e., the intention is to estimate the boundary values of a field given measurements of the exterior field value. The imaging methods most commonly used at the moment are the beamforming and nearfield acoustical holography families. In both of these procedures, the basis functions used to represent the boundary value are fixed without regard to the source types to be recovered. It may easily be imagined that some source types can be represented very efficiently by some specific set of basis functions, while for other source types the approximation is poor even when a large number of terms are used. In this paper, the efficient representation of the boundary value is considered to be essential, as is the ability to take advantage of preknowledge: e.g., regions where the source velocity is known to be zero. A general procedure is introduced here which makes use of local basis functions of the type used in image processing applications. The effectiveness of the procedure is illustrated through a planar source imaging example.

11:25

2aNCb6. Comparison of Helmholtz equation least squares and other methods in diagnosing stationary and nonstationary vehicle noises. Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

This paper gives a brief overview of current near-field acoustic holography (NAH) techniques. In particular, it compares the analyses of vehicle transaxle noise under run-up and coast-down conditions using spatial transformation of sound field (STSF) and Helmholtz equation least-squares (HELS) based NAH. It also demonstrates an application of HELS to visualizing brake squeals and compares its advantages and disadvantages to those of a laser vibrometer.

11:45

2aNCb7. Theory of patch inverse boundary element method (IBEM). Nicholas Valdivia, Earl G. Williams, Peter Herdic (Naval Res. Lab., 4555 Overlook Ave., SW, Washington, DC 20375), Jacob Klos, Daniel Palumbo, Richard Silcox (NASA Langley Res. Ctr., Hampton, VA 23681), and Bernard Slanka (The Boeing Co., Noise Eng. Lab., Seattle, WA 98124)

Near-field acoustical holography (NAH) based on boundary element methods require the measurement of the near-field pressure field over a conformal and close surface in order to recover the acoustic field on a nearby surface. In this work a new technique called patch IBEM is introduced. This new technique eliminates the need to measure over an entire closed surface, but instead the measurements are taken over a small area to reconstruct the acoustic field. Patch IBEM is used for the reconstruction of the acoustic field over a fuselage panel in a Boeing 757 aircraft, using measurement taken at Boeing's interior noise test facility (INTF). The measurements taken include a 120 mikes planar array positioned conformal to the fuselage panel surface and a 50 mikes spherical array. Results will be showed when the fuselage panel is excited by a shaker. [Work supported by ONR and NASA.]

Session 2aNCc

NOISE-CON and Noise: From Noise Control to Product Design

Richard H. Lyon, Chair
RH Lyon Corp, 60 Prentiss Ln., Belmont, MA 02478-2021

Chair's Introduction—10:00

10:05

2aNCc1. Noise, vibration and harshness (NVH) criteria as functions of vehicle design and consumer expectations. Daniel R. Raichel (CUNY Grad. Ctr. and Eilar Assoc., 2727 Moore Ln., Fort Collins, CO 80525-2192, draichel@comcast.net)

The criteria for NVH design are to a large degree determined by the types of vehicles and the perceived desires of the purchasers of vehicles, as well as the cost of incorporating NVH measures. Vehicles may be classified into specific types, e.g., economy car, midsize passenger, near-luxury and luxury passenger cars, sports cars, vans, minivans, and sports utility vehicles of varying sizes. The owner of a luxury sedan would expect a quiet ride with minimal vibration and harshness—however, if that sedan is to display sporting characteristics, some aspects of NVH may actually have to be increased in order to enhance a feeling of driver exhilaration. A discussion of the requirements for specific types of vehicles is provided, with due regard for effects on the usability of installed sound/video systems, driver and passenger fatigue, feel of steering mechanisms and other mechanical components, consumer market research, etc. A number of examples of vehicles on the market are cited.

10:25

2aNCc2. The acoustical design of vehicles: A new tool for benchmarking and target sound. Klaus Genuit (HEAD Acoust. GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany), Brigitte Schulte-Fortkamp (Inst. of Tech. Acoust., Tech. Univ. Berlin, Einsteinufer 25, 10587 Berlin) and Andre Fiebig (HEAD Acoust., Hersogenrath, Germany)

Evaluating sound quality leads to difficulties. Conventional tests neglect the context-sensitivity of human perception evaluation; they only reflect artificial realities which do not correspond with perceptual authenticity. Since the subjective evaluation is influenced by different modifiers, it is necessary to extend methods in order to be able to survey and represent various objective variables as well as subjective variables sufficiently. It is known that experiments in laboratories only constitute artificial realities and disregard important parameters like context, relevance, or external validity. The test design lacks the consideration of the usual interaction between the self-stimulated sound and the driver as well as the context, in which the driver experiences and evaluates the perceived sound. A 3-D sound simulation system (H3S) has been developed for the realistic reproduction of airborne sound in driving simulators. Mobile H3S is capable of simulating a vehicle soundscape—consisting of engine sound, tire, and wind noise—according to a real driving situation. A vehicle equipped with this interactive technique can be driven regularly on the road while the acoustic impression for the driver originates from a different or modified vehicle. The soundscape adapts online to changes in speed, engine rpm, and load caused by the vehicle controls.

10:45

2aNCc3. Noise control, sound, and the vehicle design process. Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952)

For many products, noise and sound are viewed as necessary evils that need to be dealt with in order to bring the product successfully to market. They are generally not product "exciters" although some vehicle manufacturers do tune and advertise specific sounds to enhance the perception of their products. In this paper, influencing the design process for the "evils," such as wind noise and road noise, are considered in more detail. There are three ingredients to successfully dealing with the evils in the design process. The first of these is knowing how excesses in noise effects the end customer in a tangible manner and how that effects customer satisfaction and ultimately sells. The second is having and delivering the knowledge of what is required of the design to achieve a satisfactory or even better level of noise performance. The third ingredient is having the commitment of the designers to incorporate the knowledge into their part, subsystem or system. In this paper, the elements of each of these ingredients are discussed in some detail and the attributes of a successful design process are enumerated.

11:05

2aNCc4. NASA Glenn Research Center Acoustical Testing Laboratory: Five year retrospective. Beth A. Cooper (Acoust. Testing Lab., NASA John H. Glenn Res. Ctr. at Lewis Field, 2100 Brookpark Rd., M.S. 86-10, Cleveland, OH 44135), James C. Akers, and Paul J. Passe (Analex Corp., NASA John H. Glenn Res. Ctr. at Lewis Field, Cleveland, OH 44135)

In the five years since the NASA Glenn Research Center Acoustical Testing Laboratory (ATL) opened its doors in September, 2000, it has developed a comprehensive array of services and products that support hearing conservation goals within NASA and industry. The ATL provides acoustic emission testing and noise control engineering services for a variety of specialized customers, particularly developers of equipment and science experiments manifested for NASA's manned space missions. The ATL aggressively supports the vision of a low-noise on-orbit environment, which facilitates mission success as well as crew health, safety, and comfort. In concert with these goals, the ATL also produces and distributes free educational resources and low-noise advocacy tools for hearing conservation education and awareness. Among these are two compact discs of auditory demonstrations (of phenomena in acoustics, hearing conservation, and communication), and presentations, software packages, and other educational materials for use by engineers, audiologists, and other hearing conservation stakeholders. This presentation will highlight ATL's construction, history, technical capabilities, and current projects and will feature demonstrations of some of the unique educational resource materials that are distributed by the ATL.

2a TUE. AM

Session 2aNCd

NOISE-CON and Noise: Issues in Aircraft Noise Analysis

Andrew S. Harris, Chair
Andrew S. Harris Inc., 19 University Ln., Manchester, MA 01944

Chair's Introduction—10:00

10:05

2aNCd1. Federal actions on control noise from airport operation.
Andrew S. Harris　(Andrew S. Harris, Inc., 19 University Ln., Manchester, MA 01944)

The problem of aircraft noise impacts in communities from the operation of airports is addressed from two perspectives: federal actions to establish mechanisms to control noise and actions by airport proprietors to develop noise control programs. Aircraft noise from the operation of civil airports became a community problem with the introduction of first-generation passenger jets such as the Boeing 707 and the Douglas DC-8. Federal Aviation Regulation Part 36 set source limits for new jet aircraft. However, there was no broadly accepted information about the compatibility of environmental noise and community activities and no uniform method to reduce impacts of aircraft noise until further federal action. The U.S. EPA published compatibility information under mandates of the Noise Control Act of 1972 (NCA72). The Federal Aviation Administration promulgated procedures for development and implementation of Noise Compatibility Programs for Airports as Federal Aviation Regulation Part 150 (FAR Part 150) in 1981. The effectiveness of programs developed by airport proprietors under FAR Part 150 is addressed as are the reactions of aircraft operators to concerns that airport programs could adversely impact their ability to conduct business. Congressional actions to limit actions by airport proprietors are also considered as are current regulatory issues.

10:25

2aNCd2. Error of estimation of community reaction to aircraft noise.
Sanford Fidell　(Fidell Assoc., Inc., 23139 Erwin St., Woodland Hills, CA 91367)

Errors and uncertainties of measurement, estimation, and prediction of aircraft noise exposure and of community reaction to it can be so great as to render dosage-effect analyses of community response to aircraft noise unreliable. Biases and other errors may arise during routine monitoring of aircraft noise, via oversimplification and misrepresentation of exposure and sound propagation conditions in predictive noise modeling, through sampling and interviewing procedures, and from poor statistical association in functional relationships between community response and acoustic predictor variables. Some of the more notable uncertainties afflicting the prediction of aircraft noise and its effects are described in this presentation.

10:45

2aNCd3. Observed cause-effect relationships between measured aircraft sound exposure levels and various predictor variables.
Richard D. Horonjeff　(Consultant in Acoust. and Noise Control, 81 Liberty Square Rd. #20-B, Boxborough, MA 01719), David A. Senzig (Senzig Eng., Winchester, MA 01890), and Grant S. Anderson　(Harris Miller Miller & Hanson Inc., Burlington, MA 01803)

Sound exposure levels of in-service jet aircraft departures at a single airport were measured along with a number of potential explanatory variables to determine the extent of cause-effect relationships under typical uncontrolled operating conditions. Explanatory variables included aircraft

type, installed engine type, aircraft gate weight, aircraft operator, 10-meter wind velocity, temperature, relative humidity, and barometric pressure. The two factors having the greatest influence on measured SELs (within aircraft type) were (1) a linear association with gross weight and (2) a logistic association with air density (or density altitude). The logistic relationship with air density shows an asymptotic functional form at extreme density values suggestive of full takeoff power under low air density conditions, and lower power settings under high air densities. Together, gross weight and air density alone explained 80 percent of the measured SEL variance. The remaining variables appeared to show little affect for the aircraft types investigated, with the exception of headwind component at speeds in excess of 15 knots. These relationships, along with the integrated noise model's (INM) ability to predict them, are presented. Also presented are differences in computed annual average SELs using single, annual average predictor values of the variables versus separate, seasonally averaged ones.

11:05

2aNCd4. A pragmatic re-analysis of sleep disturbance data.　Grant S. Anderson and Nicholas P. Miller　(Harris Miller Miller & Hanson Inc., 15 New England Executive Park, Burlington, MA 01803)

The acoustics literature documents several field studies in which aircraft noise was measured in people's bedrooms while their awakening response was simultaneously monitored. Nearly all the studies produced a dose-response relationship between a single event noise metric and an awakening response. Virtually all of these dose-response relationships pertain to (1) the indoor noise dose produced by a single aircraft flyover and (2) the chances that the noise dose will awaken an average person. These dose-response relationships are too limited for application to a full night of operations and to a realistic population of varying individual sensitivities to awakening. Accounting for multiple aircraft during the night is obviously necessary. Less obvious, but important is the need to account for person-to-person variation in awakening sensitivity to noise. This paper presents a re-analysis of original data from one of the field studies. The re-analysis generalizes the dose-response relationships for behavioral awakening to account for multiple aircraft during the night. Then it further generalizes the results to account for person-to-person variations in awakening. The resulting algorithms are applied to the output of the FAA's Integrated Noise Model, and contours of percent of the population awakened are computed for a full night of aircraft operations.

11:25

2aNCd5. Passive interior noise reduction analysis of King Air 350 turboprop aircraft using boundary element method/finite element method (BEM/FEM).　Indranil Dandaroy, Joseph Vondracek, Ron Hund, and Dayton Hartley　(Raytheon Aircraft Co., 9709 East Central, M.S. B6, P.O. Box 85, Wichita, KS 67201)

The objective of this study was to develop a vibro-acoustic computational model of the Raytheon King Air 350 turboprop aircraft with an intent to reduce propfan noise in the cabin. To develop the baseline analysis, an acoustic cavity model of the aircraft interior and a structural dynamics model of the aircraft fuselage were created. The acoustic model

was an indirect boundary element method representation using SYS-NOISE, while the structural model was a finite-element method normal modes representation in NASTRAN and subsequently imported to SYS-NOISE. In the acoustic model, the fan excitation sources were represented employing the Ffowcs Williams-Hawkings equation. The acoustic and the structural models were fully coupled in SYSNOISE and solved to yield the baseline response of acoustic pressure in the aircraft interior and vi-bration on the aircraft structure due to fan noise. Various vibration absorb-ers, tuned to fundamental blade passage tone (100 Hz) and its first har-monic (200 Hz), were applied to the structural model to study their effect on cabin noise reduction. Parametric studies were performed to optimize the number and location of these passive devices. Effects of synchrophas-ing and absorptive noise treatments applied to the aircraft interior were also investigated for noise reduction.

Session 2aNCe
NOISE-CON and Noise: Workshop on Noise Policy Developments I

William W. Lang, Cochair
29 Hornbeck Ridge, Poughkeepsie, NY 12603

George C. Maling, Jr. Cochair
60 High Head Rd., Harpswell, ME 04079

The purpose of the Workshop is to discuss the latest developments on noise policy since the first Workshop was held in Baltimore, MD in 2004 July. The Workshop will focus on several areas: community (environmental) noise, manufacturing and product noise, and occupational noise. The discussions of community noise will include aviation/airport noise, surface transportation noise, industrial noise, and local citizens' concerns. Specialists in these areas will discuss recent actions related to noise policy. In those areas where no recent action has been taken in the USA, the speakers will be asked to summarize the current status of the implementation of key policies in their areas of expertise. Following the presentations, there will be a focused discussion among the presenters and members of the audience. The objective of the Workshop is to bring the attendees up to date on what is being done to stimulate the rejuvenation of noise policies in America.

I. Developments on Community Noise Policy

Chair's Opening Remarks − 10:00

10:10−11:10

Invited Panel Presentations
11:10−12:00
Discussion
Workshop Continues in Session 2pNCe

Session 2aNS
Noise, NOISE-CON and ASA Committee on Standards: Specifying Uncertainties in Acoustic Measurements I

William J. Murphy, Cochair
National Inst. for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH 45226-1998

Ralph T. Muehleisen, Cochair
Illinois Inst. of Technology, Civil and Architectural Engineering, 3201 S. Dearborn St., Chicago, IL 60616

Chair's Introduction—9:20

Invited Papers

9:25

2aNS1. Microphone calibration and measurement uncertainties. George S. K. Wong and Lixue Wu (Inst. for Natl. Measuring Standards, Natl. Res. Council, Ottawa, ON K1A 0R6, Canada)

In acoustical measurement, the uncertainty of the front end components such as the microphone has to be taken into account together with the instrument measuring chain that follows. In this presentation, examples of uncertainties of microphone calibration

1889   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1889**

are present with some discussions given on the tolerance limits provided by national and international standards for sound level meters. It is not surprising that one may find the uncertainty in a sound level measurement may be much higher than one expected.

9:50

2aNS2. Use of the Monte Carlo method for uncertainty analysis of acoustic models and measurements. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., 3201 S. Dearborn St., Chicago, IL 60616)

The Monte Carlo method is often used to estimate uncertainties in models or measurements when analytic formula are not easily determined and/or when the inputs do not have independent errors. In this work, the use of the Monte Carlo method to estimate uncertainties in several common acoustic measurements is presented. The determination of uncertainty distributions of the input parameters, the generation of correlated input errors, and the use of the Monte Carlo method for determination of output uncertainty is discussed.

10:15

2aNS3. Uncertainty in outdoor noise measurement and prediction. D. Keith Wilson (U.S. Army Eng. Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03755)

Standards for outdoor noise are intended to ensure that (1) measurements are representative of actual exposure and (2) noise prediction procedures are consistent and scientifically defensible. Attainment of these worthwhile goals is hindered by the many complexities of sound interaction with the local atmosphere and terrain. The paradigm predominant in current standards might be described as measuring/predicting ''somewhat worse than average'' conditions. Measurements/predictions are made for moderate downward refraction conditions, since that is when noise annoyance is most often expected to occur. This paradigm is reasonable and practical, although one might argue that current standards could implement it better. A different, potentially more rigorous, paradigm is to explicitly treat the statistical nature of noise imissions as produced by variability in the atmospheric environment and by uncertainties in its characterization. For example, measurements and prediction techniques could focus on exceedance levels. For this to take place, a better conceptual framework must be developed for predictions that are averaged over environmental states, frequency bands, and various time intervals. Another increasingly important issue is the role of computer models. As these models continue to grow in fidelity and capability, there will be increasing pressure to abandon standard calculations in many applications.

10:40

2aNS4. Psychophysical uncertainty estimates for real ear attenuation at threshold measurements in naive subjects. William J. Murphy (Natl. Inst. for Occupational Safety and Health, Hearing Loss Prevention Team, 4676 Columbia Pkwy., M.S. C-27, Cincinnati, OH 45226-1998.), David Byrne (Natl. Inst. for Occupational Safety and Health, Hearing Loss Prevention Branch, Bruceton, PA 15236), Brad Witt, and Jesse Duran (Howard Leight Industries, San Diego-Acoust. Lab., San Diego, CA 92154)

The International Standards Organization (ISO) TC43 SC1 has directed the working groups to use the ''Guide to the Expression of Uncertainty in Measurement'' (GUM) in the development of new acoustical standards. In light of this resolution, working group 17 on hearing protection has added an annex to a proposed standard ISO4869-7 describing a subject fit method for estimation of noise reduction. Data from two laboratories, NIOSH Pittsburgh Research Lab and Howard Leight Industries, were analyzed to examine the within-test variance and between-test variance when subjects were tested using a Bekesy tracking paradigm. The NIOSH data demonstrated that the within-test variance, calculated from the reversals of the Bekesy tracks did not agree with the between test variability of the subjects' repeated measures. This result prompted a revision of the testing software to determine the within-test variance from the midpoint of the Bekesy tracks. The revised software has been installed at both NIOSH PRL and Howard Leight facilities. This paper will compare the within-test and between-test variances determined with reversal levels and the midpoints from these two labs. The variability has been modeled with a simulated subject and Monte Carlo techniques.

11:05

2aNS5. The ultimate uncertainty—Subjective assessment? Brigitte Schulte-Fortkamp (Tech. Acoust., TA 7, Tech. Univ. Berlin, Einsteinufer 25 D-10587 Berlin, Germany)

The evaluation of acoustic ambiences is influenced by the context, the point in time, and the place in which sounds appear as well as by people's minds. It is necessary to take into account the nature of the social world and to realize that values and interests become part of the process of research. With their help people's behavior, perspectives, and experiences can be explored. Interactions and interrelations between acoustic parameters and phenomena in everyday life can be investigated with regard to people's expectations. The focus is on people's everyday life, these data have primacy; the theoretical framework is not predetermined but derives directly from the data and is context-bound. The context of the participants' lives and work affects their behavior so that it is important to keep in mind that the participants are grounded in their history and temporality. The context of people's lives, events, and actions is studied in everyday-''real life''-settings. Their culture does not only consist of the physical environment but also of particular ideologies, values, and ways of thinking. Human response in measurements depends on the way people interpret and make use of their experiences and on the world which they live in.

11:35–12:00
Panel Discussion

TUESDAY MORNING, 18 OCTOBER 2005   CONRAD A, 8:00 TO 11:25 A.M.

Session 2aPA

Physical Acoustics and Engineering Acoustics: Thermoacoustics: What our Customers Want and Why They Want It, Part 1

David Gardner, Chair
Los Alamos National Lab., Material Div., MS/K764, Los Alamos, NM 87545

Chair's Introduction—8:00

Invited Papers

8:05

2aPA1. Technology push, market pull, and the Valley of Death. Gregory W. Swift   (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., Los Alamos, NM 87545, swift@lanl.gov)

The Valley of Death is the gap between fundamental research and product development, where apparently promising technologies can stall or disappear. Fundamental researchers may hope for potential applications of their work, and they try to push technology based on their research. Businesses may hope that new technology might serve their market needs, and they try to find promising new technologies that can be pulled toward practical use. The valley between the researchers and the businesses can be surprisingly twisted and thorny, despite government attempts to build roads across it. The histories of cryogenic engineering in the late 20th century and of thermoacoustics work at Los Alamos offer examples of both useful and misguided strategies in this valley. Although global thermoacoustics R&D has not (yet?) been as successful as cryogenic engineering, thermoacoustics has thus far avoided some of the worst pitfalls in the valley.

8:30

2aPA2. The role of thermoacoustics in the world of commercial cooling. John A. Corey   (CFIC / Qdrive, 302 10th St., Troy, NY 12180)

The science of thermoacoustics has been with us for nearly 30 years, but as yet few applications have made their way to the marketplace. Acoustic Stirling cryocoolers (also called pulse-tube Stirling or high-frequency pulse-tube coolers) have been the most successful commercial thermoacoustic devices, because they address a region of the cooling market in terms of temperature and cooling power that is not well served by existing technology. This talk will explore how thermoacoustics might fare in attempting to compete with existing technologies in refrigeration and air conditioning, what niche markets make the most sense as entry points, and how thermoacoustics compares to conventional (kinematic or free-piston) Stirling machines. In particular, why there are relatively few commercial Stirling devices in the marketplace (although Stirling cycle machines have been around for over 150 years) will be discussed, and what lessons learned with Stirlings are applicable to thermoacoustics.

8:55

2aPA3. Quality at a price you can afford. Steven Garrett, Robert Smith, and Matthew Poese   (ThermoAcoustics Corp., 546 Aguajito Rd., Carmel, CA 93923)

The transition of thermoacoustics from the laboratory to the commercial sector involves a process that is very different from science. In the laboratory, when an experiment fails to produce the expected result, it is possible to make modifications, followed by further measurements, to isolate the discrepancy. After discussions with nearly 40 different companies over the past 5 years, there is rarely any feedback about why a given company decides not to proceed. In such an open-loop environment, we are left with only speculation regarding this obscure process. This talk will postulate some reasons that have led to our failure, thus far, to team with an early adaptor of the technology. These include their general lack of familiarity with acoustics, difficulties surrounding licensing of existing and future intellectual property controlled by a university, and the drift away from the maintenance of an active in-house research capability in industry [see H. W. Chesbrough, Open Innovation (Harvard, 2003)]. The sequence of events that lead to formation of the ThermoAcoustics Corporation, an investor-funded start-up, will be related, along with the changes that licensure of intellectual property have made to our ability to deal with industry on a business-to-business basis without intervention of university tech-transfer professionals.

2aPA4. Practical issues in adopting a traveling wave thermoacoustic cooler for use in a food storage refrigerator. Philip S. Spoor (CFIC / Qdrive, 302 10th St., Troy, NY 12180)

CFIC/QDrive has developed a food storage refrigerator for the Army based on thermoacoustic technology. This "Phase II" SBIR project is a continuation of a "Phase I" effort that explored using a standing-wave thermoacoustic cooler for the refrigerator. The standing-wave cooler was found to be too inefficient with too low a power density to be practical, so it was switched to an acoustic Stirling, or traveling-wave thermoacoustic (regenerator based) cooler for Phase II. The major challenges of this project were adapting the Stirling-style cooler to a food storage application, and not the fundamentals of the cooler itself (the one exception being the issue of acoustic streaming). The challenges include: Running at 60 Hz (without frequency-shifting electronics), heat exchange without circulating fluids, dynamic balance, guarantee of long life, efficiency, and compactness (power density). How these challenges were met and how they drove the design, in most cases away from what would be ideal for the cycle itself, will be discussed. Time permitting, how the additional pressure of low unit cost would affect this type of product development will also be discussed. [Research supported by the U. S. Army through a Small Business Innovation Research (SBIR) grant.]

2aPA5. Being everything to anyone: Applicability of thermoacoustic technology in the commercial refrigeration market. Matthew E. Poese, Robert W. M. Smith, and Steven L. Garrett (Appl. Res. Lab., Penn State Univ. and ThermoAcoustics Corp., P.O. Box 30, State College, PA 16804, poese@psu.edu)

This talk will compare electrodynamically driven thermoacoustic refrigeration technology to some common implementations of low-lift vapor-compression technology. A rudimentary explanation of vapor-compression refrigeration will be presented along with some of the implementation problems faced by refrigeration engineers using compressor-based systems. These problems include oil management, compressor slugging, refrigerant leaks and the environmental impact of refrigerants. Recently, the method of evaluating this environmental impact has been codified to include the direct effects of the refrigerants on global warming as well as the so-called "indirect" warming impact of the carbon dioxide released during the generation (at the power plant) of the electrical power consumed by the refrigeration equipment. It is issues like these that generate commercial interest in an alternative refrigeration technology. However, the requirements of a candidate technology for adoption in a mature and risk-averse commercial refrigeration industry are as hard to divine as they are to meet. Also mentioned will be the state of other alternative refrigeration technologies like free-piston Stirling, thermoelectric and magnetocaloric as well as progress using vapor compression technology with alternative refrigerants like hydrocarbons and carbon dioxide.

10:10–10:25 Break

Contributed Papers

2aPA6. Second law performance of thermoacoustic systems. Insu Paek, James E. Braun, and Luc Mongeau (Ray W. Herrick Labs., School of Mech. Eng., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031, paek@purdue.edu)

Thermoacoustic cooler prototypes developed so far can be divided into three categories based on their cold-side operating temperatures: cryogenic, refrigeration, and the third is the air-conditioning. For most prototypes developed so far, the coefficient of performance (COP), defined as the cooling power divided by the acoustic power, was well above unity. The thermodynamic second law efficiency, defined as the COP to the reversible COP, however, varied between 0.03 and 0.22 for different prototypes. The objective of this work was to evaluate the best application area for thermoacoustic cooling from the point of view of the thermodynamic second law efficiency. Model predictions in conjunction with empirical heat transfer correlation models for thermoacoustic environments were used. The results suggest that refrigeration applications are better suited for thermoacoustic cooling than cryogenics or air-conditioning.

2aPA7. Determination of an equivalent pore size from acoustic flow measurements. Linde Clark, Jin Liu, and Steven Garrett (Grad. Prog. in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804-0000, sxg185@psu.edu)

The hydraulic radius, r_h, is defined as the ratio of a channel's cross-sectional area to its perimeter. This parameter is important for specification of the performance of a porous medium that can be used as a regenerator in a Stirling engine or refrigerator. It is easy to calculate r_h for pores of regular geometry, but difficult in more complex media. Two techniques

which use oscillating flow to determine this parameter will be presented and compared. One technique extracts r_h by finding the low velocity limit of the standard expression for viscous pressure drop in the Poiseuille flow regime. The other involves a plot of the nondimensional viscous flow resistance, $\Delta p_{\mathrm{vis}}/\Delta x \omega \rho u$, versus the reciprocal of the viscous penetration depth, $1/\delta_\nu$, in the laminar flow regime. When $r_h < \delta_\nu$, the flow behavior is frequency independent and the dynamics is characterized by r_h only. When $r_h > \delta_\nu$, the flow resistance is frequency dependent and the dynamics is characterized by both r_h and δ_ν. It is possible to identify an effective hydraulic radius by equating it to the value of δ_ν where that transition occurs. [Work supported by ONR.]

2aPA8. Capillary-tube-based extension of thermoacoustic theory for a random medium. Heui-Seol Roh, Richard Raspet, and Henry E. Bass (NCPA, Univ. of Mississippi, Coliseum Dr., University, MS 38677)

Thermoacoustic theory for a single capillary tube is extended to random bulk medium on the basis of capillary tubes. The characteristics of the porous stack inside the resonator such as the tortuosity, dynamic shape factor, and porosity are introduced for the extension of wave equation by following Attenborough's approach. Separation of the dynamic shape factor for the viscous and thermal effect is adopted and scaling using the dynamic shape factor and tortuosity factor is demonstrated. The theoretical and experimental comparison of thermoviscous functions in reticulated vitreous carbon (RVC) and aluminum foam shows reasonable agreement. The extension is useful for investigations of the properties of a stack with arbitrary shapes of non-parallel pores.

11:10

2aPA9. Numerical simulations comparing nonlinear acoustic effects in straight and tapered tubes. Brian C. Tuttle and Victor W. Sparrow (Grad. Prog. in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802)

In their efforts to improve the efficiency of an orifice pulse tube refrigerator, Olson and Swift [Cryogenics **37**, (769–776 (1997)] reported on the benefit of using an optimal taper angle in the construction of the conical pulse tube. They concluded that this geometry effectively sup-pressed nonlinear acoustic streaming which was interfering with the desired heat transfer within the device. To investigate further the effects of tube geometry on streaming, a numerical simulation is developed for a tube with varying degrees of taper angle. This 2-D finite-difference model, based on the work of Sparrow and Raspet [J. Acoust. Soc. Am. **90**, 2683–2691 (1991)], compares nonlinear acoustic propagation in a straight cylindrical tube with that in a tapered conical tube. Development of the model equations in cylindrical and spherical coordinates as well as refinement of the computational grid near the boundaries are considered. [Work supported in part by ONR.]

Note: A panel discussion is scheduled at the end of session 2pPA.

TUESDAY MORNING, 18 OCTOBER 2005

CONRAD C, 8:00 A.M. TO 12:00 NOON

Session 2aPP

Psychological and Physiological Acoustics: Psychological Acoustics (Poster Session)

Robert S. Schlauch, Chair

Univ. of Minnesota, Speech-Language-Hearing Sciences, 115 Shevlin Hall, 164 Pillsbury Ave., SE, Minneapolis, MN 55455

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

2aPP1. Determining which primate audiograms are reliable for comparing sensitivity in different species. Mark N. Coleman (Stony Brook Univ., Health Sci. Tower A-8069, Stony Brook, NY 11794, mcoleman@ic.sunysb.edu)

This study investigated all known audiograms of nonhuman primates going back to the study by Elder in 1934. In total, audiograms have been published for 28 species of nonhuman primates, with macaques being the most intensively studied. The aim of this project was to determine which audiograms represent reliable measures of sensitivity by examining inter-laboratory, intraspecific, and individual variability. Auditory thresholds were taken from the audiograms by digitally scanning each graph and then extracting amplitude and frequency data using SIGMASCAN PRO image measurement software. The precision of this method was verified by comparing extracted data with published threshold values from numerous studies. The results indicate that several patterns are consistently illustrated regardless of which techniques were used, although there is some degree of variability in audiograms produced in different laboratories. For example, several studies have found that anthropoids have better low-frequency hearing than prosimians, and nearly all monkeys show the 4-kHz dip. Another interesting finding was that a lack of low-frequency sensitivity in certain taxa (e.g., patas monkeys) cannot simply be explained as the result of using headphones instead of loudspeakers. With judicious usage, audio-grams can be useful for comparing biologically meaningful differences in primate hearing patterns.

2aPP2. Thresholds for a tone masked by constant, remote-frequency maskers. Lynne A. Werner (Dept. Speech & Hearing Sci., Univ. of Washington, 1417 N.E. 42nd St., Seattle, WA 98105-6246) and Lori J. Leibold (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

In an experiment examining informational masking in infants, Leibold (2004) reported that infants and adults exhibited masking of a tone by two tones of constant, remote frequencies. Neither informational nor energetic masking would be expected for adults under these conditions. In the present experiment, adults' detection of a tone masked by two constant frequency tones or two narrow bands of noise were compared. Fifteen young adults participated. The target tone was 1000-Hz, 300-ms duration with 16-ms rise/fall. Unmasked and two masked thresholds were estimated. The maskers were either tones of 581 and 2920 Hz or 50-Hz-wide noise bands centered at 556 and 2895 Hz. The duration of the maskers was 300 ms with 16-ms rise/fall. A one-interval adaptive procedure was used to estimate thresholds. The maskers repeated at 600-ms intervals. The target tone was presented simultaneously with the maskers on signal trials. Listeners received feedback whenever they correctly detected the target. About 75% of listeners demonstrated masking in both conditions, but the threshold in the two-noise condition was generally higher than in the two-tone condition. Thus, introducing a timbre difference between the target and masker did not reduce the amount of masking.

2aPP3. Effects of reduced spectral uncertainty and masker fringes with multitonal maskers. Lori J. Leibold, Donna L. Neff, and Walt Jesteadt (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131)

This study examined the degree to which masker spectral uncertainty contributes to the high thresholds often observed with random-frequency, multitonal, simultaneous maskers. Masked thresholds for four normal-hearing adults were measured for a 1000-Hz, 200-ms signal presented with broadband-noise or multitonal maskers with 2 or 10 components. The multi-tonal maskers were 70 dB SPL and had a 160-Hz spectral gap cen-

tered at the signal. Two sets of conditions, using a 2AFC adaptive task, manipulated the degree of spectral uncertainty (masker randomized with each presentation vs fixed throughout a block of trials) and the temporal offset of the masker (masker equal to signal duration vs. masker increased to 400 ms, with a 100-ms fringe preceding and following the signal). A third set of ''molecular'' conditions, using a yes-no adaptive task, measured thresholds with five samples of multitonal maskers, either randomizing samples across trials or using one sample throughout trial blocks. For both large-set and molecular conditions, results showed as much as 30-dB release from masking when spectral uncertainty was reduced for fixed compared to random conditions. Fixed-frequency conditions, however, showed a further masking release with the masker fringe, with effects as large as 12 dB, variable across listeners. [Work supported by NIDCD.]

2aPP4. Unmasking the harmonics of a complex tone. Matthew J. Goupell and William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 4230 BPS Bldg., East Lansing, MI 48824)

Alternately eliminating and reintroducing a harmonic of a complex tone can cause that harmonic to stand out as a pure tone—separately audible from the rest of the complex-tone background. In the psychoacoustical literature the effect is known as ''enhancement.'' Further, during those intervals in which a harmonic is eliminated, listeners can hear out a neighboring harmonic—separately audible from the background. The latter effect has been called ''unmasking'' or ''monaural unmasking.'' The first quantitative evidence for the monaural unmasking effect has been obtained using a pitch matching experiment. For a 200-Hz complex tone with equal-amplitude (45 dB SPL) harmonics in sine phase, harmonics out to the 16th can be unmasked and matched. Pulsing harmonic n almost always unmasks $n+1$ but for some listeners and some harmonics, $n-1$ is systematically unmasked instead. Changing the relative phases among the harmonics greatly changes the probability that $n+1$ or $n-1$ will be unmasked. Interpretation of the data is somewhat complicated by pitch shifts, normally repelling the unmasked harmonic away from the pulsed harmonic. Classical place pitch models from the 1970s account for some of the features of unmasking, but not for the phase effects. [Work supported by the NIDCD under grant DC00181.]

2aPP5. Critical ratio for tones, modulated tones and noise, and speech. Andrew Lovitt and Jont Allen (Beckman Inst., MC 251, 405 N. Mathews, Urbana, IL 61801)

The critical ratio (CR) is a very basic concept first introduced in 1930 [Fletcher and Munson, (1933)], but rarely studied today. In this study we have remeasured the CR for a host of signals, including long duration tones, AM modulated tones at various modulation depths, at 3 Hz modulation, at several levels, modulated noise at two modulation depths levels, and speech signals of several bandwidths. A simple peak detection model explains most of the data. Some subject differences are also of interest.

2aPP6. Across-frequency loudness matches with pure tones and narrow-band noises. Melanie J. Gregan and Robert S. Schlauch (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. S.E., Shevlin 115, Minneapolis, MN 55455, grega005@umn.edu)

Loudness matches between stimuli that differ in frequency are difficult, presumably due to the strong pitch differences which make it challenging to focus on loudness exclusively. It is hypothesized that the use of stimuli with a less salient pitch percept (but that still permit relatively frequency-specific excitation of the cochlea) would allow for more accurate across-frequency loudness matches. This would permit more stable comparisons of loudness growth between frequency regions (corresponding to normal and elevated thresholds) as is often the need in studies involving hearing-impaired subjects. In this study, normal-hearing listeners were evaluated in loudness matching tasks with two types of stimuli: pure tones and narrow bands of noise (with corresponding center frequen-

cies). Listeners matched the loudness of a stimulus in one frequency region with a stimulus in another frequency region. The stimuli were either two pure tones, two narrow-band noises, or a pure tone and a narrow-band noise. The results will be evaluated based on predictions from an excitation pattern model and transitivity to determine if more accurate predictions are possible using narrow-band noise stimuli. [Work supported by the University of Minnesota.]

2aPP7. Informational masking and loudness. Elin M. Roverud and Robert S. Schlauch (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, 115 Shevlin Hall, Minneapolis, MN 55455, rove0011@umn.edu)

In a recent study, Neff and Odgaard [J. Acoust. Soc. Am. **116**, 3051–3061 (2004)] found that distracter tones with frequencies sampled randomly from Gaussian distributions interfered with frequency discrimination for target tones centered among the distracters. When the distracter tones were fixed in frequency, the interference was minimal. The goal of the present study was to examine the effect of distracter tones on the perceived loudness of a target tone. Magnitude estimations for 2.0-kHz tones as a function of level were made in isolation, in the presence of fixed-frequency distracters, and in the presence of distracters with frequencies drawn randomly from Gaussian distributions, as in Neff and Odgaard (2004). The number of distracters was either 1, 2, or 4. In a follow-up study, loudness for a frequency-modulated target tone was measured as a function of level in isolation and in the presence of unmodulated distracter tones that were also either fixed in frequency or randomly selected from a Gaussian distribution. Results will be discussed in terms of energetic and informational masking based on the loudness predictions of an excitation-pattern model. [Work supported by the UROP program, University of Minnesota.]

2aPP8. Introduction and practical use of an algorithm for the calculation of pitch strength. Markus Fruhmann (AG Technische Akustik, MMK, TU Munich, Germany, fruhmann@ei.tum.de)

Pitch strength is a psychoacoustic magnitude describing how strong the pitch of a sound can be perceived. Introduced in 1979 by Fastl and Stoll, a large variety of experimental data has been collected, but no model has been established so far to describe the data found. In the talk, some data to give an overview of pitch strength will be shown. Based on a transform taking into account the properties of the human auditory system, namely the Fourier-t-transform, an algorithm was developed to calculate pitch strength. Calculated data will be compared to the corresponding experimental data showing the potential of the algorithm. Furthermore, the measure of pitch strength will be discussed in the context of tonality which describes the perceivability of single tonal components within a sound. Tonality is subject of several industrial standards (e.g., DIN 45 681) to judge the sound quality of noises from mechanical structures and machines. Pitch strength data show several similarities when compared to tonality, but as a more global percept it also shows differences especially in the cases where virtual pitch is evoked.

2aPP9. Development of comodulation masking release. Cynthia M. Zettler and Marsha G. Clarkson (Dept. of Psych., Georgia State Univ., Univ. Plaza, Atlanta, GA 30303)

The comodulation masking release (CMR) paradigm permits assessment of across-channel auditory processing. To evaluate the development of this phenomenon, CMR was measured in 81 children ($M = 8.95$ years) and 20 adults ($M = 23.3$ years). A single-interval, maximum-likelihood procedure was used to estimate thresholds in a reference condition (signal plus on-signal AM masker) and a modulated masker condition (reference condition plus eight comodulated flanking bands). CMR was defined as improvement in signal detection in the modulated masker condition relative to the reference condition. Although total

CMR did not change from 7 to 10 years of age, total CMR for children (4.02 dB) was significantly less than that for adults (10.85 dB), suggesting that across-channel processing develops beyond 10 years of age. In addition, thresholds for children were significantly higher than those for adults in both the reference condition (by 6.47 dB) and the modulated masker condition (by 13.91 dB), supporting the likelihood that processing efficiency also develops beyond 10 years of age. These results suggest that both processing efficiency and across-channel cues utilized in CMR are slowly developing phenomena. [Work supported by a grant from the Georgia State University Research Program Enhancement Program.]

2aPP10. Modulation detection interference with periodic complex modulators. Stanley Sheft and William A. Yost (Parmly Hearing Inst., Loyola Univ. Chicago, 6525 North Sheridan Rd., Chicago, IL 60626)

Past studies of modulation detection interference (MDI) have considered the possibility that the effect may in part relate to auditory grouping of the probe and masker. Alternatively as a cross-channel effect, MDI may be representative of informational masking. The present experiments varied modulator waveshape in attempts to distinguish the effects of auditory grouping and informational masking on MDI. In the first experiment, both the slope and duty cycle of the masker modulator varied across conditions with the task to detect either 4- or 10-Hz sinusoidal probe modulation. The effect of masker-modulator waveshape on the detection of 4-Hz probe modulation was generally consistent with change in the perceptual prominence of the masker. The 10-Hz results showed little effect of these manipulations. The second experiment utilized interrupted modulation patterns to allow for either concurrent or sequential envelope fluctuation of the probe and masker. Results showed no effect of modulation concurrency. In both experiments, significant departures from energetic masking in the modulation domain were obtained. Collectively, these results were taken to support an interpretation of the interference in terms of informational masking rather than grouping based on common modulation. [Work supported by NIDCD.]

2aPP11. Effects of rate-of-change pattern on the discrimination of unidirectional gliding tones with fixed frequency and time transition spans. Pei-Chun Li, Shih-Tsang Tang, and Shuenn-Tsong Young (Inst. of Biomed. Eng., Natl. Yang-Ming Univ., No. 155, Sec. 2, Li-Nung St., Taipei, Taiwan)

Listening experiments were designed to test the three hypotheses for detection and discrimination of glides in frequency: (1) end point sampling; (2) a weighted average method; and (3) decision based on changes in the low-frequency side of the excitation pattern. Forty-eight frequency and time transition spans were chosen. The center frequencies of testing signals were 500 Hz, 2 KHz, and 6 KHz, with frequency spans of 0, 0.5, 1, and 2 ERBs and durations of 20, 50, 200, and 400 ms. For a given transition span, the frequency difference limens for five different gliding patterns were measured, including linear chirp, quadratic and inverse quadratic sweep chirps, and FM modulated sweeps. All test patterns had the following features: (1) they are unidirectional gliding tones with increasing instantaneous frequency throughout the whole duration; (2) the phase functions are at least first-order differentiable, thus the interference of click sounds caused by fast frequency transitions are alleviated. The results showed that none of these hypotheses was able to explain the mechanism for detection and discrimination of glides in frequency well. It is speculated that, at least, both sides of the excitation patterns should be compared, and that the time-related cues could have an effect as well.

2aPP12. Detection and discrimination of monaural and binaural increments and decrements of intensity with a temporal fringe. Mark A. Stellmack, Neal F. Viemeister, and Andrew J. Byrne (Dept. of Psych., Univ. of Minnesota, 75 East River Rd., Minneapolis, MN 55455, stell006@umn.edu)

Detection thresholds are larger for brief, monaural intensity increments in long-duration stimuli than for decrements. This experiment assessed the contribution of a decrement in one ear to direction discrimination of an interaural intensity difference (IID), where a brief intensity increment was presented at one ear and an equal-magnitude intensity decrement at the other. In two-interval monaural tasks, thresholds were measured for detection of an increment, a decrement, and discrimination of the two. In single-interval binaural tasks, thresholds were measured for identifying ear of presentation of an increment, a decrement, and for identification of direction of IID achieved by presenting an increment and decrement to opposite ears. In all cases, thresholds were measured for a 5-ms probe segment as a function of its temporal position in a 50-ms stimulus (broadband noise). The major results: (1) Thresholds for discriminating the ear of presentation of a decrement were much larger than those for simply detecting the decrement monaurally; (2) Interaural thresholds were slightly better when both increment and decrement were present relative to increment alone. These results indicate that the decrements contributed to binaural discrimination, although by themselves the same decrements were detectable but, at best, weakly lateralizable. [Work supported by NIDCD DC00683.]

2aPP13. Roles of frequency information and spatial information in auditory search. Yôiti Suzuki, Masato Saze (Res. Inst. of Elec. Commun. and Grad. School of Information Sci., Tohoku Univ., Sendai, Japan, yoh@ais.riec.tohoku.ac.jp), Yoichi Sugita (Natl. Inst. of Adv. Industrial Sci. and Technol., Tsukuba, Japan), Noriaki Asemi (Kobe Univ., Kobe, Japan), and Shuichi Sakamoto (Tohoku Univ., Sendai, Japan)

Effects of attention on auditory search process were examined with five spatially distributed loudspeakers, each of which radiated a component sound (either a pure tone or a 1/12-octave NBN). The stimulus consisted of five component sounds with different (center) frequencies. The listener's task was to judge whether or not a target sound was included in the distractors as quickly as possible while minimizing errors [N. Asemi *et al.*, Acta Acustica **89**, 346–354 (2003)]. The following three conditions were examined: (1) a control condition with no cue sound; (2) a condition with a preceding cue sound indicating the target sound location; and (3) a probe-signal condition in which 1/8 of the cue sounds provided inconsistent information of frequency and location (probes) while the remainder provided consistent information (primaries). The resultant response times for probes were generally longer than those for primaries. Furthermore, attention windows were observed in both frequency and spatial domains. The effect of the disparities between the cue and target was larger when both the frequency and location differed than when only one of them differed. This result also means that the detection performance becomes better if attention is directed with certain information related to the targets.

2aPP14. Coupling of head and body movements to acoustic flow in sighted adults. Thomas A. Stoffregen, ChungGon Kim (School of Kinesiology, Univ. of Minnesota, 1900 University Ave. SE, Minneapolis, MN 55455, tas@umn.edu), Kiyohide Ito (Future Univ., Hakodate, Japan), and Benoit G. Bardy (Univ. of Montpellier I, Montpellier, France)

Blindfolded sighted persons were found to detect acoustic flow patterns and use this information to control action. A moving room (a large box on wheels, with no floor, that moved in the subject's fore–aft axis) was used. Blindfolded sighted persons (1) stood comfortably or (2) moved their head backward and forward to track audible room motion. Pink noise was presented through four speakers attached to the room, or mounted on stationary stands. Room motion was a sinusoid at 0.2 Hz, 22 cm, along subject's fore–aft axis. When standing comfortably, participants exhibited

weak but consistent coupling of body sway with room motion. Tracking of room motion with head movements was robust, matching both the frequency and amplitude of room motion. This was true even when the only information about room motion came from reflected sound (i.e., when the speakers were stationary). The results suggest a strong ability of sighted persons to use acoustic flow in the perception and control of their own action. [Work supported by NSF (BCS-0236627).]

midrange point constrains learning. Subjects were able to learn optimal or near-optimal boundary placement for overlapping distributions only when the ideal boundary was near the midrange point and when those distributions were presented during the initial blocks of training in the experiment. [Work supported by NSF and NIDCD.]

2aPP15. Sensitivity to changing stimulus distribution characteristics in auditory categorization. Sarah C. Sullivan (Dept. of Psych., Univ. of Texas, Austin, Texas 78712-0187, sullivan@psy.utexas.edu), Andrew J. Lotto (Boys Town Natl. Res. Hospital, Omaha, NE 68131), Elizabeth T. Newlin, and Randy L. Diehl (Univ. of Texas, Austin, Texas 78712-0187)

This experiment examined the ability of humans to categorize sounds as a function of changing training distribution characteristics. Participants were presented non-speech sounds randomly sampled from two overlapping distributions. The sounds consisted of 25 narrow-band noise bursts varying in center frequency from 1000 to 1360 Hz. Four different distribution pairs were created by varying the ratio of stimuli in each distribution (i.e., prior probabilities), resulting in different distribution cross-over points or ideal boundaries. Subjects identified the distribution from which each sound was drawn and were provided feedback. The training distributions were switched midsession without alerting the subjects. Performance was tracked by plotting identification functions and noting boundary placement for each individual block. Previously collected data suggested that the ability to learn optimal boundaries was determined in part by the proximity of the distribution boundary to the stimulus midrange point. Results from the current experiment also suggest that distance from the

2aPP16. Model for noise-induced hearing loss using support vector machine. Wei Qiu, Jun Ye, Xiaohong Liu-White, and Roger P. Hamernik (Auditory Res. Lab., State Univ. of New York at Plattsburgh, 101 Broad St., Plattsburgh, NY 12901)

Contemporary noise standards are based on the assumption that an energy metric such as the equivalent noise level is sufficient for estimating the potential of a noise stimulus to cause noise-induced hearing loss (NIHL). Available data, from laboratory-based experiments (Lei *et al.*, 1994; Hamernik and Qiu, 2001) indicate that while an energy metric may be necessary, it is not sufficient for the prediction of NIHL. A support vector machine (SVM) NIHL prediction model was constructed, based on a 550-subject (noise-exposed chinchillas) database. Training of the model used data from 367 noise-exposed subjects. The model was tested using the remaining 183 subjects. Input variables for the model included acoustic, audiometric, and biological variables, while output variables were PTS and cell loss. The results show that an energy parameter is not sufficient to predict NIHL, especially in complex noise environments. With the kurtosis and other noise and biological parameters included as additional inputs, the performance of SVM prediction model was significantly improved. The SVM prediction model has the potential to reliably predict noise-induced hearing loss. [Work supported by NIOSH.]

Session 2aSC

Speech Communication: Perception and Production in Linguistically Diverse Populations (Poster Session)

Gary G. Weismer, Chair

Univ. of Wisconsin, Waisman Ctr., 1500 Highland Ave., Madison, WI 53705-2280

Contributed Papers

All posters will be on display from 8:45 a.m. to 11:45 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:45 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 11:45 a.m.

2aSC1. The influence of native language phonology on auditory and visual word recognition in Russian−English bilinguals. Valeriy Shafiro (Dept. of Comm. Disord. and Sci., Rush Univ. Medical Ctr., 203 Senn, 1653 W. Congress Pkwy., Chicago, IL 60612) and Anatoliy Kharkhurin (Brooklyn College and The Grad. Ctr. of The City Univ. of New York, Brooklyn, NY 11210)

Visual and auditory recognition of English monosyllabic words was examined in 21 native Russian bilinguals and 12 monolingual speakers of American English. Stimuli comprised 40 CVC minimal pairs distinguishable by four vowel contrasts. Experiment 1 tested visual word recognition following a semantic categorization task. Sixty of the tested words were previously shown in the categorization task, and 60 were new words, 40 of which represented minimal pair alternatives to previously shown words. In experiment 2 participants listened to all words from each minimal pair spoken by one male and one female speaker, and selected one word from each pair. No significant differences were found in the number of errors made by the bilingual and monolingual participants during visual word recognition. In auditory word recognition, bilingual listeners made significantly more errors on the two vowel contrasts that cannot be differentiated based on their native phonological categories. Auditory errors on these categories strongly and significantly correlated with participant's age of arrival to the U.S.A. These results demonstrate the influence of native language phonology on auditory, but not visual word recognition by Russian−English bilinguals.

2aSC2. Interaction between speech and numerical processing: A behavioral and fMRI study in bilingual adults. Jo-Fu Lotus Lin, Rajeev Raizada, Toshiaki Imada, Patricia K. Kuhl (Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98195-7988), and Toshiaki Imada (Tokyo Denki Univ., Inzai, Chiba, Japan)

Studies suggest that exact mathematical terms (numbers, multiplication tables) are stored in the language in which they were originally learned. Bilingual adults are less accurate and slower when doing exact mathematical calculations using their non-dominant language. Approximate mathematical calculations appear not to be affected by language. The aim of this study is to (a) examine the relationship between brain activation during exact and approximate mathematical tasks in bilinguals using both languages, and (2) relate brain activation during mathematical tasks to brain activation during phonetic processing in both languages. We hypothesize that in both phonetic and mathematical tasks, measures of neural efficiency [Zhang *et al.*, (2005)] will show an effect of the first versus second language. We will examine (a) the extent to which different types of calculation activate language areas, and (b) how language (number word) structures affect arithmetic processing. Two groups of bilingual adults (Chinese-English and Spanish-English) are presented with calculation problems during fMRI scans. The behavior-brain results highlight the importance of identifying the role of language, at the most basic level of phonetics, in bilingual numerical processing. Relevance of these results to a theoretical framework of language and thought will be discussed (Work funded by NSF.)

2aSC3. The perceptual magnet effect (PME) in L2 consonants. Kota Hattori (Dept. of Second Lang. Studies, Univ. of Hawaii at Manoa, 1890 East-West Rd., Honolulu, HI 96822, khattori@hawaii.edu)

The perceptual magnet effect (PME) has been investigated in conjunction with L1 vowels and consonants. However, there has been a paucity of examinations of the effect in L2 speech perception. The present study used the methodology of Iverson and Kuhl [P. Iverson and P. K. Kuhl, J. Acoust. Soc. Am. **99**, 1130–1140 (1996)] to examine whether Japanese speakers who have had exposure to English-speaking environments for more than 1 year exhibit the PME for /r/ and /l/. Eighteen synthetic stimuli were generated. Ten Japanese speakers and 12 American English speakers participated in an identification and rating task. Each participant engaged in ten experimental sessions (ten sessions of each of the 18 stimuli). They chose either /ra/ or /la/ and gave a rating with a 1–7 scale. They also underwent a similarity scaling task to determine the similarity of 306 pairs of the 18 stimuli with a 1–7 scale. The results indicated that, while the English speakers demonstrated the PME, some native speakers of Japanese did exhibit the PME and prefer some exemplars of English /r/ and /l/ over others. The results also revealed that other Japanese speakers exhibited reorganized perceptual space without the PME.

2aSC4. Contrastive and contextual vowel nasalization in Ottawa. Marie Klopfenstein (Wayne State Univ., 5057 Woodward, Detroit, MI 48202, ab2771@wayne.edu)

Ottawa is a Central Algonquian language that possesses the recent innovation of contrastive vowel nasalization. Most phonetic studies done to date on contrastive vowel nasalization have investigated Indo-European languages; therefore, a study of Ottawa could prove to be a valuable addition to the literature. To this end, a percentage of nasalization (nasal airflow/oral + nasal airflow) was measured during target vowels produced by native Ottawa speakers using a Nasometer 6200-3. Nasalized vowels in the target word set were either contrastively or contextually nasalized: candidates for contextual nasalization were either regressive or perserverative in word-initial and word-final syllables. Subjects were asked to read words containing target vowels in a carrier sentence. Mean, minimum, and maximum nasalance were obtained for each target vowel across its full duration. Target vowels were compared across context (regressive or perseverative and word-initial or word-final). In addition, contexts were com-

1897   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1897**

pared to determine whether a significant difference existed between contrastive and contextual nasalization. Results for Ottawa will be compared with results for vowels in similar contexts in other languages including Hindi, Breton, Bengali, and French.

2aSC5. Spectral characteristics of voiceless fricatives in Arabic. Ali S. Abu-Al-Makarem and Donald S. Cooper (Dept. of Commun. Disord., Bowling Green State Univ., 200 Health Ctr., Bowling Green, OH 43403)

The purpose of this study was to investigate the acoustics of the voiceless fricatives of Al-Khat spoken Arabic. The FFT of the eight Arabic voiceless fricatives followed by /a: i: u:/ was investigated in four healthy adult males. The speech samples were collected in a sound-attenuated booth using a microphone with a frequency response flat 0.5 dB from 10 Hz to 11 000 Hz, and digital recording. All recordings were sampled at a rate of 24 kHz. The FFT was computed for each waveform token using a 1024-point Hamming window. A 42.67-ms analysis window located in the middle of the fricative was used to compute each spectrum. The effects of vowel context on all examined acoustic parameters appeared minimal. The most effective acoustic parameter distinguishing fricatives place of articulation was spectral skewness, followed by spectral mean and spectral standard deviation of the FFT spectra. Spectral skewness distinguished the seven places of articulation for Arabic voiceless fricatives. Skewness increased as place of articulation moved from the front to the back of the vocal tract. This result extended the findings of Forrest *et al.* (1988) and Jongman *et al.* (2000) that spectral skewness reliably distinguished English fricative spectra from four places of articulation.

2aSC6. The production of the /d/−/t/ distinction in French at slow speaking rate. Nassima Abdelli-Beruh (Dept. Speech-Lang. Pathol., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003, nab4@nyu.edu)

This study examined how monolingual French speakers produced the /d/−/t/ distinction at normal and slow speaking rates. Target syllables were preceded by a voiceless fricative (/s/). Voicing-related differences in durations of preceding vowel, /s/, stop closure, and VOT were calculated and analyzed as a function of speaking rate (stressed syllables spoken at normal speaking rate, stressed syllables spoken at low speaking rate). Percentages of /s/ and /d/ closures with voicing were tallied in each speaking rate condition. Results from ANOVA showed that the absolute durations in sentence, preceding vowel, /s/, and closure were larger in the slow speaking rate condition than in the normal speaking rate condition. The voicing of the /s/ preceding /d/ and during the closure of /d/ occurred systematically in the normal speaking rate condition. However, in the slow speaking rate condition, most speakers failed to phonate the /s/, but partially phonated the /d/ closures. The effects of slow speaking on the /d/−/t/ distinction are compared with the effects obtained in the unstressed speaking rate condition.

2aSC7. The influence of stress on the /d/−/t/ distinction in French. Nassima Abdelli-Beruh and Radhika Aravamudhan Aravamudhan (Dept. of Speech-Lang. Pathol. and Audiol., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003, nab4@nyu.edu)

This study examined how monolingual French speakers produced the /d/−/t/ distinction in stressed and unstressed syllable-initial stops preceded by a voiceless phone (/s/). Syllables were embedded in sentences. Sentence durations and voicing-related differences in durations of preceding vowel, /s/, stop closure, and VOT were calculated and analyzed as a function of the stress condition separately for each speaker (stressed syllables spoken at normal speaking rate, unstressed syllables produced at normal speaking rate). Preliminary analyses reveal that the vowel and the voiceless fricative preceding the unstressed target syllables were longer than the vowel and the fricative preceding the stressed target syllables. Closure durations were also longer in the unstressed condition than in the stressed

condition. However, voicing-related duration differences were not systematically affected by stress. Finally, the voicing of /s/ (/s/ before /d/) and of /d/ closures, which occurred frequently in the stress condition, occurred less frequently in the unstressed condition for most of the speakers. The perceptual consequences of such results remain to be investigated.

2aSC8. Variation in Rioplatense Spanish /s/ + stop sequences. Michal Temkin Martinez (Linguist. Dept., Univ. of Southern California, GFS 301, 3601 Watt Way, Los Angeles, CA 90089-1693)

This acoustic study of Rioplatense Spanish examines the lenition of syllable-onset, voiced stops after /s/. Four Rioplatense speakers from Buenos Aires participated in an acoustic experiment by reading sentences with the target stop occurring in phrase-initial, word-initial, and word-medial positions in both a conversational and clear speech style. Stop tokens at all three places of articulation were analyzed for the presence of voicing, of a stop-closure interval, and of a release burst. The variable realization of the preceding coda /s/ in all conditions was also transcribed and analyzed, to evaluate co-occurrence patterns in the sequence. The results indicated frequent occurrences of stop spirantization in the conversational style and in phrase-medial positions. Further, variants of /s/ containing only a laryngeal gesture, such as [h], co-occur with spirantized variants of the voiced stops. An articulatory phonology account is presented of the variation in stop realization, as well as of the relationship between the /s/ variant and its following stop. [Work supported by NIH.]

2aSC9. English vowel production by native Mandarin speakers: Influences of AoA, LoR, education, perception, and orthography. Fredericka Bell-Berti (Dept. of Speech, Commun. Sci., & Theatre, St. John's Univ., 8000 Utopia Pky., Jamaica, NY 11439) and Yan Helen Yu (CUNY, New York, NY 10016)

This study investigates relations among several factors that are expected to influence vowel production in second language learning, including AoA, LoR, L2 and general education, L2 perception, and orthography. Vowel production will be examined through duration and formant frequency measurements and listener identification. The results will be analyzed in relation to educational background and language use. Among the educational factors examined are general education level, English education (in their native land and/or New York City), and sound-annotating system experiences in Mandarin (Pinyin or Zhuyin). The language-use factors include AoA, LoR, language spoken at work and at home, and perception of English vowels. The hypotheses addressed include: (1) educational background, language use, and sound-annotating system experiences in Mandarin all influence L2 English speakers perception and production of English vowels; (2) the more accurately an L2 listener discriminates a vowel contrast, the more distinctly he/she produces that contrast.

2aSC10. To play or to pray?−The influences of initial exposure age and environment on perception of Mandarin speakers of English on devoiced liquids. Sally Chen and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., 1, Roosevelt Rd., Sec. 4, Taipei 106, Taiwan)

The devoicing rule of liquids after voiceless aspirated stops presents difficulty for L2 English learners. This study investigates Mandarin speakers' perception of the rule with regards to initial exposure age and environment. Three initial exposure ages were included in this study: kindergarten, elementary, and junior high school. Except for the last group, each group was further divided into two exposure environments: Taiwan and the U.S. As a result, five groups of people were included in total, resulting in 25 subjects. Stimuli were pseudo-words in an SLV structure, recorded by one native English speaker and one non-native English speaker. Half of the stimuli began with a voiceless stop, and the other half began with a voiced stop. The native speakers pronounced the voiceless liquids accordingly while the non-native speaker only had voiced liquids. Preliminary

results showed that, in general, L2 learners took longer time in perceiving devoiced liquids. Listeners who had earlier exposure were more likely to respond faster to native speech while late learners responded better to non-native speech.

2aSC11. The role of linguistic experience on the perception of phonation. Christina Esposito (Dept. of Linguist., UCLA, Los Angeles, CA 90095)

This study investigates the role linguistic experience has on the perception of phonation and the acoustic properties that correlate with this perception. Listeners from three languages (Gujarati, which contrasts breathy versus modal vowels, Italian, which has no breathiness, and English, which has allophonic breathiness) participated in two tasks. In the Visual Sort task, breathy and modal vowels from a variety of languages (e.g., Chong, Mon, etc.) were presented as icons on a computer screen. Subjects sorted the icons (the stimuli) into two groups based on perceived similarity of the talker's voices. In the multidimensional scaling task, listeners heard pairs of Mazatec vowels, and moved an on-screen slider to indicate the perceived similarity of the vowels in each pair. Results will show that judgments were more uniform across subjects who had breathy categories present in their native language(s), and varied across subjects who lack a breathy category. The perceived similarity among the stimuli will correlate with a measurable acoustic property (H1-H2, H1-H2, H1-A3, H1-A1 or H1-A2, the average of H1-H2 compared to A1, and A2-A3). It is hypothesized that H1-H2 will be the most salient acoustic property for Gujarati listeners because this correlates with their production of breathy vowels.

2aSC12. Reference frame and scaling in the perception of pitch in a tone language. Heriberto Avelino (Dept. of Linguist., Univ. of California at Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

Listeners can make subtle judgments about the relative magnitudes of auditory objects but only rarely make judgments of the absolute size of speech stimuli. In a pioneering paper, and his first publication in 1957, Peter Ladefoged demonstrated that the relative size of the surrounding context affects the perception of vowel quality. Although, years later, Ladefoged showed that the effect is only valid under certain conditions, the effect is still a motivated perceptual phenomenon. In this paper, the role of the reference frame in the perception of pitch in Yalalag Zapotec is investigated. YZ has three phonemic tones, high, low, and falling. Minimal pairs of H and L tones were introduced by a carrier sentence. The experiment manipulated the $F0$ of the Zapotec equivalent sentence to "Please say what this word is," while the target words were kept unmodified. Other things being equal, the only difference was the $F0$ of six versions of the introductory sentence. Preliminary results are analogous to Ladefoged's original findings on the identification of vowels: listeners' pitch judgments are influenced by the overall auditory context in which they occur. The results suggest that perception of pitch for listeners of tone languages depends on allocentric frames of reference.

2aSC13. Use of spectral change and duration cues in vowel identification by monolingual and bilingual listeners: Evidence from confusion matrices. Catherine L. Rogers, Merete M. Glasbrenner, and Teresa DeMasi (Dept. of Commun. Sci. & Disord., Univ. of South Florida, 4202 E. Fowler Ave., PCD1017, Tampa, FL 33620, crogers@chuma1.cas.usf.edu)

The degree to which vowel duration and time-varying spectral change cues are used in American English vowel identification was investigated for three groups of listeners: (1) monolingual American English listeners; (2) proficient Spanish-English bilinguals; and (3) less proficient Spanish-English bilinguals. Digital manipulation and high-fidelity synthesis procedures (STRAIGHT) were used to create six versions of six target items (bead, bid, bayed, bed, bad, and bod). These consisted of the unaltered whole word and five versions of the isolated vowel: (1) natural vowel—unaltered; (2) synthetic vowel with no cue alteration; (3) synthetic vowel with neutral duration; (4) synthetic vowel with flattened formants; and (5) synthetic vowel with neutral duration and flattened formants. Isolated-vowel stimuli were presented to listeners for identification in two 120-trial blocks, followed by 48 whole-word stimuli. Significant between-group differences in percent-correct performance and a significant group by listening condition interaction were found, suggesting that cue weighting may vary across groups and that some bilinguals' phoneme representations may be less robust than those of monolinguals. Analysis of confusion matrix data will be used to further explore these hypotheses by examining patterns of confusions across group and condition and extent of individual differences across subjects within a group. [Work supported by NIH-NIDCD 1R03DC005561-01A1.]

2aSC14. Acoustic and perceptual correlates of syllable weight. Matthew Gordon, Carmen Jany, and Carlos Nash (Dept. of Linguist., Univ. of California Santa Barbara, Santa Barbara, CA 93106, mgordon@linguistics.ucsb.edu)

Differences between languages in the stress-attracting properties of various syllable types (syllable weight) are associated with phonetic differences. Certain languages that preferentially stress CVC syllables (i.e., treat CVC as heavy) fail to display substantial vowel shortening in CVC, unlike languages that treat CVC as non-stress-attracting or light [Broselow *et al.* (1997)]. Furthermore, CVC has greater energy (intensity integrated over time) in languages in which it is heavy relative to languages with light CVC [Gordon (2002)]. This paper compares multiple potential acoustic and perceptual correlates of syllable weight. A representative cross section of syllable types in words uttered by speakers of four languages was recorded. In two languages (Arabic, Hindi), CVC is heavy; in two languages (Mongolian, Malayalam), CVC is light. Three measurements were taken: duration of the syllable rime, acoustic intensity integrated over the rime, and a measure of perceptual energy of the rime incorporating various factors (e.g., temporal integration and adaptation, bandpass filtering). Results thus far indicate that a measure of prominence factoring in both intensity and duration better distinguishes languages on the basis of weight criterion than a simple measure of duration. The perceptual energy measure provides a slightly better fit than acoustic energy. [Work supported by NSF.]

2aSC15. Factors affecting the perception of Korean-accented American English. Kwansun Cho, John G. Harris (Dept. of Elec. and Computer Eng., Univ. of Florida, Gainesville, FL 32611), and Rahul Shrivastav (Univ. of Florida, Gainesville, FL 32611)

This experiment examines the relative contribution of two factors, intonation and articulation errors, on the perception of foreign accent in Korean-accented American English. Ten native speakers of Korean and ten native speakers of American English were asked to read ten English sentences. These sentences were then modified using high-quality speech resynthesis techniques [STRAIGHT; Kawahara *et al.*, Speech Commun. 27, 187–207 (1999)] to generate four sets of stimuli. In the first two sets of stimuli, the intonation patterns of the Korean speakers and American speakers were switched with one another. The articulatory errors for each speaker were not modified. In the final two sets, the sentences from the Korean and American speakers were resynthesized without any modifications. Fifteen listeners were asked to rate all the stimuli for the degree of foreign accent. Preliminary results show that, for native speakers of American English, articulation errors may play a greater role in the perception of foreign accent than errors in intonation patterns. [Work supported by KAIM.]

2aSC16. Korean students' repetition of English sentences mixed with noise at various speed. Byunggon Yang (Dept. of English Education, Pusan Natl. Univ., 30 Changjun-dong, Keumjung-gu, Pusan 609-735, South Korea) and Eunjee Kim (Dongeui Univ., Pusanjin-gu, Pusan 614-714, South Korea)

Many Korean students feel quite frustrated by native speakers' fast talk, especially in a noisy situation. In this study, it was examined how well 33 Korean university students repeat 48 recorded English sentences at normal speed, fast speed, normal speed mixed with noise, and fast speed mixed with noise. The subject's production was scored by the percent ratios of correctly produced words to the total words in each sentence. Results showed that overall the student group correctly repeated around 65% of all the words, while the native speakers demonstrated almost perfect performances. Students achieved 69% and 61% accuracy in their repetition of content and function words, respectively. The high-proficiency group outperformed the low-proficiency group, particularly in their repetition of function words. Student's accuracy of repetition remarkably dropped when both the speed and noise of normal sentences were modified. The noise condition seemed to have the strongest effect on their accuracy. Finally, it was observed that the more words a stimulus sentence had, the harder the Korean students found it to repeat correctly. Those results may be helpful for English teachers to develop better teaching materials for listening skills and to offer some criteria of proficiency level in English.

2aSC17. Identification of American English vowels by native Japanese speakers: Talker-and-token-based analysis. Takeshi Nozawa (Faculty of Economics, Ritsumeikan Univ., 1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan), Elaina M. Frieda (Auburn Univ., Auburn, AL 36849), and Ratree Wayland (Univ. of Florida, Gainesville, FL 32611)

Native speakers of Japanese identified American English vowels /i, ɪ, ɛ, æ, ɑ, ʌ/ produced by four female native speakers of American English in /CVC/ contexts. Native speakers of American English served as the control group, and they outperformed the Japanese subjects in identifying all the English vowels in every /CVC/ context. In another experiment the Japanese subjects equated these English vowels with Japanese vowels. In general, English vowels were equated with phonetically close Japanese vowels, but significant talker effect was observed. The /i/ tokens equated with the Japanese long high front vowel /ii/ were much more correctly identified as /i/ than those equated with the Japanese short high front vowel /i/. These tokens were more often misidentified as /ɪ/. The /ɑ/ and /ʌ/ tokens were predominantly equated with the Japanese low vowel /a/. The percent-correct identification of /ɑ/ and /ʌ/ was low in most of the /CVC/ contexts, and these two vowels were often misidentified as each other, and the Japanese subjects' latency before they decided what vowel they had heard was longer when /ɑ/ or /ʌ/ tokens were presented. The Japanese subjects do not seem to have salient cues to differentiate /ɑ/ and /ʌ/.

2aSC18. Does clear speech enhance the voice onset time contrast in Croatian and English? Rajka Smiljanic and Ann Bradlow (Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208)

This study investigates the effect of intelligibility-enhancing clear speech on voice onset time (VOT) in two languages with different phonetic realizations of a two-way voicing contrast. Specifically, it was explored whether the contrasts between prevoiced and short-lag categories in Croatian and short-lag and long-lag categories in English are enhanced along the VOT dimension in clear speech. Word-initial stops in words embedded in sentences read by native speakers of each language in conversational and clear speaking styles were examined. The results revealed that in English, the long-lag category (i.e., voiceless) extended somewhat in the positive VOT direction (longer aspiration), and in Croatian, the prevoiced category (i.e., voiced) extended somewhat in the negative VOT direction (longer prevoicing) in clear speech. However, in both languages, the short-lag tokens (i.e., voiced in English, voiceless in Croatian) were unchanged in clear versus conversational speech. Overall, then, the dis-

tance between the two voicing categories was rather stable across the two speaking styles in both languages. These results suggest that VOT is not a dimension of voicing category contrast enhancement. Rather, the language-specific pronunciation norms along this dimension are maintained in clear and conversational speech.

2aSC19. Tuning in and tuning out: Speech perception in native- and foreign-talker babble. Kristin Van Heukelem and Ann R. Bradlow (Linguist. Dept., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, k-van@northwestern.edu)

Studies on speech perception in multitalker babble have revealed asymmetries in the effects of noise on native versus foreign-accented speech intelligibility for native listeners [Rogers *et al.*, Lang Speech **47**(2), 139–154 (2004)] and on sentence-in-noise perception by native versus non-native listeners [Mayo *et al.*, J. Speech Lang. Hear. Res., **40**, 686–693 (1997)], suggesting that the linguistic backgrounds of talkers and listeners contribute to the effects of noise on speech perception. However, little attention has been paid to the language of the babble. This study tested whether the language of the noise also has asymmetrical effects on listeners. Replicating previous findings [e.g., Bronkhorst and Plomp, J. Acoust. Soc. Am., **92**, 3132–3139 (1992)], the results showed poorer English sentence recognition by native English listeners in six-talker babble than in two-talker babble regardless of the language of the babble, demonstrating the effect of increased psychoacoustic/energetic masking. In addition, the results showed that in the two-talker babble condition, native English listeners were more adversely affected by English than Chinese babble. These findings demonstrate informational/cognitive masking on sentence-in-noise recognition in the form of linguistic competition. Whether this competition is at the lexical or sublexical level and whether it is modulated by the phonetic similarity between the target and noise languages remains to be determined.

2aSC20. Second language vowel training using vowel subsets: Order of training and choice of contrasts. Kanae Nishi and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Our previous vowel training study for Japanese learners of American English [J. Acoust. Soc. Am. **117**, 2401 (2005)] compared training for two vowel subsets: nine vowels covering the entire vowel space (9V condition); and the three more difficult vowels (3V condition). Trainees in 9V condition improved on all vowels, but their identification of the three more difficult vowels was lower than that of 3V trainees. Trainees in 3V condition improved identification of the trained three vowels but not the other vowels. In order to further explore more effective training protocols, the present study compared two groups of native Korean trainees using two different training orders for the two vowel subsets: 3V then 9V (3V-9V); and 9V then 3V (9V-3V). The groups were compared in terms of their performance on all nine vowels for pre-, mid-, and post-test scores. Average test scores across the two groups were not different from each other. A closer examination indicated that group 3V-9V did not improve on one of the three more difficult vowels, whereas group 9V-3V improved on all three vowels, indicating the importance of training subset order. [Work supported by NIH DC-006313 and DC-02229.]

2aSC21. Lexical status effects on compensation for fricative assimilation. Claudia Kuzla, Holger Mitterer, and Anne Cutler (Max Planck Inst. for Psycholinguist., P.O. Box 310, 6500 AH Nijmegen, The Netherlands)

In German, word-initial voiced fricatives may be devoiced if they follow /t/. This assimilation could make recognition of words beginning with /v/ harder, since there may be competing words beginning with /f/, but will not affect words beginning with /z/, since initial /s/ in German is illegal. Previous research has shown that speakers indeed produce more

assimilation across word boundaries than across phrase boundaries for /z/, but inhibit the assimiliation across word boundaries for /v/. In the present study, German listeners identified the fricative continua /f–v/ and /s–z/, across word-versus-phrase boundaries, in viable-versus-nonviable contexts for assimilation. Less voicing was required for a /v,z/ judgment in viable than in nonviable assimilation contexts. This context effect was larger after a word boundary than after a phrase boundary. Within the viable-context condition, a prosodic effect appeared for /f–v/, with less voicing required for /v/ judgments after a word than a phrase boundary, but no such effect appeared for /s–z/. This asymmetry reverses the difference observed in production. Thus, listeners adjust phoneme category boundaries to compensate for prosodically conditioned variation where such adjustment is functional for word recognition, but show less compensation where adjustment would have no functional consequences.

2aSC22. Lexical statistics of competition in L2 versus L1 listening. Anne Cutler (Max Planck Inst. for Psycholinguist., P.O. Box 310, 6500 AH Nijmegen, The Netherlands)

Spoken-word recognition involves multiple activation of alternative word candidates and competition between these alternatives. Phonemic confusions in L2 listening increase the number of potentially active words, thus slowing word recognition by adding competitors. This study used a 70,000-word English lexicon backed by frequency statistics from a 17,900,000-word corpus to assess the competition increase resulting from two representative phonemic confusions, one vocalic (ae/E) and one consonantal (r/l), in L2 versus L1 listening. The first analysis involved word embedding. Embedded words (*cat* in *cattle*, *rib* in *ribbon*) cause competition, which phonemic confusion can increase (*cat* in *kettle*, *rib* in *liberty*). The average increase in number of embedded words was 59.6 and 48.3 temporary ambiguity. Even when no embeddings are present, multiple alternatives are possible: *para-* can become *parrot, paradise*, etc., but also *pallet, palace* given /r/–/l/ confusion. Phoneme confusions (vowel or consonant) in first or second position in the word approximately doubled the number of activated candidates; confusions later in the word increased activation by on average 53 third, 42 confusions significantly increase competition for L2 compared with L1 listeners.

2aSC23. Some novel allophonic and phonemic phenomena in Biscayan Basque. Rebeka Campos-Astorkiza (Linguist. Dept., Univ. of Southern California, University Park GFS 301, Los Angeles, CA 90089-1693, rebekaca@usc.edu)

An acoustic study of novel allophonic and phonemic phenomena in the isolate language Basque is presented. The focus is on speakers of the Biscayan dialect. First, Basque shows a spirantization process by which voiced plosives are produced as approximants, particularly intervocalically. Interestingly, we find that Basque /ld/ sequences, where spirantization is not expected [Hualde (1991) Basque Phonology], are realized as a lateral approximant followed by a voiced lateral fricative. Second, in this variety of Basque, the historical three-way contrast among sibilants (two alveolars and one postalveolar) has been reduced to a two-way distinction. The original contrast, still found in other varieties, between a laminal alveolar and an apical alveolar has merged with different results depending on the continuancy of the sibilants. Third, Basque presents a contrast between trill and flap intervocalically. However, elsewhere this is neutralized, and the precise realization of this segment varies from trill to frication. Finally, the Basque five-vowel inventory allows for almost any sequence of two vowels. The same vowel sequence might be a diphthong (tautosyllabic) or a hiatus (heterosyllabic) depending on the lexical item. That is, diphthongs and hiatus are contrastive. [Submitted for the student paper competition.]

2aSC24. Acquisition of stress and pitch accent in English−Spanish bilingual children. Sahyang Kim, Jean Andruski (Dept. of Audiol. and Speech-Lang. Pathol., Wayne State Univ., Detroit, MI 48202), Geoffrey S. Nathan, Eugenia Casielles, and Richard Work (Wayne State Univ., Detroit, MI 48202)

Although understanding of prosodic development is considered crucial for understanding of language acquisition in general, few studies have focused on how children develop native-like prosody in their speech production. This study will examine the acquisition of lexical stress and postlexical pitch accent in two English−Spanish bilingual children. Prosodic characteristics of English and Spanish are different in terms of frequent stress patterns (trochaic versus penultimate), phonetic realization of stress (reduced unstressed vowel versus full unstressed vowel), and frequent pitch accent types (H^* versus L^*+H), among others. Thus, English−Spanish bilingual children's prosodic development may provide evidence of their awareness of language differences relatively early during language development, and illustrate the influence of markedness or input frequency in prosodic acquisition. For this study, recordings from the children's one-word stage are used. Durations of stressed and unstressed syllables and $F0$ peak alignment are measured, and pitch accent types in different accentual positions (nuclear versus prenuclear) are transcribed using American English ToBI and Spanish ToBI. Prosodic development is compared across ages within each language and across languages at each age. Furthermore, the bilingual children's productions are compared with monolingual English and Spanish parents' productions.

2aSC25. Prosodic resolution of a syntactic ambiguity in Korean learners of English. Hyekyung Hwang and Amy J. Schafer (Univ. of Hawaii at Manoa, 569 Moore Hall, 1890 East-West Rd., Honolulu, HI 96822)

Research has discussed disambiguating effects of the relative strength of prosodic boundaries in native-language sentence processing (e.g., Carlson *et al.*, 2001; Schafer *et al.*, 2000), but not in second language processing. In this study, Korean learners of English (L2ers) and native English speakers (L1ers) listened to the initial portion of sentences with a temporary ambiguity in clause-boundary location (e.g., when that moves the square), and chose one of two visually presented disambiguating continuations [(1) it will . . . (2) will . . .]. The stimuli were quasispontaneously produced by untrained speakers, with prosody that varied in boundary strength at the correct and competing clause-boundary locations. As expected, pronunciations with the strongest prosodic boundary at the correct clause-boundary location successfully disambiguated most tokens for both L1ers and L2ers. Contrary to predictions of Carlson *et al.*, 2001, both groups were sensitive to the size of the correct versus competing boundary strength difference: e.g., a contrast between a correct-location intonation phrase boundary and a competing-location word boundary was more disambiguating than a contrast between an intonation and an intermediate phrase boundary. Additionally, L2ers showed some difficulty in utilizing intermediate phrase boundaries as disambiguating cues, suggesting a greater learning component to intermediate than intonation phrase use for L2ers.

2aSC26. New and contrastive focus in Taiwan Mandarin. Jia-rong Li (Dept. of Foreign Lang. and Lit., Natl. Chiao Tung Univ., 1001 Ta Hsueh Rd., Hsinchu, Taiwan)

This study investigated the influence of information structure on surface acoustical parameters in Taiwan Mandarin. There were two experiments, i.e., GN and GC experiment. In GN experiment, the noun phrases placed in initial (NP1) or medial (NP2) position in a Mandarin sentence were elicited through a game. Subjects answered experimenter's questions, e.g., ''What is on pink truck's right side?'' according to colored pictures in a 4 by 4 matrix. New or given information was placed on either NP1 or NP2. In GC experiment, subjects corrected or confirmed experimenter's description, such as Pink truck is yellow IDs right side, by placing either contrastive information on the color term and given information

on noun (cg), given information on the color term and contrastive information on noun (gc), given information on both color term and noun (gg), or contrastive information on both color term and noun (cc). Results showed that the $F0$ range expansion and duration lengthening were observed on syllables carrying new and contrastive information. Moreover, duration lengthening is more pronunced than $F0$ range expansion. Furthermore, syllables with contrastive information were lengthened more than syllables with new information.

2aSC27. Acoustic correlates of Japanese pitch accent. Yukiko Sugiyama (Dept. of Linguist., Univ. at Buffalo, The State Univ. of New York, 609 Baldy Hall, North Campus, Buffalo, NY 14260, ys8@buffalo.edu)

Acoustic correlates of Japanese pitch accent were investigated using bimoraic/disyllabic pairs of words that are identical except for their accent patterns. For example, /hana/ ''flower'' and /hana/ ''nose'' have the same phonological shape and pitch levels (low-high) yet they differ in that ''flower'' is accented but ''nose'' is unaccented. The most notable acoustic difference between the two accent patterns is realized by the $F0$ of the following word when there is one. There have been debates about whether the two accent patterns are acoustically different when the following word is excluded. It has been proposed that the $F0$ maximum is higher, and/or the $F0$ movement is larger for the accented words. While previous studies used only a few pairs of words, this study tests all 20 pairs of words found by searching a computerized dictionary [Amano and Kondo (1999)] that had a relatively high word familiarity. The $F0$ maximum and movement

were measured in the 20 pairs of words as produced by native speakers of Tokyo Japanese (males and females) in isolation and two frame sentences. Implications of their results for the nature of accent in Japanese and the perception of accent will be discussed.

2aSC28. The prosody of the compound words in Hong Kong Cantonese. Eric Zee (Dept. of CTL, City Univ. of Hong Kong, 83 Tat Chee Ave., Kowloon, Hong Kong, eric.zee@cityu.edu.hk)

The bisyllabic, trisyllabic, and quadrisyllabic compounds in Hong Kong Cantonese were analyzed for their prosodic characteristics, including temporal structure, $F0$ contour, and intensity contour. Results show that the durations of the syllable-initial fricatives and the voice onset time of the aspirated stops and aspirated affricate of the component syllables vary according to the positions in the compounds in which the syllables occur. As for the unaspirated stops and affricate, they are not affected by the positions of the component syllables. The rime durations which also vary according to their positions in the compounds are longest for the final syllables to be followed by the preceding syllables in decreasing order. In the case of the quadrisyllables, the difference in rime duration between the initial and second component syllables is insignificant. On the basis of the findings, temporal models for the compounds are constructed. The $F0$ contours of the citation tones on the component syllables may be subject to the $F0$ declination effect. The shapes of the citation tones may change, but in most cases the citation forms are retained. With a few exceptions, the intensity curves for the rimes co-vary with the $F0$ contours.

TUESDAY MORNING, 18 OCTOBER 2005   MARQUETTE ROOM, 8:15 A.M. TO 12:00 NOON

Session 2aUW

Underwater Acoustics, Signal Processing in Acoustics and Engineering Acoustics: Sonar Performance and Signal Processing in Uncertain Environments I

Lisa M. Zurk, Cochair
Portland State Univ., Electrical and Computer Engineering Dept., 1900 SW Fourth Ave., Portland, OR 97207

David R. Dowling, Cochair
Univ. of Michigan, Dept. of Mechanical Engineering, 2019 Automotive Lab. 2133, 1231 Beal Ave., Ann Arbor, MI 48109-2133

Chair's Introduction—8:15

Invited Papers

8:20

2aUW1. Seabed variability and its influence on acoustic prediction uncertainty. Charles W. Holland (Appl. Res. Lab., The Penn State Univ., North Atherton St., State College, PA 16801), Brian Calder, Barbara Kraft, Larry Mayer (Univ. of New Hampshire, Durham, NH 03824), John Goff (Univ. of Texas, Austin, TX 78705), and Chris Harrison (NATO Undersea Res. Ctr., La Spezia, Italy)

Kevin LePage (Naval Research Laboratory, Washington, DC), Robert I. Odom (University of Washington, Applied Physics Laboratory), Irina Overeem, James Syvitski (University of Colorado, INSTAAR, Boulder, CO) and Lincoln Pratson (Duke University, Durham, NC). The weakest link in performance prediction for naval systems operating in coastal regions is the environmental data that drive the models. In shallow-water downward refracting environments, the seabed properties and morphology often are the controlling environmental factors. In order to address the issue of uncertainty in seabed properties, we focused on two overarching goals: (1) assess and characterize seafloor variability in shelf environments, (2) determine the impact of the seafloor variability on acoustic prediction uncertainty. Our inherently multidisciplinary approach brought marine geology/geophysics and ocean acoustics together at the intersection of geoacoustic modeling. This talk will review results from a 3-year collaboration under the ONR Capturing Uncertainty DRI. [Work supported by the Office of Naval Research.]

2aUW2. Some issues concerning numerical simulation as a predictive tool for assessing sonar performance in uncertain environments. Steven Finette (Acoust. Div., Naval Res. Lab, Washington, DC 20375, finette@wave.nrl.navy.mil)

One goal of numerical simulation in ocean acoustics is the accurate prediction of acoustic field properties in complex environments. In this context, the objective assessment of both sonar performance and associated signal-processing schemes operating within uncertain littoral environments pose significant challenges for numerical simulation. Some of the issues involved with using numerical models for the prediction of sonar performance in the presence of incomplete environmental information are addressed. To achieve credibility as a prediction tool, the simulation results should be held to the same standards as those associated with experimental results, i.e., the numerical predictions should include error bars or confidence intervals that represent a quantitative measure of the environmental uncertainty that is always present. Providing a rational, quantitative assessment of system uncertainty in complex environments is a nontrivial task. Incomplete environmental information occurs at multiple space-time scales within both the oceanographic and acoustic modeling phases; this uncertainty can propagate, interact, and amplify in the course of transmitting and processing the signal. It is argued that environmental uncertainty should be directly incorporated into the numerical simulation process so that it links directly to the physics of wave propagation. [Work supported by NRL Base funding through ONR.]

9:10

2aUW3. Assessing uncertainty in ocean acoustic inverse problems. Robert I. Odom and Andrew A. Ganse (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

The goal of solving ocean acoustic inverse problems is not just to find a model fitting the data. We can always fit an N-1th order polynomial to N data points, but it does not tell us much about the uniqueness, or limits placed on resolution in the presence of noise. The real goal is to develop more information than just a model that fits the data. For truly linear problems, constraints on the model information and uniqueness are characterized by the null space of the operator which maps the model to the data, the statistics are Gaussian, and the resolution matrices have well-defined meanings. Unfortunately most inverse problems of interest in ocean acoustics are nonlinear. We are faced with methods such as simulated annealing or genetic algorithms, which tell us nothing about the statistics and little about the uniqueness of our model solution, or exhaustive but numerically intensive Monte Carlo analyses. As an alternative we have examined a nonlinear filter, that is an extension of the Kalman and extended Kalman filters. We discuss the background of the ocean acoustic inverse problem for bottom properties, and how it can be addressed by employing an exact nonlinear filter. [Work supported by ONR.]

9:35

2aUW4. Sonar detection performance prediction of a moving source in an uncertain environment. Jeffrey Krolik and Vivek Varadarajan (Dept. of Elec. and Computer Eng., Duke Univ., Box 90291, Durham, NC 27708)

This paper addresses the problem of passive sonar detection of a moving acoustic source when both the signal spatial wavefront and background noise field are uncertain. The classic sonar equation is derived for a stationary source-receiver geometry and known environment. For dynamic scenarios, however, array gain (AG) and transmission loss (TL) can fluctuate in time and frequency over the integration time and bandwidth used. This is caused by signal multipath group and phase velocity differences at the receiver array. Moreover, for adaptive methods, noise covariance matrix snapshot support is often limited. This paper presents methods for estimating TL and AG under these conditions from a moving source of opportunity. These estimates are then used to predict adaptive detection performance. Results for probability of detection versus signal-to-noise ratio using real horizontal array data from the Swellex-96 experiment are shown to compare favorably with theoretically predicted detection performance. [Work supported by ONR.]

10:00–10:15   Break

Contributed Papers

10:15

2aUW5. A study of the effect of sound velocity variability on received signal level. H. John Camin, David L. Bradley, R. Lee Culver, and Leon H. Sibul (Appl. Res. Lab. and Grad. Program in Acoust., P.O. Box 30, State College, PA 16804, hjc130@psu.edu)

A detailed survey of the sound velocity profiles (SVP) in the Strait of Gibraltar was performed in 1996 over several tidal cycles. The Strait of Gibraltar is a particularly complicated environment where tides and internal waves create strong depth-, range-, and time-dependent SVP structure. In such an environment, to obtain a mean SVP, spatial and temporal averages, obscure the detail of the profile. Such averaging displays considerably more variability and randomness than a set of SVPs all taken at the same location and point in the tidal cycle. The maximum entropy (ME) method is used to estimate the probability density function (pdf) of the environmental uncertainty in two extremes: a known point in the tidal cycle with low variability, and an unknown point using the time-averaged SVPs. The uncertainty is then propagated 15 km across the strait using a parabolic equation code in order to predict the pdf's of the received signal level. Thus, the essential environmental uncertainty is represented by the distribution of the received signal. The predictions are compared with available acoustic data. [Gibraltar data and model results generously provided by Chris Tiemann, ARL:UT, and Peter Worcester, SIO. Work supported by ONR Code 321US.

10:30

2aUW6. Estimating mode coupling effects on shallow-water propagation uncertainty. Kevin D. LePage (NRL Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

The introduction of unknown aberrations to the background sound speed of the ocean waveguide and sediments, as well as unknown free-surface excursions and bottom bathymetry, causes uncertainty in the characteristics of acoustic time series. Not all of this uncertainty can be adequately explained by simply assuming that energy in various paths is perturbed only in phase and travel time. Data and simulations both clearly show that significant mode coupling is introduced by even small aberrations introduced over significant ranges. This talk estimates the effects of mode coupling introduced by statistically characterized sound-speed defects on the statistical properties of time series in shallow water. In this approach EOF decompositions of the sound-speed field are used, along with the differential form of the coupled -modeequations and perturbation

expressions for mode-coupling coefficients. Results are compared to previously derived expressions for the statistical properties of time series obtained under the adiabatic approximation [Work supported by ONR.]

10:45

2aUW7. Angle-of-arrival estimation in atmospheric turbulence.
Sandra Collier (U.S. Army Res. Lab., ATTN: AMSRD-ARL-CI-ES, 2800 Powder Mill Rd., Adelphi, MD 20783-1197), Vladimir Ostashev (NOAA Environ. Technol. Lab., Boulder, CO 80305), and D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755-1290)

An angle-of-arrival estimator is derived for acoustic waves that propagate in atmospheric turbulence. The estimator directly accounts for the effects of random fluctuations in the propagation environment. The first and second moments of the sound field, as derived from the theory of wave propagation in a random medium, and realistic atmospheric turbulence models are utilized. Results using this estimator are compared to results using conventional beamformers that do not account for the effects of atmospheric turbulence.

11:00

2aUW8. Probability density function predictions of acoustic uncertainty in an underwater sound channel.
Kevin R. James and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 2019 Lay Auto Lab., 1231 Beal Ave., Ann Arbor, MI 48109)

The utility of forward acoustic propagation models is often limited by incomplete knowledge of environmental parameters. For this reason, information about the resulting uncertainty of acoustic field calculations is a valuable addition to the solutions generated by current propagation routines. However, extensive Monte Carlo simulations are presently the only means for quantifying this uncertainty. This presentation explores alternative methods for obtaining a complete description of the uncertainty of a time-harmonic acoustic field solution in the form of a probability density function (PDF). The isobaric contours of the field solution, including the uncertain parameters as independent variables, produce the relationship between acoustic pressure and the uncertain variables. In an ocean waveguide, the isobaric contours can be obtained approximately from a modal sum Green's function using techniques similar to those that yield the waveguide invariant. Thus, information from a single field calculation can be used to transform PDFs for the uncertain environmental parameters into an uncertainty distribution for the calculated field solution. Results are shown for an ideal underwater sound channel and compared to results from Monte Carlo simulations. [Work sponsored by the ONR, Code 321.]

11:15

2aUW9. Experimental demonstration of time-reversed reverberation focusing in a rough waveguide. Application to target detection.
Karim G. Sabra, Philippe Roux, Hee-Chun Song (MPL, Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0238, ksabra@mpl.ucsd.edu), William Hodgkiss, William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238), Tuncay Akal (Earth and Marine Sci. Res. Inst., Kocaeli, Turkey), and Mark R. Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

For most shallow-water waveguides, the backscattered energy measured in a monostatic configuration is dominated by ocean bottom reverberation. A selected time-gated portion of the measured reverberation signals is used to provide a transfer function between a time-reversing array and a corresponding range interval on the bottom. Ultrasonic and at-sea experiments demonstrate focusing capabilities along the rough bottom interface of a time-reversing array using these reverberation signals only. The iterative time-reversal technique facilitates robust focusing along the ocean bottom, with little signal-processing effort involved and *a priori* information on the environment, and the enhancement of detection and localization of proud or buried targets in complex shallow-water environments. A passive implementation of the iterative time-reversal processing is used to construct reflectivity maps, similar to a sonar map, but with an enhanced contrast for the strongest reflectors (or scatterers), at the water–bottom interface. Ultrasonic and at-sea experiments show that targets lying on the seafloor located up to 400 wavelengths from the time-reversing array were detected over the bottom reverberation.

11:30

2aUW10. Effects of uncertain geoacoustic parameters and coastal shipping densities on shipping noise directionality.
Thomas J. Hayward and Richard M. Heitmeyer (Naval Res. Lab., Washington, DC 20375)

The bearing-elevation directionality of low-frequency shipping noise is influenced both by the sediment geoacoustic parameters and by the coastal shipping density. This study examines the effects of geoacoustic parameter and shipping density uncertainties on noise directionality through simulations for a North Pacific site. The simulations are based in part on stochastic models of the spatial variations of geoacoustic parameters that incorporate: (1) conditional probability densities of the geoacoustic parameters given mean sediment grain size, derived from measurement compilations reported by Hamilton *et al*. [J. Acoust. Soc. Am. **72**, 1891], and (2) location-dependent probability densities of the mean grain size, derived from field measurements. [Carlson *et al*., U.S. Geological Survey Map MF 876]. The simulations also incorporate coastal shipping noise source uncertainties based on models of commercial fishing vessel locations and ship-radiated noise. [Wales and Heitmeyer, J. Acoust. Soc. Am. **111**, 1211] The results quantify the effects of the geoacoustic and coastal shipping uncertainties on received low-frequency bearing-elevation noise directionality. Variances in the noise power at horizontal arrival due to geoacoustic uncertainties substantially exceed those associated with uncertainties in the shipping noise source distributions and levels. These results are interpreted in terms of down-slope propagation effects. [Work supported by ONR.]

11:45

2aUW11. Determining the sensitivity of measured acoustic quantities on variability in geophysical and oceanographic conditions.
Sean Pecknold, Paul Hines, and John Osler (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, sean.pecknold@drdc-rddc.gc.ca)

Sonar is used to remotely investigate the underwater environment and to detect and track vessels therein, either by their own acoustic emissions or through scattering from active transmission. The acoustic signals available to an observer are a function of the transmitter's relative position, course, and signature, and the local environment. In order to make effective use of the received acoustic signals, an observer requires a thorough understanding of the propagation and scattering characteristics of the overall underwater environment as well as the implications of those characteristics on the analysis techniques applied to the received signals. Defense R&D Canadas (DRDC) Rapid Environmental Assessment (REA) Program aims to provide a capability for accumulation and interpretation of environmental information in a tactical timeframe. A main objective of the REA program is to explore the nature of geophysical and oceanographic variability and to quantify its effect on acoutic signals. This effect is analyzed via modeling and through the use of data from the joint DRDC/NURC sea trial BASE 04 (Broadband Acoustic Sonar Experiment 2004). This trial, which took place in May and June of 2004 in the Malta Plateau and Medina Bank areas, included several experiments measuring active sonar propagation in an uncertain environment.

Session 2pAB

Animal Bioacoustics and Psychological and Physiological Acoustics: Cognition in the Acoustic Behavior of Animals II

Caroline M. Delong, Chair

Brown Univ., Dept. of Neuroscience, Box 1953, Providence, RI 02912

Chair's Introduction—1:00

Invited Papers

1:05

2pAB1. Measuring the "acoustic gaze" of echolocating bats, and using it to access internal processes in an active sensing system. Kaushik Ghose and Cynthia Moss (NACS, Psych., ISR, Univ. Maryland, College Park, MD 20742, kghose@umd.edu)

The sonar beam of an echolocating bat spatially filters the information it receives from the environment. It is proposed that the sonar beam direction can serve as an index to the cognitive state of an echolocating bat. In this respect the sonar beam may be considered an analog to visual gaze. Gaze direction spatially restricts visual information, and has been used as an index to study cognitive states of visual animals. A method to measure the sonar beam direction of a flying echolocating bat is outlined. Laboratory studies show, when presented with a single target, the bat tightly locks its sonar beam to the target before starting the high repetition rate (buzz) phase of insect pursuit. While maneuvering through a small gap, echolocating bats inspect the edges of the gap before flying through it. When presented with two targets, the bat may probe both targets with the center of its beam before changing its flight path to attack one. These studies support the hypothesis that the sonar beam of the echolocating bat may be used as an index of overt, spatial attention, much like gaze direction in naturally behaving visual animals.

1:25

2pAB2. Object representation in echolocating bats: Recognition of targets from different orientations. Caroline M. DeLong, Rebecca Bragg, and James A. Simmons (Dept. of Neurosci., Brown Univ., Box 1953, Providence, RI 02912, Caroline_Delong@brown.edu)

Big brown bats (*Eptesicus fuscus*) can discriminate among objects using wideband FM sonar sounds. Bats could perceive and represent only the acoustic dimensions of echoes (e.g., delay, frequency, amplitude), or they could use those dimensions to reconstruct object features (e.g., distance, shape, size). To investigate this question, bats were presented with a two-alternative (left/right) forced-choice sonar discrimination task. The stimuli were a one-cylinder monopole target and a two-cylinder dipole target that was presented at all aspect angles. Acoustic dimensions of the dipole echoes change depending on aspect angle. If the bats represent only acoustic dimensions, they should have difficulty selecting the dipole. If the bats represent object features, they should be able to select the dipole independent of aspect angle. Bats can select the dipole over a wide range of aspect angles, suggesting that they construct object features from echo acoustic dimensions. The bats' error patterns may indicate the echo dimensions they use to construct object features. [Work supported by NIH and ONR.]

1:45

2pAB3. Classification across the senses: Auditory-visual cognitive performance in a California sea lion (*Zalophus californianus*). Kristy L. Lindemann, Colleen Reichmuth-Kastak, and Ronald J. Schusterman (UCSC, Long Marine Lab., 100 Schaffer Rd., Santa Cruz, CA 95060)

The model of stimulus equivalence describes how perceptually dissimilar stimuli can become interrelated to form useful categories both within and between the sensory modalities. A recent experiment expanded upon prior work with a California sea lion by examining stimulus classification across the auditory and visual modalities. Acoustic stimuli were associated with an exemplar from one of two pre-existing visual classes in a matching-to-sample paradigm. After direct training of these associations, the sea lion showed spontaneous transfer of the new auditory stimuli to the remaining members of the visual classes. The sea lion's performance on this cross-modal equivalence task was similar to that shown by human subjects in studies of emergent word learning and reading comprehension. Current research with the same animal further examines how stimulus classes can be expanded across modalities. Fast-mapping techniques are used to rapidly establish new auditory-visual relationships between acoustic cues and multiple arbitrary visual stimuli. Collectively, this research illustrates complex cross-modal performances in a highly experienced subject and provides insight into how animals organize information from multiple sensory modalities into meaningful representations.

2p TUE. PM

2:05

2pAB4. Echoic functional class formation by a bottlenose dolphin. David Kastak and Ronald Schusterman (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060)

A bottlenose dolphin was trained to perform separate two-choice simple discrimination tasks using echolocation and vision. The targets were air-filled shapes constructed from PVC pipe. Targets varied in length and width but not depth, and were separated into classes of letters and numbers. Target echoes were measured, and classes were set up so that they could not be grouped on the basis of perceptual similarities between class members. The subject showed successful learning set formation (win-stay, lose-shift) in visual and echoic discrimination reversals consisting of specific target pairings. The subject also successfully grouped targets into classes as shown by performance in a discrimination reversal procedure involving sets of targets rather than individual pairings. The subject was unable to group the same targets using vision instead of echolocation. No evidence of bimodal performance enhancement or immediate cross-modal transfer was shown. The results suggest that dolphins can arbitrarily classify dissimilar targets using echolocation, but that performance can be independent of visual information. Successful cross-modal class formation likely results from high levels of experience associating echoic and visual stimuli, therefore is more likely to be shown in the context of a conditional discrimination rather than a simple discrimination task. [Work supported by ONR.]

2:25

2pAB5. Rhythm perception and production by the bottlenose dolphin. Heidi E. Harley, Sara E. Crowell (Div. of Social Sci., New College of Florida, Sarasota, FL 34243, harley@ncf.edu), Wendi Fellner, Kim Odell, and Leslie Larsen-Plott (Walt Disney World Resort, Lake Buena Vista, FL 32830)

Rhythm is an important component of many natural communication systems, but it has rarely been the focus of laboratory studies of nonhuman species. Recent cognitive studies with a bottlenose dolphin confirm that a dolphin can discriminate among six different 14-kHz 4-s acoustic rhythms at 94% accuracy, and can transfer that discrimination across multiple frequency (93%) and tempo (16%−93%) shifts. In addition, a dolphin has learned to produce six different rhythms in an object-labeling paradigm. Original training required the dolphin to produce the rhythms using a pneumatic switch that led to the in-air projection of computer-generated tones. However, the dolphin spontaneously began to produce the rhythms vocally as well. To date, the dolphin has accurately labeled five objects with unique rhythms at 87% accuracy using the switch and at 83% accuracy using his own vocalizations. Confusions at the various tempos in the perception study and the variability of some characteristics and stability of others in the production study provide insight into how dolphins represent rhythm and have implications for natural communication in this species.

2:45−3:00   Break

3:00

2pAB6. What's in a voice? Cues used by dolphins in individual recognition of signature whistles. Laela S. Sayigh (Biological Sci., UNCW, Wilmington, NC 28403, sayighl@uncw.edu), Vincent M. Janik (Univ. of St. Andrews, UK), and Randall S. Wells (Chicago Zoological Society, c/o Mote Marine Lab, Sarasota, FL 34236)

Cues that bottlenose dolphins may use for individual recognition of signature whistles are (1) the individually distinctive frequency modulation patterns of whistles; and (2) voice cues caused by the anatomy of the vocal apparatus. Experiments were designed to determine whether dolphins use either or both of these cues in recognizing whistles. Temporarily held wild dolphins listened to whistles of a close relative and of a known conspecific of the same sex and similar age. To test the hypothesis that dolphins recognize the frequency modulation patterns of whistles, signature whistles were synthesized and all general voice features removed. In playbacks to 14 individuals, dolphins turned significantly more often towards the speaker if they heard the synthetic signature whistle of a close relative than that of another individual. To test the hypothesis that dolphins may also be using voice cues to recognize whistles, natural variant (nonsignature) whistles were played back, which are highly variable in contour. Preliminary analysis of seven playbacks showed no difference in responses to variant whistles of kin versus nonkin. Thus, the frequency modulation pattern of signature whistles alone provides information on the identity of the caller, and voice cues are likely not used by dolphins to identify individuals.

Contributed Papers

3:20

2pAB7. The big brown bat's perceptual dimension of target range. James A. Simmons (Dept. of Neurosci., Brown Univ., Providence, RI 02912, james_simmons@brown.edu)

Big brown bats determine the distance to targets from echo delay, but information actually is entered onto the bat's psychological delay scale from two sources. The first is the target-ranging system itself, from the time that elapses between single-spike neural responses evoked by the broadcast and similar responses evoked by echoes at different delays. These responses register the FM sweeps of broadcasts or echoes, and the associated system of neural delay lines and coincidence detectors cross correlates the spectrograms along the time axis. The second source is the echo spectrum, which relates to shape expressed as range profile. The target-ranging system extracts this by fanning out to encompass parallel representations of many possible notch frequencies and notch widths in echoes. Bats perceive delay separations of $5-30$ μs and have a resolution limit of about 2 μs, but interference amplifies small delay separations by transposing them into large changes in notch frequency, so only perception of intervals smaller than 5 μs is surprising. Experiments with phase-shifted echoes show that the psychological time scale can represent two different delays originating entirely in the time domain when they are at least as close together as 10 μs. [Work supported by NIH and ONR.]

3:35

2pAB8. Methodological considerations of acoustic playbacks to test the behavioral significance of call directionality in male northern elephant seals (*Mirounga angustirostris*). Marla M. Holt, Stephen J. Insley (UC Santa Cruz Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060), Brandon L. Southall (UCSC Long Marine Lab. and NOAA Fisheries Acoust. Program, Silver Spring, MD 20910), and Ronald J. Schusterman (UC Santa Cruz Long Marine Lab., Santa Cruz, CA 95060)

While attempting to gain access to receptive females, male northern elephant seals form dominance hierarchies through multiple dyadic interactions involving visual and acoustic signals. These signals are both highly stereotyped and directional. Previous behavioral observations suggested that males attend to the directional cues of these signals. We used *in situ* vocal playbacks to test whether males attend to directional cues of the acoustic components of a competitors calls (i.e., variation in call spectra and source levels). Here, we will focus on playback methodology. Playback calls were multiple exemplars of a marked dominant male from an isolated area, recorded with a directional microphone and DAT recorder and edited into a natural sequence that controlled call amplitude. Control calls were recordings of ambient rookery sounds with the male calls removed. Subjects were 20 marked males (10 adults and 10 subadults) all located at Año Nuevo, CA. Playback presentations, calibrated for sound-pressure level, were broadcast at a distance of 7 m from each subject. Most responses were classified into the following categories: visual orientation, postural change, calling, movement toward or away from the loudspeaker, and re-directed aggression. We also investigated developmental, hierarchical, and ambient noise variables that were thought to influence male behavior.

TUESDAY AFTERNOON, 18 OCTOBER 2005

LA SALLE ROOM, 1:00 TO 5:35 P.M.

Session 2pAO

Acoustical Oceanography and Animal Bioacoustics: Ocean Ecosystem Measurements

Whitlow W. L. Au, Cochair
Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

D. Vance Holliday, Cochair
BAE System, 4669 Murphy Canyon Rd., San Diego, CA 92123-4333

Chair's Introduction—1:00

Invited Papers

1:05

2pAO1. Advancing from pattern to process in Hawaii's near-shore pelagic ecosystem. Kelly Benoit-Bird (College of Oceanic and Atmospheric Sci., Oregon State Univ., 104 COAS Admin. Bldg., Corvallis, OR 97331, kbenoit@coas.oregonstate.edu) and Margaret McManus (Univ. of Hawaii, Honolulu, HI 96822)

Micronekton comprising a near-shore sound-scattering layer around the Hawaiian Islands have been shown to exhibit diel horizontal migrations, moving onshore until midnight and them moving offshore into deep waters where they remain in the day. The reasons for this daily movement pattern have remained unexamined. Moored echosounders were used to describe micronekton migration over a 5-week period. A moored acoustic Doppler current profiler (ADCP), thermistor chains, point current meters, and an autonomous vertical profiler provided a description of the physical circulation. Periodic four-frequency echosounder and ADCP surveys were used to further characterize the micronekton's movement along with water column physics. During surveys, vertical profiles measured zooplankton abundance using the Tracor acoustic profiling system (TAPS) and integrated net tows, micronekton identity and size using an optical imaging system, and primary productivity with a fluorometer. It was found that micronekton are not simply tracers of water mass movement indicating active migration. Near-shore waters had significantly higher biomass of zooplankton during nighttime hours than offshore waters, suggesting micronekton have greater access to food resources as a result of their migration. Primary productivity was also significantly higher inshore, indicating bottom-up control of this system. These results suggest the underlying reasons behind this daily horizontal movement.

1:25

2pAO2. Bioacoustic absorption spectroscopy: The promise of classification by fish size and species. Orest Diachok (Poseidon Sound, 3272 Fox Mill Rd., Oakton, VA 22124), Paul Smith (Southwest Fisheries Sci. Ctr., La Jolla, CA), Stephen Wales (Naval Res. Lab., Washington, DC), and Carla Scalabrin (Ifremer, Brest, France)

A recently completed experiment, BAS II, was designed to test the validity of inferences of fish length from measurements of bio-alpha in frequency (250–5000 Hz)-depth (8–53 m) space. This experiment was conducted in the vicinity of sardine and anchovy spawning grounds at a 63-m-deep site where oceanic, biological, and geologic parameters were well calibrated. Classifications of absorption lines associated with 15-cm-long sardines, 11-cm-long anchovies, and 6-cm-long juvenile anchovies (recorded at night when most fish were dispersed) were consistent with trawl and historical data. Classification of an absorption line, possibly associated with juvenile sardines, was inconclusive, due to inconsistencies between trawl and historical data. Very low frequency lines may have been due to bubble cloud resonances associated with sardine schools, or larger fish which avoided sampling. The scintillation index

associated with bio-alpha lines, which is hypothetically driven by the rate of change of number densities, may offer additional classification clues, since fish dynamics may be species specific. Time series measurements over a few months would permit further testing and refinement of classification methods, a prerequisite for applications oriented BAS-based measurement systems, including observatories. [Work was supported by ONR.]

1:45

2pAO3. An explicit approach to detecting and characterizing submersed aquatic vegetation using a single-beam digital echosounder. Bruce M. Sabol (U.S. ERDC (EE-C), 3909 Halls Ferry Rd., Vicksburg, M.S. 39180, Bruce.M.Sabol@erdc.usace.army.mil)

There has been a longstanding need for an objective and cost-effective technique to detect, characterize, and quantify submersed aquatic vegetation at spatial scales between direct physical sampling and remote aerial-based imaging. Acoustic-based approaches for doing so are reviewed and an explicit approach, using a narrow, single-beam echosounder, is described in detail. This heuristic algorithm is based on the spatial distribution of a thresholded signal generated from a high-frequency, narrow-beam echosounder operated in a vertical orientation from a survey boat. The physical basis, rationale, and implementation of this algorithm are described, and data documenting performance are presented. Using this technique, it is possible to generate orders of magnitude more data than would be available using previous techniques with a comparable level of effort. Thus, new analysis and interpretation approaches are called for which can make full use of these data. Several analyses' examples are shown for environmental effects application studies. Current operational window and performance limitations are identified and thoughts on potential processing approaches to improve performance are discussed.

2:05

2pAO4. Passive acoustic monitoring of biological activity on coral reefs and in nearby waters. Marc O. Lammers, T. Aran Mooney (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734), Russell E. Brainard (Pacific Islands Fisheries Sci. Ctr., Honolulu, HI 96814), and Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

Monitoring the changing state of coral reef habitats is a challenging task that is exacerbated when the reefs in question are in remote locations. Physical sensors provide a wide range of measurements of local environmental variables, but do not give an indication of biological activity. The preliminary findings of an effort to use the ambient sound field as a means of characterizing and monitoring biological activity on coral reefs and surrounding waters are reported. Moored recording systems were developed to sample the sound field of reefs on Oahu, Hawaii for 1-min periods, at 30-min intervals, for 10 days at a time. Snapping shrimp produce the dominant acoustic energy on the reefs examined and exhibit clear diel acoustic trends. Peaks in activity consistently occur during crepuscular periods. At frequencies below 2 kHz, many fish sounds occur, which also exhibit distinct temporal variability. Cetacean sounds are also common, indicating the occurrence of an apex predator in the area. Many sounds can be detected automatically, making the examination of the sound field an efficient means of tracking acoustically active species. The results indicate that acoustic monitoring may be an effective means of tracking biological activity at locations where traditional surveys are impractical.

2:25

2pAO5. Intra- and interannual measurements of zooplankton biomass in the Gulf of Alaska using bioacoustical sensors. D. V. Holliday (BAE Systems, 4669 Murphy Canyon Rd., San Diego, CA 92123, van.holliday@baesystems.com), J. M. Napp (Alaska Fisheries Sci. Ctr., Seattle, WA 98115), C. F. Greenlaw (BAE Systems, San Diego, CA 92123), and P. J. Stabeno (Pacific Marine Environ. Lab., Seattle, WA 98115)

Sensors that monitor temperature and salinity in the sea are commonly deployed for long periods on moorings. Meteorological measurements and surface conditions such as wave height are also routinely available from surface buoys. A few moorings include a variety of optical sensors, e.g., fluorometers and spectral backscattering sensors. Ocean currents, shear, and parameters that describe mixing are sensed over long periods by autonomous acoustic Doppler current profilers (ADCP). Bioacoustical devices, used to detect and study marine zooplankton and micronekton communities, are relatively new additions to the growing suite of ''moorable'' sensors. In each of 2002, 2003, and 2004, several months of continuous bioacoustical measurements of zooplankton were made concurrently with measurements of several physical parameters and meteorological variables at a single mooring site located south of Seward, Alaska in the coastal Gulf of Alaska. Volume scattering strength spectra were collected at 20-min intervals with a TAPS-8 sensor at 20-m depth. The data allow scientists to examine complex inter- and intra-annual variations in the estimated biomass-size spectra for organism shapes typical of small zooplankton (e.g., copepods) and elongate micronekton (e.g., krill) in a unique new way. [Work sponsored by NOAA's Coastal Ocean Program.]

2:45

2pAO6. Lessons learned from multifrequency acoustic studies of zooplankton and micronekton in the western Antarctic Peninsula and the Gulf of Maine. Andone C. Lavery, Gareth L. Lawson, and Peter H. Wiebe (Woods Hole Oceanograph. Inst., M.S. 11, Woods Hole, MA 02543)

A series of acoustic surveys of zooplankton and micronekton have been performed in the Gulf of Maine (GOM), off the northeast United States, and along the western Antarctic Peninsula (WAP). Similar techniques were used to survey these regions, including multifrequency acoustic backscatter (43, 120, 200, 420, 1000 kHz), MOCNESS, CTD, VPR, and in some instances physical microstructure measurements. The GOM is characterized by heterogeneous zooplankton communities in which biomass is dominated by abundant millimeter sized copepods, but the scattering is frequently dominated by a smaller number of strong scatterers, such as shelled pteropods and gas-bearing siphonophores. Heterogeneous zooplankton communities are also observed in the WAP, but patches of comparatively large (40 mm) Antarctic krill are present and often dominate the scattering. In both regions, striking patterns are evident in the backscatter that can be related to the biological community structure and physical processes. Differences in community structure, however, strongly affect the quantitative inferences that can be made based on the acoustic data. Combining direct biological

and environmental information with recently developed scattering models has allowed dominant scatterers to be identified and inferences to be made regarding the physical factors influencing backscatter variability, though only under limited conditions. Highlights from these studies and lessons learned regarding our ability to interpret multifrequency acoustics are presented.

3:05–3:25   Break

3:25

2pAO7. A new tool for real-time acoustic species identification of delphinid whistles.   Julie N. Oswald   (Scripps Inst. of Oceanogr., Univ. of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, joswald@ucsd.edu), Shannon Rankin, Jay Barlow (Natl. Marine Fisheries Service, NOAA, La Jolla, CA 92037) and Marc O. Lammers   (Hawaii Inst. of Marine Biol., Kailua, HI 96734)

Acoustic species identification studies generally focus on postprocessing of field recordings using multivariate statistics such as classification tree analysis (CART). When CART was used to classify whistles of nine delphinid species recorded in the eastern tropical Pacific ocean ($n = 908$), 51% were correctly classified to species (versus 11% expected by chance). These results led to the development of a new automated system to allow real-time acoustic species identification in the field. This MATLAB-based tool automatically extracts variables from whistles manually selected from a real-time scrolling spectrograph (ISHMAEL software) and runs classification statistics (CART and discriminant function analysis). Output includes predicted species identification with a certainty score and alternate identifications with certainty scores. The original CART results were based on classification of one whistle at a time. With the new system, identification decisions are based on multiple whistles, resulting in higher classification success. This new tool allows species distribution data to be collected using a towed hydrophone array when visual efforts are compromised (e.g., darkness, high sea states). It may also prove valuable for studies utilizing bottom-mounted recorders, where visual species identification is not possible. Additionally, through automation, this tool reduces the time required for analysis of voluminous acoustic data.

3:45

2pAO8. Foraging behavior of fish-eating sperm whales in the Gulf of Alaska in the presence and absence of fishing vessels.   Aaron Thode   (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093), Jan Straley, Kendall Folkert   (Univ. of Alaska Southeast, Sitka, AK 99835), Victoria O'Connell   (Alaska State Fish and Game, Sitka, AK 99835), and Christopher Tiemann   (Univ. of Texas, Austin, TX 78713)

Historical whaling records indicate that sperm whales off southeast Alaska incorporate fish into their diets, particularly black cod (*Anoploploma fimbria*). Since 1995 this fact has become relevant to fisheries' concerns in the form of increased depredation encounters between longline fishermen and over 40 sperm whales. Since 2002 the SE Alaska Sperm Whale Avoidance Project (SEASWAP) has been studying this phenomenon using fishermen reports, photo-ID, biopsy, and (since 2004) passive acoustics using both towed arrays and autonomous recorders placed on longline deployments. By using acoustic multipath the range and depths of foraging whales can be determined. Findings to date indicate that, under natural conditions, sperm whales are foraging at mid-depth in the water column (e.g., 250 m in 500-m-deep water), and that their dive cycle durations are similar to those reported in other oceans. This information is being compared with depth measurements of black cod at various stages of its life cycle. There is increasing evidence that distinctive acoustic cues made by longline vessels lead to changes in diving and acoustic behavior by the animals, when the animals are less than 10 nautical miles away. [Work supported by the North Pacific Research Board.]

Contributed Papers

4:05

2pAO9. Model-based passive acoustic tracking of sperm whale foraging behavior in the Gulf of Alaska.   Christopher Tiemann   (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713, tiemann@arlut.utexas.edu), Aaron Thode   (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), Jan Straley, Kendall Folkert   (Univ. of Alaska Southeast, Sitka, AK 99835), and Victoria O'Connell   (Alaska State Fish and Game, Sitka, AK 99835)

In 2004, the Southeast Alaska Sperm Whale Avoidance Project (SEASWAP) introduced the use of passive acoustics to help monitor the behavior of sperm whales depredating longline fishing operations. Acoustic data from autonomous recorders mounted on longlines provide the opportunity to demonstrate a tracking algorithm based on acoustic propagation modeling while providing insight into whales' foraging behavior. With knowledge of azimuthally dependent bathymetry, a 3D track of whale motion can be obtained using data from just one hydrophone by exploiting multipath arrival information from recorded sperm whale clicks. The evolution of multipath arrival patterns is matched to range-, depth-, and azimuth-dependent modeled arrival patterns to generate an estimate of whale motion. This technique does not require acoustic ray identification (i.e., direct path, surface reflected, etc.) while still utilizing individual ray arrival information, and it can also account for all waveguide propagation physics such as interaction with range-dependent bathymetry and ray refraction.

4:20

2pAO10. A scattering analysis of echoes due to biosonar signals emitted by foraging beaked whales.   Benjamin A. Jones, Timothy K. Stanton, Andone C. Lavery, Mark P. Johnson, Peter T. Madsen, and Peter L. Tyack   (Woods Hole Oceanogr. Inst., M.S. 16, Woods Hole, MA 02543-1050)

Blainville's beaked whales (*Mesoplodon densirostris*) hunt their prey by echolocation at depths of more than 500 meters. These whales use a FM upswept, ultrasonic click, of greater than an octave bandwidth to search for, localize, and close on individual prey which generally consist of mesopelagic fishes and squid. It is well known that acoustic scattering from organisms of varying morphology (e.g., swimbladder-bearing or fluidlike) is strongly frequency dependent. However, it is unknown if the broadband nature of the whales' outgoing signal, and the frequency dependence of the echoes, is a key component in the classification and selection of their prey. Non-invasive, acoustic ''Dtags,'' which sample stereo acoustic data at a rate which satisfies the high-frequency Nyquist criterion for the animal's transmit signal, were affixed to beaked whales. The Dtags successfully recorded transmitted signals and associated echoes. Structure was observed in the frequency content of echoes from isolated targets in the water column which may be used for classification by the whales. An analysis of the echoes identified as possibly due to prey has demonstrated that multiple classes of frequency responses are present. These results will be compared with the frequency responses of possible prey types.

2pAO11. Where should we eat?: Predator and prey spatial relationships in the Antarctic krill ecosystem. Joseph D. Warren (Marine Sci. Resource Ctr., Stony Brook Univ., 239 Montauk Hwy., Southampton, NY 11968), Jarrod Santora (College of Staten Island, CUNY, Staten Island, NY 10214), and David Demer (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037)

Acoustic surveys of the distribution and abundance of Antarctic krill were conducted by two different vessels in February 2005 in the near-shore waters of Livingston Island, Antarctica. Multiple frequency acoustic backscatter, video observations, and net tow data were used to identify the acoustic scatterers and to estimate the abundance of krill. Hydrographic, meteorologic, and net tow data were used to examine physical forcing functions affecting the spatial distribution of krill. In addition, the behavior and distribution of several krill predators (penguins, petrels, and whales) were also measured during the survey to explore biological factors affecting the krill ecosystem. Both horizontal and vertical patterns of habitat utilization by prey (krill) and these predators were examined. The survey was interrupted midway through by a large storm event that limited ship-based observations for a day, but provided a picture of the ecosystem before and after the disturbance. [Work supported by the NSF Office of Polar Programs and NOAA.]

2pAO12. High-frequency acoustic tracking of odontocetes at the Southern California Offshore Range (SCORE). Paul Hursky, Ahmad T. Abawi, Michael B. Porter (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), John A. Hildebrand, Sean M. Wiggins, Melissa Soldevilla, and Allan W. Sauter (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

The SCORE site, located near San Clemente Island, is known to be abundant in marine mammals, and is the site of frequent naval operations, making it an ideal site to study what impact man-made activities may have upon these creatures. In August 2004, an experiment was performed to collect both visual and acoustic observations of odontocetes at the SCORE site. Odontocete vocalizations consist of clicks and whistles. The clicks are very wideband (often 100 kHz), so times of arrival can be measured with extremely high resolution, which makes clicks attractive waveforms for multipath ranging and depth estimation. Although the whistles are narrowband, their fundamental frequency often ranges over tens of kilohertz, so these too can be processed as if they were wideband. We will present results of processing excerpts from several weeks of data recorded on a high-frequency horizontal line array deployed from the FLIP research vessel. Bearing tracks of individuals and groups from a number of different species will be correlated and compared with visual observations. In addition, we will discuss the possibilities for range and depth estimates derived from the multipath arrivals observed in the clicks and whistles.

2pAO13. Factors which influence acoustic surveys of marine mammals. Tracey L. Rogers, Michaela B. Ciaglia (Australian Marine Mammal Res. Ctr., Zoological Parks Board of NSW/Faculty of Veterinary Sci., Univ. of Sydney, Mosman, Australia), and Douglas H. Cato (Defence Sci. & Technol. Organisation, Pyrmont, Australia)

Traditionally, many marine mammal populations have been estimated by visual surveys. These count the animals that are available—either seals hauled-out on the ice or whales at the water's surface. Corrections are then made to include the animals that were not seen either because they were in (seals) or under (whales) the water. However when the majority of the animals in a population are not available to a visual survey this approach may be less effective. So we investigated whether acoustic surveys offered promise for estimating the distribution and abundance of Antarctic pack-ice seals. Four acoustic surveys were conducted (October 1996, 1997; December 1997, 1999) between longitudes 600E and 1500E. Surveys were bounded to the south by fast-ice, shelf-ice or the Antarctic continent and to the north by the edge of the pack-ice. No crabeater seals were heard. Leopard and Ross seals were highly vociferous in December coinciding with their breeding season. To predict the area surveyed we modeled transmission loss and measurements of received background levels. To identify the number of seals calling we modeled calling behavior. A preliminary estimate of 0.13 male leopard seals/km^2 was calculated which is in the high-density range described from the literature.

2pAO14. Environmental noise levels affect the activity budget of the Florida manatee. Jennifer L. Miksis-Olds, Percy L. Donaghay (Grad. School of Oceanogr., Univ. of Rhode Island, Narragansett, RI 02882, jmiksis@gso.uri.edu), James H. Miller (Univ. of Rhode Island, Narragansett, RI 02882), and Peter L. Tyack (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Manatees inhabit coastal bays, lagoons, and estuaries because they are dependent on the aquatic vegetation that grows in shallow waters. Food requirements force manatees to occupy the same areas in which human activities are the greatest. Noise produced from human activities has the potential to affect these animals by eliciting responses ranging from mild behavioral changes to extreme aversion. This study quantifies the behavioral responses of manatees to both changing levels of ambient noise and transient noise sources. Results indicate that elevated environmental noise levels do affect the overall activity budget of this species. The proportion of time manatees spend feeding, milling, and traveling in critical habitats changed as a function of noise level. More time was spent in the directed, goal-oriented behaviors of feeding and traveling, while less time was spent milling when noise levels were highest. The animals also responded to the transient noise of approaching vessels with changes in behavioral state and movements out of the geographical area. This suggests that manatees detect and respond to changes in environmental noise levels. Whether these changes legally constitute harassment and produce biologically significant effects need to be addressed with hypothesis-driven experiments and long-term monitoring. [For Animal Bioacoustics Best Student Paper Award.]

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: Transducers and Imaging

Robert J. McGough, Chair

Michigan State Univ., Electrical and Computer Engineering, 2120 Engineering Bldg., East Lansing, MI 48824

Contributed Papers

1:30

2pBB1. Comparison of high intensity focused ultrasound (HIFU) exposures using empirical and backscatter attenuation estimation methods. John Civale, Gail ter Haar, Ian Rivens, and Jeff Bamber (Phys. Dept., Inst. of Cancer Res., Royal Marsden Hospital, Sutton, Surrey, SM2 5PT, UK)

Currently, the intensity to be used in our clinical HIFU treatments is calculated from the acoustic path lengths in different tissues measured on diagnostic ultrasound images of the patient in the treatment position, and published values of ultrasound attenuation coefficients. This yields an approximate value for the acoustic power at the transducer required to give a stipulated focal intensity *in situ*. Estimation methods for the actual acoustic attenuation have been investigated in large parts of the tissue path overlying the target volume from the backscattered ultrasound signal for each patient (backscatter attenuation estimation: BAE). Several methods have been investigated. The backscattered echo information acquired from an Acuson scanner has been used to compute the diffraction-corrected attenuation coefficient at each frequency using two methods: a substitution method and an inverse diffraction filtering process. A homogeneous sponge phantom was used to validate the techniques. The use of BAE to determine the correct HIFU exposure parameters for lesioning has been tested in *ex vivo* liver. HIFU lesions created with a 1.7-MHz therapy transducer have been studied using a semiautomated image processing technique. The reproducibility of lesion size for given *in situ* intensities determined using BAE and empirical techniques has been compared.

1:45

2pBB2. Ultrasound imaging for high intensity focused ultrasound therapy. Shahram Vaezy, Vesna Zderic, and Lawrence Crum (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98195)

We have investigated the use of ultrasound-based imaging to develop a non-invasive, safe and real-time method of guidance and monitoring for HIFU therapy. HIFU application was synchronized with the imaging frame rate to allow interference-free visualization of the region of interest. The regions treated with HIFU at intensities above the cavitation threshold appear immediately (within 30–60 ms) as hyperechoic structures in ultrasound images, confirmed to be due to microbubble activity at the focus. This HIFU dose is shown to produce minimal tissue damage, in the form of capillary rupture in an area of approximately 0.5 mm around the focus. The appearance of hyperechoic regions is potentially useful for pretreatment targeting to ensure that the focus is located at the correct spot. The hyperechoic structures persist for 1–2 min, while the microbubbles still remain at the focus, before being absorbed. Microbubbles at the focus also provide a shield for HIFU, to prevent thermal damage of the healthy tissues located post-focally. Challenges of ultrasound-image-guided HIFU are in monitoring the tissue thermal behavior in and around the focus, and follow-up monitoring of tissues treated with HIFU. Ultrasound imaging is developing into a promising method for guidance of HIFU therapy. [Work funded by DoD, NIH.]

2:00

2pBB3. Ophthalmic imaging with a 40-MHz annular array. Jeffrey A. Ketterling, Sarayu Ramachandran (Riverside Res. Inst., 156 William St., New York, NY 10038, ketterling@rrinyc.org), Ronald H. Silverman (Weill Medical College of Cornell Univ., New York, NY 10021), and Orlando Aristizábal (Skirball Inst. of Biomolecular Med. and New York Univ. School of Medicine, New York, NY 10016)

High-frequency annular arrays provide a means for improving the depth of field versus single-element transducers. An annular array with five equal-area rings has been designed and an experimental system has been assembled to test it. The transducer is fabricated from a 9-μm polyvinylidene fluoride (PVDF) membrane bonded to a copper-clad polyimide film. The transducer has a total aperture of 6 mm, a geometric focus of 12 mm, and a nominal center frequency of 40 MHz with a 6-dB fractional bandwidth of 33%. The device is pressed into a spherically curved geometry and then backfilled with epoxy. Radio-frequency data were acquired for all 25 transmit/receive ring pairs, and then postprocessed with a synthetic focusing algorithm to improve the image depth of field (DOF). The resulting processed data improved DOF by a factor of 6 compared to a single-element transducer of the same total aperture and geometric focus. To test the system, rf data were acquired from excised bovine eyes, human cadaver eyes, and anesthetized rabbits. The tests revealed enhanced resolution away from the geometric focus and a more uniform image brightness than can be achieved with high-frequency, single-element devices. [This work was supported by NIH Grant EY014371.]

2:15

2pBB4. Multiplanar angular spectrum approach for fast simulations of ultrasound therapy arrays. Xiaozheng Zeng and Robert McGough (Michigan State Univ., 2120 Engineering. Bldg., East Lansing, MI 48824)

Pressure field modeling with ultrasound phased arrays designed for thermal therapy is extremely time-consuming due to the large computational grids required and because each array contains a large number of elements. The angular spectrum approach (ASA) computes wave propagation in parallel planes with a two-dimensional fast Fourier transform in considerably less time than point-by-point direct integral methods. The conventional ASA, which performs well with planar and small focused radiators, encounters numerical problems in simulations of large phased arrays. These errors are caused in part by the narrowing of the beam in the focal region, which produces a significant increase in the spatial frequency content. The numerical errors are reduced with the inclusion of additional source planes within the computational domain. These are distributed such that the distance between adjacent source planes is small near the focal zone, and the spacing between source planes grows progressively larger as the distance from the focus increases. This approach, which compensates for the error produced by spatial aliasing, reduces the computation time required for simulations of large ultrasound phased arrays while maintaining the necessary accuracy for thermal therapy simulations. [Supported in part by NIH Grant R01-CA093669.]

2p TUE. PM

1911   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1911**

2:30

2pBB5. Polyvinylidene fluoride (PVDF) vibration sensor for stethoscope and contact microphones. Minoru Toda and Mitchell Thompson (460 E. Swedesford Rd., Ste. 3005, Wayne, PA 19087)

This paper describes a new type of contact vibration sensor made by bonding piezoelectric PVDF film to a curved frame structure. The concave surface of the film is bonded to a rubber piece having a front contact face. Vibration is transmitted from this face through the rubber to the surface of the PVDF film. Pressure normal to the surface of the film is converted to circumferential strain, and an electric field is induced by the piezoelectric effect. The frequency response of the device was measured using an accelerometer mounted between the rubber face and a rigid vibration exciter plate. Sensitivity (voltage per unit displacement) was deduced from the device output and measured acceleration. The sensitivity was flat from 16 Hz to 3 kHz, peaking at 6 kHz due to a structural resonance. Calculations predicting performance against human tissue (stethoscope or contact microphone) show results similar to data measured against the metal vibrator. This implies that an accelerometer can be used for calibrating a stethoscope or contact microphone. The observed arterial pulse waveform showed more low-frequency content than a conventional electronic stethoscope.

TUESDAY AFTERNOON, 18 OCTOBER 2005 CONRAD C, 1:30 TO 2:30 P.M.

Session 2pEDa

Education in Acoustics: Acoustics Education Prize Lecture

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Chair's Introduction—1:30

Invited Paper

1:35

2pEDa1. Speech neglect: A strange educational blind spot. Katherine Safford Harris (Haskins Labs., 300 George St., New Haven, CT 06511 and Grad. School, City Univ. of New York, New York, NY)

Speaking is universally acknowledged as an important human talent, yet as a topic of educated common knowledge, it is peculiarly neglected. Partly, this is a consequence of the relatively recent growth of research on speech perception, production, and development, but also a function of the way that information is sliced up by undergraduate colleges. Although the basic acoustic mechanism of vowel production was known to Helmholtz, the ability to view speech production as a physiological event is evolving even now with such techniques as fMRI. Intensive research on speech perception emerged only in the early 1930s as Fletcher and the engineers at Bell Telephone Laboratories developed the transmission of speech over telephone lines. The study of speech development was revolutionized by the papers of Eimas and his colleagues on speech perception in infants in the 1970s. Dissemination of knowledge in these fields is the responsibility of no single academic discipline. It forms a center for two departments, Linguistics, and Speech and Hearing, but in the former, there is a heavy emphasis on other aspects of language than speech and, in the latter, a focus on clinical practice. For psychologists, it is a rather minor component of a very diverse assembly of topics. I will focus on these three fields in proposing possible remedies.

TUESDAY AFTERNOON, 18 OCTOBER 2005 CONRAD C, 3:00 TO 4:30 P.M.

Session 2pEDb

Education in Acoustics: Take "Fives"

Ian M. Lindevald, Chair
Truman State Univ., Div. of Science/Physics, Kirksville, MO 63501

Your favorite Acoustics teaching devices. Bring a short demonstration, teaching device, or video, to share with your colleagues. No abstract required. A brief descriptive handout is encouraged. Sign up at the door for a 5-min slot. If you bring more than one demo sign up for two non-consecutive slots.

Session 2pMU

Musical Acoustics: Patents in Musical Acoustics

George Brock-Nannestad, Cochair
Patent Tactics, Resedavej 40, DK-2820 Gentofte, Denmark

George Augspurger, Cochair
Perception Inc., P.O. Box 39536, Los Angeles, CA 90039

Chair's Introduction—1:00

Invited Papers

1:05

2pMU1. Two centuries of French patents as documentation of musical instrument construction. Haury Jean (14 TER, rue de Mouzaia, 75019 Paris, France, jhaury@noos.fr)

The French Patent Office I.N.P.I. has preserved the originals of ca. 12 000 French patents filed between 1791 and present days that are concerned with music-related inventions. As an I.N.P.I. pilot project, these were identified, collected, and classified by the present author, and the actual database named ''Musique & Brevets'' is going to be expanded with English, American, and German material, bringing currently a knowledge base up to 1900. It is expected to be made available on an I.N.P.I. website. This is an unequaled initiative that covers all branches of musical instrument manufacture, mechanical musical instruments, early recording and reproducing of music, but also educational material and methods for printing music. There already exists a number of websites presenting inventions on musical instruments, but these are restricted to one particular instrument and its related patents. ''Musique & Brevets'' intends to be exhaustive and make links between patents filed in different countries at the same time. The paper will present the content of the database, the access to texts and drawings of the patents via specific links, and their importance for the study of history and construction of musical instruments.

1:30

2pMU2. The good, the bad, and the bogus in acoustical patents. George L. Augspurger (Percept., Inc., Box 39536, Los Angeles, CA 90039)

Patent reviews published in the Journal of the Acoustical Society of America are intended to provide useful information from the standpoint of an informed layman. From that perspective at least 50% of acoustical patents are superfluous and a good many are outright fraudulent. Why do we see so many patents that arguably should never have been issued? Part of the answer lies in common misconceptions about the purpose of a U.S. patent and the protection it provides. Additional confusion arises from the common strategy of writing patent text to obfuscate rather than enlighten. Examples of junk patents, vanity patents, and bogus patents will be presented to illustrate current trends in the patent process.

1:55

2pMU3. Story of the patent intitulated (in translation from the French), "Feature of wind instrument permitting the generation of special effects in real time." Rene Causse (I.R.C.A.M. (CNRS UMR991) 1 place Igor Stravinsky, F-75004 Paris, France)

This presentation will focus on the history of a patent taken out in 1996 by my colleagues F. Laloe, A. Ducoureau, F. Terrier, and myself (France Tlcom patent). This patent is related to a system that allows the wind instrument's pitch to be controlled by means of a mechanical device comprising a foot pedal for the musician and a cable connected to the embouchure of the instrument. There is no predefined semitone or quarter-tone position. The system extends the current musical range and performance capabilities of the instrument chosen for a realization, a clarinet, by inserting in the instrument an additional length. It is possible to change not only the pitch of tones but also to soften their timbre, producing a tone similar to that of early instruments. This type of special effect is well suited to contemporary, jazz, and popular music. The device is adapted to any type of clarinet and the principle can conceivably be extended to other wind instruments. In this presentation, the device and its derivatives shall be described, and its musical applications illustrated by sound examples. A critical approach to the different levels of reading of the patent will be presented.

2p TUE. PM

2:30

2pMU4. A recent U.S. patent process for a musical instrument. James I. Baecker   (Harmos Music, Ltd., 1950 Crestview Circle, Excelsior, MN 55331, james@harmosmusic.com)

The ins and outs of going from a novel musical instrument concept to issuance of a U.S. patent. The technical work performed included the development of a musical instrument based on space-frame body construction and a definition of the instrument body's resonance characteristics. The result required a description of the invention and communication with the patent attorney and conveying a correct perception of the invention to the U.S. Patent Office. This presentation describes several technical, practical, legal, and commercial issues encountered during the patent process by the inventors and their business entity. On 7 September 2004, U.S. Patent No. 6,787,688 for a musical instrument was issued and assigned by the inventors to Harmos Music, Ltd.

2:55

2pMU5. Access to patents as sources to musical acoustics inventions. George Brock-Nannestad   (Patent Tactics, Resedavej 40, DK-2820 Gentofte, Denmark)

Patents are important sources for the development of any technology. The paper addresses modern methods of access to patent publications relating to musical acoustics, in particular the constructions of instruments and components for instruments, methods for tuning, methods for teaching, and measuring equipment. The patent publications available are, among others, from the U.S., England, France, Germany, Japan, Russia, and the date range is from ca. 1880 to the present day. The two main searchable websites use different classification systems in their approach, and by suitable combination of the information it is possible to target the search efficiently. The paper will demonstrate the recent transfer of inventions relating to physical instruments to electronic simulations, and the fact that most recent inventions were made by independent inventors. A specific example is given by discussing the proposals for improved pipe organ and violin constructions invented in Denmark in the 1930s by Jarnak based on patented improvements for telephone reproducers.

TUESDAY AFTERNOON, 18 OCTOBER 2005   SALON E, 1:00 TO 4:20 P.M.

Session 2pNCa

NOISE-CON and Noise: Tire/Pavement Noise and Quiet Pavement Applications

Judith L. Rochat, Cochair
U.S. Dept. of Transportation, Volpe Ctr., 55 Broadway, Cambridge, MA 02142

Paul R. Donavan, Cochair
Illingworth & Rodkin Inc., 505 Petaluma Blvd., South, Petaluma, CA 94952-5128

1:00

2pNCa1. Roadmap to quieter highways. Christopher Corbisier (Federal Hwy. Administration, 400 7th St., SW, HEPN-20, Washington, DC 20590)

Research in Europe, as documented by an FHWA/AASHTO European Scan Tour held in May 2004, and recent activity in Arizona and California, has fostered much interest in "quiet pavements." On September 14–16, 2004, an FHWA sponsored Roadmap to Quieter Highways workshop was held at Purdue University. Participants were from the disciplines of pavement, safety, and noise from FHWA, State departments of transportation, industry (paving associations, general contractors, tire, and vehicle manufacturers), and academia. After several breakout sessions in the areas of policy, construction, maintenance, analysis (measurement and prediction), research, and design, the group had identified the knowledge gaps and developed a plan to fill those gaps. Several activities have been implemented based on the Roadmap to Quieter Highways. An Expert Task Group was formed to provide a draft provisional standard for the measurement methodologies, e.g., source, wayside, pavement absorption. A Tire/Pavement 101 workshop is being developed to educate pavement practitioners in noise concepts and noise practitioners in pavement concepts. A Tire/Pavement Noise clearinghouse is being developed as a one-stop location for all current tire/pavement noise or quiet pavement activities. Several research studies have been started and a second workshop will be held in 2006 to assess progress of the Roadmap.

1:20

2pNCa2. Noise absorption characteristics and modeling for quiet pavements. Hasson Tavossi   (Dept. of Physical & Environ. Sci., Mesa State College, 1100 North Ave., Grand Junction, CO 81501)

In this study, various techniques for noise attenuation by quiet pavements and its modeling are investigated in order to reduce traffic noise in urban environments. Noise absorption in selective frequency ranges by pavements can reduce its reach, the degree of noise annoyance, and general background noise level in the environment. Different methods of noise energy absorption are considered, including internal attenuation by multiple scatting, and selective frequency absorption by cavity resonance inside the pavement structure. Noise energy dissipation in the undesirable frequency ranges, by multiple scattering and absorption in the pavement and by sound-absorbing barriers, are also considered. Dispersion and scattering of high-frequency components of the noise result in the long-range reach of the low-frequency components of the traffic-generated noise. Selective attenuation by quiet pavements of the low-frequency components of the noise reduces its rage and improves significantly the overall noise pollution in the urban environments. A general background of the modeling and the experimental findings on sound attenuation by absorption, cavity resonance, and multiple scattering techniques, at low and high frequencies, will be presented.

1:40

2pNCa3. Studies of the effect of aging of "quiet" pavements on tire/pavement noise. James A. Reyff and Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952)

One of the issues with using quieter pavements to abate traffic noise is their continued acoustic performance over the life cycle of the surface. Aging effects can be assessed in two manners: (1) long term monitoring of the noise performance of an individual section of roadway; (2) measurement of multiple sections of pavement of the same construction, but different ages. Long term monitoring of Interstate 80 near Davis (I-80 Davis) began in 1998, just prior to the placement of a dense graded leveling course and open graded asphalt overlay. The pavement surface is now approaching 7 years old and continues to show a traffic noise reduction of about 5 dBA over the existing condition. As support of the Arizona Quiet Pavement Program (QPPP), similarly constructed sections of asphalt rubber friction course (ARFC) on Arizona's interstate highways were measured for tire/pavement performance using the close proximity (CPX) method and the on-board sound intensity method. The construction dates for the pavements ranged from 1988 to 1999. The total range in noise level was 7 dB with some indication of degrading performance with age.

2:00

2pNCa4. Overview of the Arizona Quiet Pavement Program. Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952) and Larry Scofield (American Concrete Pavement Assoc., Mesa, AZ 85210)

The Arizona Quiet Pavement Pilot Program (QP3) was initially implemented to reduce highway related traffic noise by overlaying most of the Phoenix metropolitan area Portland cement concrete pavement with a one inch thick asphalt rubber friction coarse. With FHWA support, this program represents the first time that pavement surface type has been allowed as a noise mitigation strategy on federally funded projects. As a condition of using pavement type as a noise mitigation strategy, ADOT developed a ten-year, $3.8 million research program to evaluate the noise reduction performance over time. Historically, pavement surface type was not considered a permanent solution. As a result, the research program was designed to specifically address this issue. Noise performance is being evaluated through three means: (1) conventional roadside testing within the roadway corridor (e.g., far field measurements within the right-of-way); (2) the use of near field measurements, both close proximity (CPX) and sound intensity (SI); and (3) far field measurements obtained beyond the noise barriers within the surrounding neighborhoods. This paper provides an overview of the program development, presents the research conducted to support the decision to overlay the urban freeway, and the status of current research.

2:20

2pNCa5. Reduction of traffic and tire/pavement noise: 1st year results of the Arizona Quiet Pavement Program—Site III. James A. Reyff, Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952)

The Arizona Quiet Pavement Pilot Program overlaid major freeway segments in the Phoenix area with an Asphalt Rubber Friction Course (ARFC). The overlay was placed on various Portland Cement Concrete Pavement (PCCP) textures. Traffic noise reductions were evaluated by performing wayside traffic noise measurements and tire/pavement source level measurements. First year results for three different study sites are presented in this paper. Depending on the texture of the initial PCCP and microphone locations, reductions of up to 12 dBA in wayside traffic noise levels were measured. Similar reductions of tire/pavement source levels were measured. Results of the two methods are compared. Traffic conditions monitored during the measurements were modeled using the Federal Highway Administration's Traffic Noise Model (TNM 2.5) to compare modeled levels to those measured for PCCP and AFRC overlay conditions. The model under predicted levels for PCCP conditions and over predicted levels for AFRC conditions. The magnitude of under or over prediction varied with distance. The effect of propagation was examined and was

aided by simultaneous measurements of wind conditions made by Arizona State University. TNM 2.5 was used to identify sound wall heights that were equivalent to the traffic noise reductions provided by the AFRC overlay.

2:40–3:00   Break

3:00

2pNCa6. Quieting of Portland cement concrete highway surfaces with texture modifications. Paul Donavan (Illingworth & Rodkin, Inc., 505 Petaluma Blvd. South, Petaluma, CA 94952)

In recent years, various types of asphalt surfaces have become identified as "quiet pavements" due to their ability to reduce tire/pavement noise and, ultimately, traffic noise. Often lost in this perception is the fact that substantial reductions in tire/pavement noise can also be made by texture modifications to existing Portland Cement Concrete (PCC) or by novel constructions. PCC surfaces have been found to span a range of as much as 16 dB. As a result, there is the potential to achieve large noise reductions depending on the existing and final surfaces. In California, grinding of bridge decks and elevated structures has been found to reduce tire/pavement source levels 3 to 10 dB with comparable reductions in wayside measurements. In Arizona, grinding of PCC has reduced source levels up to 9 dB relative to some transversely tined surfaces. Measurements conducted in Europe using the same measurement methodology indicated a range of 11 dB including more novel porous PCC surfaces. In this paper, measurement results and case histories are reviewed for situations where PCC modifications were successful and unsuccessful in producing quieter pavement.

3:20

2pNCa7. Review of the development of concrete pavement surface types and their acoustic performance with time. Larry A Scofield (American Concrete Pavement Assoc., 807 W. Keating Ave., Mesa, AZ 85210)

Roadways have been constructed out of concrete pavements in the United States since the late 1800s. Numerous surface textures have been produced including burlap drag, astroturf, uniformly and random transverse tined, longitudinally tined, and both profile and whisper diamond grinding processes. Each surface evolved for specific reasons during the historical development of Portland Cement Concrete Pavements (PCCP).This paper reviews the development of these various surface texture types and the reasons for their evolution. In addition, results from both far field and near field acoustic testing are presented to evaluate the "acoustic" performance of these surfaces over time.For five surface types, 1/24 octave analysis were conducted on CPX data obtained with a single tire. The spectrum of each of these surface types is presented for comparison. For three of the surfaces, speed gradient testing, ranging between 25–75 MPH was conducted. 1/24 octave analysis of each of these runs was conducted so that that any speed induced spectrum shifts could be observed.The results to date indicate that the adverse tonal properties of some PCCP surfaces can be eliminated through diamond grinding and prevented by not constructing transverse tined PCCP.

3:40

2pNCa8. Noise characteristics of hot mix asphalt and Portland cement concrete pavements in United States. Douglas I. Hanson (AMEC Earth and Environ., 1405 W. Auto Dr., Tempe, AZ 85284)

In today's society, traffic noise is a serious problem that generally is considered an environmental pollution because it lowers the standard of living. Research in Europe and in the United States has indicated that it is possible to build pavement surfaces that will reduce the level of noise generated on roadways. In January of 2002 the National Center for Asphalt Technology initiated a research study with the objective to develop safe, quiet and durable asphalt pavement surfaces. As a part of that study over 300 pavement surfaces [both Portland Cement Concrete (PCCP) and Hot Mix Asphalt (HMA)] throughout the United States have been tested

using a close-proximity noise trailer. The study has shown that in general PCCP surfaces have a higher noise level than HMA surfaces. But, it has also shown that by properly choosing the surface texture of the PCCP surface significant reductions in the noise level of a PCCP surface can be achieved. The study has shown that it is possible to construct low-noise HMA mixes and that in general the smaller the nominal maximum size for those mixes (regards whether they are dense graded, SMA or OFGC mixes) the lower the noise level.

4:00

2pNCa9. Variability of pavement noise benefit by vehicle type. Judith L. Rochat and David R. Read (U.S. Dept. of Transportation, Volpe Natl. Transportation Systems Ctr., 55 Broadway, Cambridge, MA 02142)

The Volpe Center Acoustics Facility, in support of the California Department of Transportation (Caltrans), is participating in a long-term study to assess several types of pavement for the purpose of noise abatement. On a four-mile stretch of a two-lane highway in Southern California, several asphalt pavement overlays are being examined. Acoustical, meteorological, and traffic data are collected in each pavement overlay section, where microphones are deployed at multiple distances and heights. Single vehicle pass-by events are recorded primarily for three vehicle types: automobiles, medium trucks, and heavy trucks. Data are analyzed to determine the noise benefit of each pavement as compared to the reference dense-graded asphaltic concrete (DGAC); this includes a modified Statistical Pass-By Index as well as average Lmax values for each vehicle type. In addition, 1/3-octave band data are examined. Automobiles and heavy trucks are the focus of this paper, where benefits due to pavement will be presented for three pavement types: open-graded asphaltic concrete (OGAC) of 75 mm thickness, open-graded asphaltic concrete (OGAC) of 30 mm thickness, and rubberized asphaltic concrete, Type O (open) (RAC) of 30 mm thickness. Average Lmax values and spectral data show that noise benefits due to pavement can vary by vehicle type.

TUESDAY AFTERNOON, 18 OCTOBER 2005

SALON B, 1:00 TO 2:40 P.M.

Session 2pNCb

NOISE-CON, Noise and Signal Processing in Acoustics: Array Methods for Noise Source Visualization II

J. Stuart Bolton, Cochair

Purdue Univ., Ray W. Herrick Labs., 140 South Intramural Dr., West Lafayette, IN 47907-2031

Courtney B. Burroughs, Cochair

Noise Control Engineering, 1241 Smithfield St., State College, PA 16801

1:00

2pNCb1. A means to see noise sources. Yang-Hann Kim and Joung-Woo Choi (Ctr. for Noise and Vib. Control (NOVIC), Dept. of Mech. Eng., Korea Adv. Inst. of Sci. and Technol. (KAIST), Sci. Town, Daejon, 305-701, Korea)

There are many candidates to visualize noise/sound sources. For example, acoustic holography is one of them. However, to get acoustic holography, we need many microphones and associated data acquisition and processing unit. Especially, when noise is impulsive, the holography system requires two-dimensional microphone array. Depending on the type of noise sources, we have to select the best means that can provide us what we want to have, which can be noise sources' locations or entire sound distribution in space. This paper addresses various methods that can predict where noise sources are, or how noise is distributed in space. This includes beam forming methods to determine steady or impulsive noise sources, conventional acoustic holography, time domain acoustic holography and moving frame method that can show moving noise sources. We compare these methods and show a guideline that can lead us to select the best means to see noise sources. Various industrial examples are introduced. Time domain acoustic holography of plasma display panel that shows very interesting vibration pattern, beam forming method that depicts short and small noise generated by thermal expansion of a TV set, and moving frame method that shows tire noise are some of the examples.

1:20

2pNCb2. An advanced noise source identification technique using the inverse boundary-element method. Ferdy Martinus (Mech. Eng. Dept., Univ. of Kentucky, Lexington, KY 40506)

The application of the inverse BEM method as an advanced, non-contact noise source identification technique will be discussed. In the first and most conventional experiment, the inverse BEM is used to reconstruct the vibration pattern on the surface of a motorcycle engine oil pan cover. In the second experiment, the inverse BEM is used to determine the distribution of particle velocity on the open end of a rectangular duct. Measurements of the particle velocity using the two-microphone method are compared to those predicted using the inverse BEM. In the third experiment, the inverse BEM is used to identify the mechanism of an aeroacoustics noise problem (a vortex tone) created by flow over a circular rod. The spectrum of the tonal sound within the flow field is reconstructed using the inverse BEM, and a near field spatial reconstruction of the sound pressure shows clearly the nature of the vortex source. This paper will also show that only a relatively small number sound pressure measurements are required for the inverse BEM and that these measurements need not be in the near field of the source nor conformal to it to obtain good results.

1:40

2pNCb3. Beamforming-based partial field decomposition in nearfield acoustical holography. Eui Seok Hwang (Daewoo Electron. Co., LTD, Gyonggi-Do, Korea) Jong Cheon Sun, and Yeon June Kang (Acoust. & Vib. Lab., Seoul Natl. Univ., San 56-1 Shinlim-Dong Kwanak-Gu, Seoul 151-744, Korea)

In this paper, a methodology for partial field decomposition is presented that is based on the beamforming algorithm for the application of acoustical holography to the composite sound field generated by multiple incoherent sound sources. The unique aspect of the method is the use of the estimated sources identified via beamforming as the reference sources in the decomposition procedure. Numerical simulations are performed with particular emphasis on the robustness of the proposed technique under various situations. The experimental results show that the method identifies the sources very effectively regardless of the distance from source to reference and also has potential to be used for application to distributed sources.

2pNCb4. Two-microphone nearfield acoustical holography. Yong Thung Cho, J. Stuart Bolton, Yong-Joe Kim, (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47906-2031), and Hyu-Sang Kwon (Korea Res. Inst. of Standards and Sci., Yuseong, Daejeon 305-600, South Korea),

In conventional Nearfield acoustical holography (NAH), a large number of measurements is required to cover the entire source area. In addition, when the complete source comprises several incoherent sub-sources, the sound field must be decomposed into coherent partial fields before projection. The latter operation is made possible by using a fixed array of reference microphones and computing the reference cross-spectral matrix and the transfer functions between the reference microphones and the field microphones. Here it is shown that by combining the transfer functions from the reference to the field microphones obtained from one measurement with reference cross spectra measured in another, the sound field radiated any combination of different level (fixed-directionality) sub-sources can be visualized from measurements made with only enough reference microphones to represent the incoherent sources. The implementation of the procedure will be illustrated using an array of two loudspeakers, each fed by an independent white noise source. Comparisons are made of directly measured partial fields and those generated using only the reference information. By using this procedure, it is possible to avoid repetitive measurements when visualizing the sound fields radiated by different combinations of subsources having different levels, so long as the directionality of each subsource is fixed.

2pNCb5. Identification of weak spots in the sound insulation of walls using a spherical microphone array. Bradford N. Gover and John S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

A beamforming microphone array can be useful for the identification of directions of arrival of stronger components of the sound field. When these directions can be traced back to locations in the room containing the array, they can indicate the source of the detected sound. A previously developed spherical array [J. Acoust. Soc. Am. **112**, 1980–1991 (2002)] has been used for the identification of localized weak spots in otherwise highly insulating walls. A sound source and the array were placed on opposite sides of a wall sample constructed between two reverberation rooms. The omnidirectional impulse response was measured to each of the 32 array microphones, and subsequent beamforming resulted in 60 directional impulse responses at the array position, distributed over all directions. This set of responses was analyzed to identify directions of peak sound transmission through the wall. The walls tested had STC ratings greater than 50, intentionally modified to contain weak spots or defects, most not severe enough to affect the STC. The results obtained with the array approach were compared to a brute force scan with a microphone located 0.25 m from the wall. The array approach was capable of detecting even minor defects in walls.

TUESDAY AFTERNOON, 18 OCTOBER 2005

SALON A, 1:00 TO 5:20 P.M.

Session 2pNCc

NOISE-CON and Noise: Product Noise and Vibration Control – Case Studies I

Michael J. Lucas, Chair
Ingersoll-Rand, Industrial Systems, P.O. Box 1600, Davidson, NC 28036

2pNCc1. Automotive engine air intake system with variable noise control. David J. Moenssen, Mark D. Hellie, John D. Koston, and Christopher E. Shaw (Visteon Corp., One Village Center Dr., Van Buren Twp., MI 48111)

Engine air intake systems are routinely tasked with delivering a specific target sound which involves meeting an overall noise level and, in many cases, desired frequency content over the entire engine speed range. In order to meet these targets, it is generally necessary to incorporate one or more reactive tuning devices, such as Helmholtz resonators, into the intake system. Traditional devices provide deep attenuation at their designed frequency, but they also introduce undesirable sideband resonances at a higher and a lower frequency. Even after the addition of several devices, it may still not be possible to match the desired intake noise targets due to their deep attenuation and sideband amplification. The subject of this work is to introduce an electronically controlled variable noise control (VNC) device for engine air intake systems which is capable of adjusting the air intake system's frequency response as commanded by the engine operating conditions. The VNC device permits the desired amount of attenuation of peaks in the air intake noise without introducing undesirable sideband resonances. In addition, because the tuning is controlled

electronically, the VNC device can deliver a target-specific response using the same hardware across multiple vehicle programs.

2pNCc2. Application of system engineering processes to analyze and predict engine cooling fan system noise for off-highway machines. Christopher P. Masini (Sauer-Danfoss, 2800 E. 13th St., Ames, IA 50010) and J. Adin Mann III (Iowa State Univ., Ames, IA 50011)

System Engineering processes were applied to create a Cooling Fan System Noise Analysis Tool for a back-hoe loader machine. The Cooling Fan System Noise Analysis Tool combined elements of aeroacoustic theory, Fan Law, sound power measurements and particle image velocimetry into a single computer analysis tool. The cooling fan system consisted of a cooling fan, multiple radiators in front of the cooling fan, a shroud, a mock engine behind the cooling fan, and a simulated engine compartment. A vortex flow structure was measured in front of the cooling fan. The cooling fan system sound power spectrum was measured. The radiated sound power spectrum for the vortex interaction with the fan blades was calculated. Measured and predicted cooling fan system sound power results were compared. The overall structure and approach will be presented along with an overview of the theory and initial results.

1:40

2pNCc3. Air flow measurement techniques applied to noise reduction of a centrifugal blower. John W. Laage, Ashli J. Armstrong, Daniel J. Eilers, Michael G. Olsen, and J. Adin Mann III (Mech. Eng., Iowa State Univ., 2032 Black Eng. Bldg., Ames, IA 50011)

The air flow in a centrifugal blower was studied using a variety of flow and sound measurement techniques. The flow measurement techniques employed included Particle Image Velocimetry (PIV), pitot tubes, and a five hole spherical probe. PIV was used to measure instantaneous and ensemble-averaged velocity fields over large area of the outlet duct as a function of fan position, allowing for the visualization of the flow as it leave the fan blades and progressed downstream. The results from the flow measurements were reviewed along side the results of the sound measurements with the goal of identifying sources of noise and inefficiencies in flow performance. The radiated sound power was divided into broadband and tone noise and measures of the flow. The changes in the tone and broadband sound were compared to changes in flow quantities such as the turbulent kinetic energy and Reynolds stress. Results for each method will be presented to demonstrate the strengths of each flow measurement technique as well as their limitations. Finally, the role that each played in identifying noise sources is described.

2:00

2pNCc4. Meeting 2006, outdoor noise directive (OND) noise levels for a diesel engine driven air compressor: A case study in noise reduction. David F. Rowe, Jr. (Ingersoll-Rand Utility Equipment, 501 Sanford, Ave., Mocksville, NC 27028)

In January 2006, the noise limits for many products in the European Union will drop by 2–3 dBA, as directed by 2000/14/EC ''Noise Emission in the Environment by Equipment Used Outdoors,'' commonly called the ''Outdoor Noise Directive,'' or ''OND.'' Air compressors are among the products addressed by this directive. At Ingersoll-Rand, significant effort has been directed at meeting the challenge of reducing noise on a variety of diesel engine driven air compressor platforms, ranging from 15 to 350 kW diesel engine power ratings. This paper presents a case study of the noise reduction on a 750 cfm ($21\ \mathrm{m}^3/\mathrm{min}$) air compressor operating at 300 psig (21 bar), to meet the 2006 OND noise limit of 100 LwA.

2:20

2pNCc5. Double throat pressure pulsation dampener for oil-free screw compressors. Michael J. Lucas (Ingersoll-Rand Co., P.O. Box 1600, Davidson, NC 28036)

This paper describes a recent invention at Ingersoll-Rand for reducing the pressure pulsations in an oil-free screw compressor. Pressure pulsation is a term used in the air compressor industry to describe the rapid change in pressure with time measured in the downstream piping of the air compressor. The pulsations are due to the rapid opening and closing of the screws as the compressed air is eject from the compressor into the piping system. The pulsations are known to produce excessive noise levels and high levels of vibration in the piping system. Reducing these pulsations is critical to achieving a quiet running compressor. This paper will describe the methodology used to analyze the data and show both computational and experimental results achieved using the pulsation dampener. A patent for this design has been filed with the US patent office.

2:40

2pNCc6. Noise reduction techniques in the design of a pneumatic-driven hand held power tool. Christian M. Skinner (Ingersoll-Rand Co., 1467 Rte. 31 South, Annandale, NJ 08801)

Pneumatic-driven hand-held power tools generate noise in the workplace. Current legislation in Europe and the USA aims at protecting workers against noise exposure. In the United States, the Occupational Safety and Health Administration (OSHA) requires that employers create a hearing conservation program if the noise exposure exceeds 85 dB(A). In the European Community under the Directive 2003/10/EC, employers are re-

quired to provide hearing protection if the noise exposure within the working environment exceeds 80 dB(A) and must require hearing protection to be worn if the noise exposure exceeds 85 dB(A). This paper examines the sources of noise which contribute to the overall noise from a hand-held power tool. A test plan was developed to identify these individual sources of noise and to determine if structure-borne noise or airborne noise is the dominant source relative to the overall noise level. The measurements were performed per International Standards Organization (ISO) 15744. This paper will describe the methodology used to identify the noise sources and reduce the overall noise of a hand-held power tool.

3:00–3:20   Break

3:20

2pNCc7. Acoustic characterization of pneumatic percussion tools. Jorge Muract, Rahul Kadam, Ricardo Burdisso, and Marty Johnson (Vib. and Acoust. Labs, Virginia Tech, 143 Durham Hall, VA 24061-0238)

Pneumatic powered percussion tools are extensively used in construction industry. They are one of the noisiest machines in the construction industry generating noise levels well above 110 dBA which are well beyond the permissible exposure limit (PEL) of 85 dBA. The paper presents comprehensive analysis of the noise generated from these percussion tools. Noise tests were carried out on different percussion tools ranging from small chipping hammers to rock drills from two major construction equipment manufacturing companies. These tests were carried out in an anechoic room as well as in simulated operating conditions to determine the overall sound power levels. A spherical array of microphones was used to obtain an accurate estimate of the overall sound power levels and the directivity. The overall sound power radiation was found to be in the range of 105–115 dBA. An advanced 63 microphone phased array was used to successfully locate and identify the major sources of noise from these tools. The outcome of the tests is illustrated in detail in the paper. Further the paper will suggest noise control methods to reduce overall sound power radiation and discuss potential performance levels.

3:40

2pNCc8. Sound source identification technique using the inverse boundary element method. Jiaohua Shi (Caterpillar Inc. Peoria, IL 61656) and Ferdy Martinus (Univ. of Kentucky, Lexington, KY 40506)

Numerical acoustic prediction requires the knowledge of the sound source, in most cases in terms of surface vibration data. Traditional approaches to obtain this information are by performing a structural finite-element analysis or by direct measurements. Both of these approaches are time consuming and may not be feasible, particularly when the source dimension is large, the excitation cannot be identified accurately or when due to safety and time concerns, direct measurements are not preferred. In this paper, an advanced source identification technique using the inverse boundary element method is presented. This approach is advantageous since only a small number of sound-pressure measurements are needed and only a boundary element model is required, therefore eliminating the necessity of building a finite-element model. As an example, the reconstruction of surface vibration on a diesel engine is presented. The reconstructed surface vibration was then used to predict the acoustic responses in the field in term of sound-pressure and radiated sound power. It will be shown that the predicted sound-pressure and radiated sound power agrees well with the measured values. The reconstructed surface vibration was also compared to several direct vibration measurements obtained using accelerometers.

4:00

2pNCc9. Piping noise transmission loss calculations using finite element analysis. Richard Eberhart, Fred W. Catron, Allen C. Fagerlund, Denis G. Karczub, (Fisher Controls Intl., 301 South 1st Ave., Marshalltown, IA 50158), and J. Adin Mann III (Iowa State Univ., Ames, IA 50010)

The prediction of noise radiated by piping downstream of a control valve is subject to various uncertainties. One of the significant sources of uncertainty is the pipe-wall transmission loss. Due to the difficulties in experimentally measuring pipe-wall transmission loss accurately, and

practical difficulties of taking into account pipe length and boundary conditions, an analytical approach for the calculation of transmission loss is required. The feasibility of uncoupled structural-acoustic finite element based calculations of transmission loss is being investigated for this purpose. By developing the use of finite element based calculations of transmission loss, it is hoped to provide a simple analysis procedure to quantify the effects of pipe length and boundary conditions on the noise level downstream of control valves in practical piping systems. It should also assist in the refinement of analytical/statistical calculations of transmission loss and noise radiation.

4:20

2pNCc10. Evaluation of transmission loss using the boundary element method for mufflers with two inlets. C. Jiang, T. W. Wu (Dept. of Mech. Eng., Univ. of Kentucky, Lexington, KY 40506), and C. Y. R. Cheng (Fleetguard, Inc., Stoughton, WI 53589)

Mufflers with multiple inlets and multiple outlets are commonly used in exhaust systems which provide the benefits of lower restriction and higher engine performance. In this paper, the direct mixed-body boundary element method (BEM) is used to predict the transmission loss (TL) of mufflers with two inlets and one or two outlets. To do this, the impedance matrix of the muffler is first computed by the BEM. The transmission loss (TL) can then be evaluated from the impedance matrix which relates the sound pressure at the inlets and the outlets to their corresponding normal particle velocities. A complex-number ratio of the sound pressure at the second inlet to the sound pressure at the first inlet will need to be supplied before the TL can be evaluated. However, the exact value of this ratio may not be known *a priori* in real applications, let alone the ratio would be a function of frequency. Therefore, an estimate has to be made and the TL needs to be presented with different possible values of the ratio. Numerical test cases are given in the paper, and results are verified by comparing to the five-point method and the six-point method.

4:40

2pNCc11. Whistler edition. Steve Gleason (Thermo King Corp., 314 W 90th St., Minneapolis, MN 55420)

In 1995 a sound-reduced version of a large transport refrigeration unit was introduced as one of the quietest options for delivering temperature-sensitive goods to stores and restaurants. It became known as the Whisper Edition. After a couple of years the service department received a few calls about these units whistling. Lab tests found a 5-kHz tone that overwhelmed the 30-hp diesel engine noise. The noise was related to the blower inlet displacement, and a simple modification seemed to resolve the problem. Sporadic reports of whistles resurfaced after a couple more years, and an investigation showed a dimension on the blower inlet needed tighter control. Once those changes were made, recommendations for positioning the blowers in the factory and field solved most problems, and replacing blower inlets addressed the few remaining problems in the field. By now the unit was known in the service department as the Whistler Edition. The problem resurfaced yet again in the factory after a couple more years, and it was there where they recognized a slight tooling ridge on the inlet ring played a role. Subsequent tests showed the ''whistle'' was actually noise from vibrations in the ring induced by airflow.

5:00

2pNCc12. Case study of modifications to a wood "I" beam framed floor-ceiling assembly which did not meet minimum International Building Code (IBC) impact insulation class (IIC) or sound transmission class (STC). Elzo F. Gernhart (PAC Intl., Inc., P.O. Box 5369, Aloha, OR 97007)

A direct comparison, using three identical floor-ceiling assemblies, which did not meet minimum building codes standards for impact insulation class as built. Each floor-ceiling assembly received different corrective measures in an effort to correct the deficiency. All corrective measures were limited to the bottom side of the assembly, since hardwood flooring was installed as the finished floor. All the floor-ceiling assemblies consisted of light weight 16-in. deep "I" beam structural framing members (joist) spaced at 24 in. center to center supporting 3/4 in.-t & g plywood with 1-1/2 in. light weight gypsum concrete covered with 1/2-in. plywood and finished with hardwood flooring. Using tired old industry standard methods, products, and systems the best impact insulation class increase was twelve points which still did not meet minimum building code standards. By installing the state of the art products, using the proper methods available today under the floor-ceiling assembly the field impact insulation class was increased 28 points to a field impact insulation class of 60 without acoustical mat on the top side of the assembly. By design the field sound transmission class was also improved (40 to 55 STC) due to product selection, installations procedures, and design.

2p TUE. PM

Session 2pNCd

NOISE-CON, Noise and Psychological and Physiological Acoustics: Sound Quality and Multimodal Responses

Klaus Genuit, Cochair
HEAD Acoustics GmbH, Ebertstrasse 30a, Herzogenrath 52134, Germany

Wade R. Bray, Cochair
HEAD Acoustics Inc., 6964 Kensington Rd., Brighton, MI 48116

1:00

2pNCd1. Aircraft sound quality for passenger comfort and enhanced product image. Gregory H. Wakefield (Dept. of Elec. Eng. and Computer Sci., Univ. of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109), Eric J. Bultemeier, Erik West, James R. Angerer, and Waman V. Bhat (The Boeing Co., Seattle, WA 98124)

Passenger cabin noise requirements for commercial airplanes are being expanded beyond the traditional focus on noise annoyance and speech intelligibility. There is increasing recognition that the passenger response to the cabin soundscape is much more complex; affecting perceptions of product quality, and impacting fatigue and comfort. Tailoring the soundscape for a preferred cabin environment requires the development of metrics that capture a range of passenger responses. In a preliminary exploration of potential metrics, a series of experiments were undertaken to investigate passenger preference for several classes of stationary and transient sounds within the passenger cabin. The design, implementation, and data analysis for these experiments is discussed, along with the application of results to enhance cabin comfort and to convey product quality. Key findings confirm what has been found in sound quality studies in other industries: sound level alone does not fully account for passenger preference.

1:20

2pNCd2. Acoustical comfort of vehicles: A combination of sound and vibration. Klaus Genuit (HEAD Acoust. GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany) Brigitte Schutte-Fortkamp (Tech. Univ. of Berlin, Berlin, Germany) and Andre Fiebig (HEAD Acoust., Herzogenrath, Germany)

As vehicles become more and more quiet, the customer's sensitivity to acoustical comfort increases. The acoustical comfort is not independent of the vibrations the driver can feel in the seat and at the steering. The passenger of a vehicle must be regarded as part of a vibro-acoustic system. Correspondingly, the subjective judgement which passengers make about their impression of levels of acoustic comfort encompasses both sound and vibration. Achievement in this field depends on obtaining knowledge about the interaction between sound and vibration and how these factors impact subjective evaluation. To save time and money prediction tools for the estimation of sound and vibration contributions into the vehicle cabin are very important in order to simulate the final comfort with respect to sound and vibration. Based on the binaural transfer path analysis in combination with the binaural transfer path synthesis a sound and vibration reproduction in a so-called SoundCar can be realized with a very good simulation of a real situation of a car. First research tests completed for the European research project OBELICS (Objective Evaluation of Interior Car Sound) have shown that the use of SoundCar may result in more reliable sound characteristic and quality evaluation.

1:40

2pNCd3. Multimodal interaction in the perception of impact events displayed via a multichannel audio and simulated structure-borne vibration. William L. Martens and Wieslaw Woszczyk (Faculty of Music, McGill Univ., 555 Sherbrooke St. W., Montreal, QC H3A 1E3, Canada)

For multimodal display systems in which realistic reproduction of impact events is desired, presenting structure-borne vibration along with multichannel audio recordings has been observed to create a greater sense of immersion in a virtual acoustic environment. Furthermore, there is an increased proportion of reports that the impact event took place within the observer's local area (this is termed "presence with" the event, in contrast to "presence in" the environment in which the event occurred). While holding the audio reproduction constant, varying the intermodal arrival time and level of mechanically displayed, synthetic whole-body vibration revealed a number of other subjective attributes that depend upon multimodal interaction in the perception of a representative impact event. For example, when the structure-borne component of the displayed impact event arrived 10 to 20 ms later than the airborne component, the intermodal delay was not only tolerated, but gave rise to an increase in the proportion of reports that the impact event had greater power. These results have enabled the refinement of a multimodal simulation in which the manipulation of synthetic whole-body vibration can be used to control perceptual attributes of impact events heard within an acoustic environment reproduced via a multichannel loudspeaker array.

2:00

2pNCd4. The acoustical design of vehicles—a challenge for qualitative evaluation. Brigitte Schulte-Fortkamp (Tech. Univ. Berlin, Einsteinufer 25 TA7 D-10587, Berlin) Klaus Genuit and Andre Fiebig (HEAD Acoust. GmbH, Ebertstrasse 30a, 52134 Herzogenrath, Germany)

Whenever the acoustical design of vehicles is explored, the crucial question about the appropriate method of evaluation arises. Research shows that not only acoustic but also non-acoustic parameters have a major influence on the way sounds are evaluated. Therefore, new methods of evaluation have to be implemented. Methods are needed which give the opportunity to test the quality of the given ambience and to register the effects and evaluations in their functional interdependence as well as the influence of personal and contextual factors. Moreover, new methods have to give insight into processes of evaluation and their contextual parameters. In other words, the task of evaluating acoustical ambiences consists of designating a set of social, psychological, and cultural conditions which are important to determine particular individual and collective behavior, attitudes, and also emotions relative to the given ambience. However, no specific recommendations exist yet which comprise particular descriptions of how to assess those specific sound effects. That is why there is a need to develop alternative methods of evaluation with whose help effects of acoustical ambiences can be better predicted. A method of evaluation will be presented which incorporates a new sensitive approach for the evaluation of vehicle sounds.

1920   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1920**

2pNCd5. Tonal strength of harmonic complex tones in machinery noise. Kyoung Hoon Lee, Patricia Davies (Ray W. Herrick Labs, Purdue Univ., 140 Intramural Dr., West Lafayette, IN 47906), and Aimee M. Surprenant (Purdue Univ., West Lafayette, IN 47901)

The sounds of machines often contain families of harmonically related sine waves that are referred to as harmonic complex tones. The perceived tonal strength of these types of sounds can adversely influence people's impressions of the sound. While a complex tone is comprised of many sine waves, usually only one prominent pitch sensation is produced. It can be argued that harmonic complex tones are perceived as a single entities, not as a sum of individual tones. A series of psychoacoustics tests was conducted to evaluate tonal prominence of harmonic complex tones. Two sounds of equal loudness were played to subjects. One was a harmonic complex tone in noise and the other was a single tone in noise. Subjects were asked to equalize the perceived tonalness of the two sounds by adjusting the tone to noise ratio of the single tone in noise. Tonalness and Terhardt's pitch perception models were applied to the pairs of sounds used in each test. The feasibility of replacing harmonic complex tones with a tonally equivalent simple sound was investigated, and strong correlations between Aures' tonality for the simple and complex tones were found.

2pNCd6. Case history: The use of narrow-band frequency analysis to accurately identify multiple, stationary environmental noise sources. Chris R. Depies (O'Neill Engineered Systems, Inc., 3205 Dana Dr., Burnsville, MN 55337)

The sound quality near an industrial facility is very important in determining community reaction. Often, the sound environment near these facilities is negatively impacted by the presence of low- and mid-frequency tones related to rotating equipment. Complainants often describe the noise as a "hum" or "whine." A case history will be presented involving a large industrial facility located in the Midwest. The facility reacted to a noise complaint registered by a resident located approximately 0.5 miles from the facility by contracting for an environmental survey to identify the facilities noise emission points and establish a mitigation plan. The use of a FFT Narrowband Sound Analyzer allows the noise control engineer to obtain a detailed sound signature of each potential industrial noise source. The spectrum is then compared with the sound signature measured at the complainant's location and/or throughout the impacted community to determine the individual noise sources that significantly contribute to the community noise level and quality. The source identification procedure, survey challenges, and related mitigation techniques will be discussed in the paper.

2pNCd7. Radiation properties and sound quality characteristics of refrigerator noise. Jin Yong Jeon, Jeong Ho Jeong, and Jin You (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

The characteristics of refrigerator noise in an anechoic chamber and in an actual environment were investigated. In order to predict the noise propagation in real apartment house, room acoustic simulations and measurements using different types of refrigerators were conducted. The sound-pressure level of the refrigerator noise in the real living room was much higher than in the anechoic chamber. In addition, an allowable sound-pressure level for refrigerator noise was determined by auditory experiments. For the stimuli of auditory experiments, the dry source of refrigerator noise was presented using a loud speaker at the position of the refrigerator. When the result of the subjective evaluation was at the level 2 (the noise rarely aware but comfortable), in which sound pressure level was about 25 dB(A), 95% of people were satisfied with the refrigerator noise. A semantic differential test using various adjectives was also conducted to evaluate the sound quality of refrigerator noise.

2pNCd8. The development of a noise annoyance scale for rating residential noises. Jong Kwan Ryu and Jin Yong Jeon (School of Architectural Eng., Hanyang Univ., 17 Haengdang-dong, Sungdong-ku, Seoul, Korea)

In this study, 5-point and 7-point verbal noise annoyance scales were developed. The 5-point annoyance scale for outside environmental noise was developed from a survey conducted in four Korean cities. An auditory experiment using residential noises such as airborne, bathroom drainage, and traffic noises was conducted to compare the effectiveness of the 5-point and 7-point scales for rating residential indoor noise. Result showed that the 7-point scale yielded more detailed responses to indoor residential noise. In addition, auditory experiments were conducted to develop a noise annoyance scale for the classification of common residential noises. The modifiers used in the scales were selected according to the method proposed by ICBEN (International Commission on the Biological 12Effect of Noise) Team 6. As a result, the difference between the intensity of 21 modifiers investigated in the survey and the auditory experiment was very small. It was also found that the intensity of the selected modifiers in the 7-point noise annoyance scale was highly correlated with noise levels, and that the intensity difference between each pair of successive levels in the 7-point annoyance scale was almost identical.

Session 2pNCe

NOISE-CON and Noise: Workshop on Noise Policy Developments II

William W. Lang, Cochair
29 Hornbeck Ridge, Poughkeepsie, NY 12603

George C. Maling, Jr., Cochair
60 High Head Rd., South Harpswell, ME 04079

This session is a continuation of morning session 2aNCe, Workshop on Noise Policy Developments I. The purpose of the Workshop is to discuss the latest developments in noise policy since the first Workshop was held in Baltimore, MD in 2004 July. This session will focus on manufacturing and product noise, education, and occupational noise

II. Developments on Manufacturing, Education and Product Noise Policy

1.00
Chair's Opening Remarks

1:10−2:00
Invited Panel Presentations

2:00−2:30
Discussion

2:30−2:50 Break

III. Developments on Occupational Noise Policy

2:50−3:00
Chair's Opening Remarks

3:00−4:00
Invited Panel Presentations

4:00−5:00
Discussion

Session 2pNCf

NOISE-CON and Noise: Measurement Techniques and Transducers

Gerard P. Carroll, Chair

Naval Surface Warfare Center, Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817

3:00

2pNCf1. Proposed piezoceramic excitation for rotational and translational structural mobility measurements. Andrew T. Kankey, Gary Koopmann (Ctr. for Acoust. and Vib., Penn State Univ., 157 Hammond Bldg., State College, PA 16802), Steven Hambric, and John Fahnline (Penn State Univ., State College, PA 16802)

A piezoceramic array is proposed that will allow for more accurate measurement of translational and rotational structural mobility compared with current methods. This lightweight, compact array could be attached to a structure and would excite a normal force as well as bending and torsional moments in the structure. This paper focuses on verifying the ability of piezoelectric ceramic patches, when aligned in certain formations, to create normal forces and moments to excite a structure and allow easy measurement of structural mobility. A simple method for calibrating the patches is suggested using reciprocity. The test structure used is a 0.125 inch thick, fixed-fixed 12 x 1 inch aluminum beam to which two 1 x 0.5 inch piezoelectric ceramic patches are attached. Depending on the orientation of the patches, a normal force, bending moment, or torsional moment is created. Accelerometers are used to measure the mobility at the center point of the beam. This method combined with the proposed array could be moved to any point on a structure to excite the structure and measure the point structural mobility. Experimental results are compared to numerical results from a finite element program to verify the representation of the piezoceramic patches.

3:20

2pNCf2. Investigation of accelerometer mass loading using a laser Doppler vibrometer. Gerard P. Carroll and Matthew A. Craun (Naval Surface Warfare Ctr., Carderock Div., Bethesda, MD 20817-5700)

Laser Doppler vibrometer (LDV) measurements were used to quantify the effects of accelerometer loading and attachment methods on the response of a shaker driven submerged damped and stiffened plate. The complex plate transfer admittance over a frequency range from 5 to 40 kHz was first measured using the LDV without the accelerometers or the accelerometer mounting pads attached to the plate. The accelerometers and the mounting pads were then attached and the LDV measurements were repeated to determine their mass loading effects. With the accelerometers attached the complex plate transfer admittance was measured using both the accelerometer and the LDV outputs. These results were compared with the intention of developing a complex correction term to be applied to the accelerometer data to correct for data errors associated with accelerometer resonances and loading effects. This however was not carried out due to observed irregularities on the LDV measurement data possibly due to acousto-optic phase modulation of the laser signal.

3:40

2pNCf3. Structural-acoustic measurements of pipes with ninety-degree elbows, under water loading. Benjamin J. Doty (Jet Propulsion Lab., Caltech, 4800 Oak Grove Dr., M.S. 157-500 Pasadena, CA 91109), Stephen A. Hambric, Stephen C. Conlon, and John B. Fahnline (Penn State Univ., State College, PA 16804)

The complicating effects of an elbow, along with internal and external fluid loading on a thick-walled pipe were experimentally examined. The important vibroacoustic behavior of a test structure consisting of a Schedule 40 pipe with a ninety-degree elbow was measured in ARL/Penn State's reverberant water tank. To characterize coupling between the vibrating structure and the surrounding acoustic fluid, several transfer functions between structural vibration and acoustic radiation were measured using force hammer impulse excitation of the structure as well excitation by an acoustic source. A reciprocal measurement method was validated and used to quantify the sound power radiated by an acoustic source inside the pipe. Results provide a rigorous characterization of the vibroacoustic behavior a test structure with simple geometry and are suitable for validation of analytical and numerical models.

4:00

2pNCf4. Measuring high frequency valve noise to evaluate interference with ultrasonic flow meters. J. Adin Mann III (Mech. Eng., Iowa State Univ., 2032 Black Eng. Bldg., Ames, IA 50011), Allen Fagerlund, Charles Depenning, and Fred Catron (Fisher Controls Intl., Marshalltown, IA 50158)

Ultrasonic flow meters are installed in lines with particular concern on the location of flow noise sources, such as valves and other geometry changes and restrictions in a piping system. The ultrasonic flow meter operates with a tone burst, typically in the range of 100 to 300 kHz. While this frequency range is far above the typical range for noise control in piping systems, there is good evidence that the flow noise sources generate sound in the operating frequency range of the ultrasonic flow meter. The goal of the work was to establish a procedure to measure the noise generated by piping elements in the frequency range of the ultrasonic flow meter operation. The flow disturbance is placed upstream of the ultrasonic flow meter. The internal noise spectrum are measured at three locations: one upstream of the flow disturbance, one between the flow disturbance and the flow meter, and one downstream of the flow meter. Some available results in the literature will be reviewed along with presenting the experimental setup and preliminary results. The results will show the capabilities of the measurements in this frequency range.

4:20

2pNCf5. Microflown based sound pressure microphone suitable for harsh environments. D. R. Yntema (Univ. of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands) and Hans-Elias de Bree (Microflown Technologies, 6900 AH Zevenaar, The Netherlands)

There are several cases where a sound field reconstruction or prediction is required under harsh conditions such as high temperature, humidity or chemical attack. A regular pressure microphone will not last long under these conditions. Electret based pressure microphones stop working well above 70 degrees centigrade and other type of pressure microphones often operate with a built in amplifier that does not function above 120 degrees centigrade. The functionality of a MEMS based Microflown acoustic particle velocity sensor in air lies in the use of two heated platinum wires that are resistant to high temperatures and chemical attack. The wires are supported by silicon that has no other function than provide support. A pressure microphone is made based upon the Microflown principle by putting it in the opening of an enclosure. In this paper a silicon and platinum based sound probe for harsh environments is created, combining particle

1923   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1923**

2p TUE. PM

velocity and pressure measurements in a harsh environment. Use of this sensor is possible up to 250 degrees centigrade, in humid and under most chemical environments. The probe realization as well as calibration measurements are presented.

4:40

2pNCf6. Sound pressure mapping with a hand held sound level meter/real time analyzer. Roger J. Anderson and Zack Kertzman (Acoust. & Noise Control, P.O. Box 100, Carver, MN 55315-0100)

For the past twenty years or so we have been using acoustic intensity to map out the surface of an object in order to have a visual representation of that noise source. More recently the trend has been to use large microphone arrays and acoustic holography. Acoustic intensity is reasonably easy to do but requires an intensity probe, a two channel analyzer some training and special software. In total the system cost can easily exceed $20,000. Acoustic holography systems provide more information but are much more expensive and need trained personnel. However, a low cost single channel SLM/RTA can be used to generate very good sound pressure contour maps that help noise control engineers understand noise sources. The other advantage is that it is simple to use requiring only an hour or less of training. Several example pressure maps will be presented and compared to acoustic intensity maps.

5:00

2pNCf7. Design and analysis of a hemi-anechoic chamber at Michigan Technological University. Jason Dreyer, Ashish Jangale and Mohan D. Rao (Michigan Technolog. Univ., Houghton, MI 49931)

A four-wheel chassis roll dynamometer test facility was installed on the campus of Michigan Technological University (MTU). The chassis dynamometer was enclosed in a soundproof hemi-anechoic room in order to conduct noise radiation measurements on test vehicles. All surfaces of the room, except the floor and control room window, were acoustically treated with donated tetrahedral acoustic cones and panels. The acoustic absorption properties of these materials were characterized through reverberation chamber and impedance tube testing, and the effects of air gaps, cone orientation, and cone mounting materials were qualitatively evaluated. The design of the wall, ceiling, and door treatments of the chamber was based on the sound absorption properties of these materials, in addition to spatial constraints and cost considerations. The treated chamber acoustics were predicted based on the amount of acoustic material that could be applied to given chamber dimensions and would still preserve the functionality of the room. These predictions were validated through evaluation of the actual room treatment based on average reverberation time at 100-Hz third-octave band, free sound field characteristic 6-dB reduction in sound pressure level (SPL) per doubling in distance from source, noise reduction at the chamber boundaries, and background SPL Noise Criteria (NC) Rating.

TUESDAY AFTERNOON, 18 OCTOBER 2005

ROCHESTER ROOM, 1:30 TO 3:30 P.M.

Session 2pNS

Noise, NOISE-CON and ASA Committee on Standards: Specifying Uncertainties in Acoustic Measurements II

Ralph T. Muehleisen, Cochair

Illinois Inst. of Technology, Civil and Architectural Engineering, 3201 South Dearborn, Chicago, IL 60616

William J. Murphy, Cochair

National Inst. for Occupational Safety and Health, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998

Chair's Introduction—1:30

Contributed Papers

1:35

2pNS1. Compact test method for the evaluation of acoustical transmission loss and insertion loss of new helmet material samples. Matthew G. Maher, Robert D. Collier, and Laura E. Ray (Thayer School of Eng., Dartmouth College, Hanover, NH 03755)

There is a need to establish a simple and accurate measurement technique for determining the transmission loss of sample materials for helmets over a frequency range of 300–3 kHz. Standard methods, e.g., ASTM E 90-02, for measuring transmission loss of building materials and structures, based on adjacent reverberation chambers, are too costly and impractical. A 1.22-m-long double-wall tube, packed with Owens Corning

R13 insulation, has been fabricated using QUIK-TUBE^{TM} cardboard concrete forms of 200 and 300 mm diameters. A circular sample of material, also 300 mm in diameter, is placed on the end of the tube and subjected to an external sound field. Transmission loss is established by external and internal microphones. This paper describes the measurement and analysis procedures and examines the associated variables and error terms. Results are presented for 16 material samples with surface weights covering a range from 0.3 to 14.7 kg/m^2 and compared with analytical predictions including mass law models. The acoustical characteristics of commercial helmet materials and liners are evaluated in the context of hearing protection systems. The transmission loss measurement procedure has the potential for meeting standardization objectives.

1:55

2pNS2. Estimation of room absorption of acoustic cones based on impedance tube measurements. Jason Dreyer and Mohan D. Rao (Mech. Eng. - Eng. Mech. Dept., Mich. Technolog. Univ., 1400 Townsend Dr., Houghton, MI 49931)

Acoustic cones are often the primary absorption treatment choice for constructing anechoic chambers. Reverberation chamber estimation of room absorption will most accurately represent the actual effect of three-dimensional treatments, such as cones. However, for low frequencies, large reverberation chambers are often required, which may not be available due to space and cost constraints. Impedance tubes measurements are more accurate than reverberation chamber measurements in terms of sound absorption coefficient. Conversely, due to the two-dimensional nature of impedance tubes, room absorption of three-dimensional objects is difficult to estimate. Absorption coefficient of the material is a function of not only surface area but also material thickness. This paper proposes a technique to estimate the room absorption of three-dimensional treatments based on measured absorption coefficient of different thickness treatment specimens by impedance tube. This extrapolation of impedance tube measured absorption coefficient to actual room absorption can be used to evaluate the effectiveness of absorptive acoustic room treatments, including effects of adhesives and air gaps, without the need of constructing the actual treatment prototypes. The validity of this impedance tube room absorption estimation was verified for a tetrahedral cone treatment using a large reverberation chamber room absorption measurement.

2:15

2pNS3. Tone prominence ratio, error sources, and proposed solutions. G. (Randy) Stanley (HEAD Acoust., Inc., 6964 Kensington Rd., Brighton, MI 48116, rstanley@headacoustics.com), Jeffrey DeMoss, and Anil Damani (Dell, Inc., Round Rock, TX 78682)

Tone prominence ratio is finding increasing use as a tool for product sound design. ECMA-74 Annex D offers a procedure for calculating tone prominence ratio and qualifying tone audibility. However, there are significant potential sources of error when utilizing the ECMA-74 procedure. These include possible misidentification of tone frequency, and ambiguities about what constitutes tone audibility. Misidentification of tone frequency can occur when closely spaced tones exist in the spectrum; whereas, other errors may result from individual subjective assessment of tone audibility. The existing ECMA method does not specify how to identify the primary tone frequency when such ambiguities exist. Real-world examples of potential error sources are presented along with proposed solutions. An unambiguous peak prominence ratio value reporting system is presented along with data to support its improved relevance in terms of hearing perception and subjective annoyance.

2:30

2pNS4. Accuracy of the International Space Station acoustical modeling. Gregory D. Pilkinton and Samuel A. Denham (The Boeing Co., 13100 Space Ctr. Blvd., Houston, TX 77059)

The Space Station Program uses numerous algorithms to simulate on-orbit conditions to define the internal (i.e., habitable) environment of the International Space Station (ISS). Since acoustics has become one of the top ten habitability concerns for the ISS, accurately predicting the acoustic environment that the crewmembers will be exposed to is important. Before each new stage of the ISS, the acoustic predictions are updated. The acoustic environment changes depending on whether hardware is delivered or returned at each stage. If the predictions show that the noise levels are above safety requirements, recommendations are made to the crewmembers to wear hearing protection. For this reason, the predictions of the USL acoustic environment need to be as accurate as possible. This paper details the uncertainty in the predictions of the on-orbit acoustic environment in one module of the ISS, namely, the United States Laboratory (USL).

2:45

2pNS5. The effects of receiving room dimensions and absorption in the measurement of impact insulation class (IIC) rating. Rose Mary Su and Gregory Tocci (Cavanaugh Tocci Assoc., Inc. 327F Boston Post Rd., Sudbury, MA 01776)

The measurement of the Impact Insulation Class (IIC) rating of any floor/ceiling construction requires the use of a certified laboratory space. In a recent investigation into the IIC rating of a new floor system, several tests were conducted in a certified laboratory. In many tests, the IIC rating was controlled by the 100-Hz 1/3-octave frequency band, despite changes in the floor/ceiling construction. The base floor/ceiling construction included a wood structural floor on open-web wood joists and a GWB ceiling below. This paper will investigate the possibility that room resonances influenced the IIC ratings of the several floor systems tested. These data will be compared with IIC data collected for the same floor construction with the GWB ceiling removed. The removal of the GWB ceiling increased the receiver room volume and exposed the glass fiber insulation to the receiver room, thus eliminating the control of the 100-Hz 1/3-octave band over the IIC rating.

3:00

2pNS6. Uncertainty calculation for noise attenuation measurements of hearing protectors. Fabiano R. Lima, Samir N.Y. Gerges, and Thiago R.L. Zmijevski (Federal Univ. of Santa Catarina (UFSC)—CTC-EMC—Campus Universitario S/N - Trindade-Florianopolis SC, Brazil)

The objective of this paper is to present a metrology study necessary for accreditation of the Industrial Noise Laboratory (LARI) at The Federal University of Santa Catarina-Brazil for hearing protector noise attenuation procedures using the Real Ear Attenuation at Threshold (REAT) method of the Brazilian National Institute of Metrology Standardization and Industrial Quality-INMETRO. A model for the calculation of measurement uncertainty was developed. The uncertainty calculation was based on the document, ''Guide to expression of uncertainty in measurement,'' by the International Organization for Standardization, first edition, corrected and reprinted in 1995, Geneva, Switzerland. The uncertainty of each source of error was estimated. The overall uncertainty of the noise attenuation measurement of hearing protectors was calculated for each 1/1 octave band frequency test and the results applied in the single number (NRRSF—noise reduction rating for subject fit) uncertainty calculation. It was concluded that the largest uncertainty is due to the determination of the subject's hearing thresholds.

3:15

2pNS7. De-mystifying the calibration of sound measuring instruments. Mariana Buzduga (Scantek, Inc., 7060 Oakland Mills Rd. Ste. L, Columbia, MD 21046)

Sometimes the users of measuring instruments express expectations for the periodic calibrations that are beyond the purpose of these services. The relationship between measurement uncertainties and traceability is not clearly understood. In addition, some confusion persists about interpreting the calibration results relative to how the instruments meet specifications and the applicable standards. This presentation explains some metrology terms, the extent and limitations of the different types of calibrations of instruments, the relations between uncertainty, tolerances and compliance with standards, and how the calibration data should be used and interpreted. The traceability of the measurements performed with a calibrated instrument will be further discussed. Finally, some aspects about choosing the calibration intervals based on the experience acquired by Scantek, Inc. Calibration Laboratory will be presented.

Session 2pPA

Physical Acoustics and Engineering Acoustics: Thermoacoustics: What Our Customers Want and Why They Want It, Part 2

David Gardner, Chair

Los Alamos National Lab., Material Div., MS/K764, Los Alamos, NM 87545

Invited Papers

1:00

2pPA1. Opportunities and challenges for novel heat engines in passenger automobiles. George Mozurkewich (Ford Res. and Adv. Eng., MD3083/SRL, P.O. Box 2053, Dearborn, MI 48121-2053, gmozurke@ford.com)

Modern automobiles provide several opportunities for potential application of novel heat engines, including air conditioning, waste-heat recovery, auxiliary power units, and motive power. The driving force behind these potential applications is the prospect of reducing the vehicle's environmental impact by increasing its overall thermal efficiency, decreasing its emissions of greenhouse gases, or both. However, because the automotive sector is mature and highly competitive, any attempt to displace existing technology faces daunting hurdles. Weight and packaging impose tight constraints. Compatibility with existing systems is essential. Reliability must be ensured, often for 150 000 miles. Last but not least, cost is paramount in an industry where pennies matter. Some implications of these factors for implementation of thermoacoustic heat engines on passenger vehicles will be surveyed.

1:25

2pPA2. Properties of thermoacoustic engines desirable for space applications. Scott Backhaus (Condensed Matter and Thermal Phys. Group, Los Alamos Natl. Lab., M.S. K764, Los Alamos, NM 87545)

The advantages of a technology that are promoted at its inception are not always those that may eventually allow it to win out over its competition. A collaboration with Northrop Grumman Space Technology on a small thermoacoustic-Stirling electric generator has shed some light on certain advantages of thermoacoustic-Stirling engines beyond "no moving parts" or "less moving parts." Some of these advantages will be discussed

1:50

2pPA3. SAGE and DELTAE for thermoacoustic modeling. Vincent Kotsubo (Ball Aerosp., 1600 Commerce St., Boulder, CO 80301)

Two of the most extensively used modeling tools in the thermoacoustics community are SAGE and DELTAE. These codes are fundamentally different from each other, with SAGE using finite differencing and DELTAE using integration of the wave equation. Generally, there is very good agreement between the two codes, with the exceptions being conventional thermoacoustics stacks, where SAGE loses accuracy, and low temperatures, where DELTAE does not include real gas properties. The difference between these codes from the user's perspective comes down to physical viewpoint and features. DELTAE presents an acoustic picture of the system displaying first-order pressure amplitudes and volume flows, whereas SAGE presents physical components, mass flows, and pressures. Both codes model mechanical components such as transducers, mass, and springs. SAGE has a click-and-drag graphical user interface, which is advantageous in presenting a picture of the system. DELTAE uses a DOS-based interface, which can be cumbersome to a casual user. A particularly powerful feature of SAGE is the optimizer, which does a multivariable search to optimize a user-selected parameter. Both codes converge very rapidly.

2:15–2:30 Break

Contributed Papers

2:30

2pPA4. Development of a thermoacoustic organ. Nathan Stodola, Ben Tillotson, and Bradford Bruno (Union College, 218e Steinmetz Hall, Schenectady, NY 12308)

A new musical instrument based on the principles of thermoacoustics has been designed and a prototype has been constructed. The instrument, referred to as the Union Thermoacoustic Organ (UTO), is unique among instruments in several aspects. The UTO creates sound by imposing a temperature difference across a stack of closely spaced channels housed between two heat exchangers in a resonant tube (this arrangement creates a thermoacoustic heat engine). The hot and cold heat exchangers are maintained at temperature by propane flames and liquid nitrogen, respectively. The primary energy source used to create the sound is a flame, so the UTO falls into the broad classification of experimental instruments known as pyrophones. However, unlike previous pyrophones the sound produced by the UTO is easily controlled, so it can play recognizable melodies and even be integrated into ensembles with conventional instruments. The prototype UTO has 14 notes (and hence 14 tubes) which are actuated by means of a mechanical keyboard. The instrument was designed by two

senior mechanical engineering students as a capstone design project. The paper describes the principles of operation of this unique instrument, discusses the design and development of the instrument, and gives examples of its musical capabilities.

2:45

2pPA5. Harmonic prime movers. Ian Rice and Gordon Smith (Dept. of Phys. and Astron., Western Kentucky Univ., 1906 College Heights Blvd., #11077, Bowling Green, KY 42101)

Western Kentucky University is undertaking the development of a thermoacoustically powered acoustic agglomerator as a means of pretreating exhaust from coal generators before entering standard industrial electrostatic precipitators. In order to fulfill the design requirement of maintaining a clean, isolated environment for the thermoacoustic components, it will be necessary to operate a thermoacoustic prime mover in a harmonic mode, which normally is not a desirable feature in typical devices. Details of a harmonic-mode prime mover will be presented. [Work supported by KY EPSCoR and the Western Kentucky Office of Sponsored Programs.]

3:00

2pPA6. Ultrasonic thermoacoustic cooler. Young S. Kwon, H. El-Gendy, and O. G. Symko (Dept. of Phys., Univ. of Utah, 115 S. 1400 E., Rm 201, Salt Lake City, UT 84112-0830)

The development of a thermoacoustic cooler in the ultrasonic range is presented. This cooler was designed to operate at a drive frequency of 24 kHz using air as the working fluid; the resonator is 7.1 mm long and it contains a cotton wool stack with copper heat exchangers at each end. Since the ultrasonic driver is a key element in this device, a major effort was made to optimize its performance and coupling to the resonator. It is a resonant piezoelectric monomorph loaded with a metallic cone for impedance matching to the resonator. By its design it is capable of intense sound levels, of order 140 dB and higher. The attained cooling power scales with the sound power levels and a COP larger than one is achieved. Studies of heat losses consist of PIV imaging of acoustic streaming (such as Eckart) in this device and back heat flow along the stack. The developed device shows much promise for rapid cooling of small samples. [Work supported by the Office of Naval Research.]

3:15

2pPA7. Thermophysical properties, as functions of pressure and temperature, for over 300 fluids, in vapor or liquid phase. J. D. Maynard, Matthew G. Hilt, and Logan Marcus (The Penn State Univ., University Park, PA 16802)

An important undertaking for thermoacoustics, particularly for new concepts involving liquids, multiphases, etc., is the collection of thermophysical properties of fluids. Recently, a computer program was developed which calculates, with a typical accuracy of a few percent, all basic properties of pure fluids and binary mixtures over a full range of temperature, pressure, and concentration (for mixtures). The properties include mass density, specific heat at constant pressure, specific heat ratio, speed of sound, thermal expansion coefficient, viscosity, thermal conductivity, and the liquid–vapor equilibrium pressure. Currently, the program contains the necessary parameters for over 300 fluids, including simple, complex, and polar fluids. A number of parameters are used to calculate properties from reduced *PVT* surfaces for the various types of fluids; other parameters are used to obtain the liquid–vapor equilibrium line, accurate values on the liquid side of the line, and transport properties. The program is accessed with a web browser interface, so that it may be operated similarly on any computer platform.

3:30–4:15
Panel Discussion

TUESDAY AFTERNOON, 18 OCTOBER 2005

DULUTH ROOM, 1:30 TO 4:50 P.M.

Session 2pSA

Structural Acoustics and Vibration and Signal Processing in Acoustics: Experimental Modal Analysis

Joseph M. Cuschieri, Cochair
Lockheed Martin Corp., Perry Technologies, 100 East 17th St., Riviera Beach, FL 33404

Daniel A. Russell, Cochair
Kettering Univ., Science and Mathematics Dept., 1700 West Third Ave., Flint, MI 48504-4898

Chair's Introduction—1:30

Invited Papers

1:35

2pSA1. The role of modal analysis in stringed instrument design. Mark French (Purdue Univ., 138 Knoy Hall, 401 North Grant St., West Lafayette, IN 47907, rmfrench@purdue.edu)

The modal response of stringed instruments clearly contributes to sound quality. Modal analysis was applied to stringed instruments shortly after commercial data acquisition system became available. Even now, the most sophisticated equipment has been used on instruments. In spite of this effort, there is no consensus on the relationship between modal behavior of an instrument and sound quality. Useful guidelines have been developed for the lower modes of guitars and violins, but these are necessary, not sufficient conditions for superior sound quality. It now appears that superior sound quality is based on more subtle interactions of the dynamic response of the acoustically coupled structure and its material properties. Understanding the modal response of the instrument is still extremely useful. The acoustic-structure interaction of the first two or three modes of the instrument clearly demonstrates the basic

1927   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1927**

physics and forms the basis of discrete models that are very helpful in considering the first-order effects of instrument modifications. The modal response of an instrument can also be critical in placing sensors to be used as pickups. Accelerometers are being used to augment piezoelectric bridge pickups and more traditional inductance pickups and their placement greatly affects the resulting sound quality.

2:05

2pSA2. A review of multishaker excitation modal testing methods. Arun Menon and Weijie Zhao (Data Phys. Corp., 2025 Gateway Pl., Ste. 260, San Jose, CA 95110)

Common modal tests employ either impact excitation with a modally tuned hammer or random excitation with a single shaker. While these methods of testing are usually adequate for simple structures, analysis of complex and in many cases nonlinear structures requires more sophisticated test methods. This paper provides a comprehensive overview of test methods less commonly used, but essential for characterizing the dynamic behavior of complex systems and structures made of composite materials. Among the methods discussed are stepped sine, MIMO, MIMO stepped sine, and normal-mode analysis. The paper discusses the suitability of each test method for various circumstances and offers guidance on the application of each method.

2:35

2pSA3. A multi-input−multioutput (MIMO) version of the algorithm of mode isolation (AMI) and comparisons with other algorithms. Matthew S. Allen (Sandia Natl. Labs., 1515 Eubank SE, Albuquerque, NM 87112, msalle@sandia.gov)

Experimental modal analysis (EMA) plays an important role in structural acoustic model validation, system diagnosis, and in the design of structural modifications and noise control measures. Recently, EMA has also been used to detect damage in structures and to characterize the elastic properties of complex materials. A multitude of algorithms has been proposed for extracting a system's modal natural frequencies, damping ratios, and mode shapes from experimentally measured frequency response functions. The overwhelming majority of these algorithms attempts to extract all of the modes simultaneously. The algorithm of mode isolation (AMI) offers a different perspective, sequentially extracting modes in the order of their dominance in the data and then refining the extracted modal parameters through an iterative procedure. This presentation will compare these two approaches using two multi-input−multioutput (MIMO) data sets, typical of many EMA applications. The first data set consists of synthetically generated, noise-contaminated data for a near-square plate, while the second consists of experimentally measured data from the Z24 highway bridge in Switzerland. In each case, the performance of AMI is compared with that of other popular algorithms.

3:05−3:20 Break

3:20

2pSA4. Experimental modal analysis for structural dynamics and vibro-acoustics design engineering. Herman Van der Auweraer and Bart Peeters (LMS Intl., Interleuvenlaan 68, B-3001 Leuven, Belgium, bart.peeters@lms.be)

System identification plays a crucial role in structural dynamics and vibro-acoustic system optimization. A number of industrially- as well as socially relevant applications will be discussed. The most popular modeling approach is based on the modal analysis concept, leading to an interpretation in terms of visualized eigenmodes. Using the modal models, design improvements can be predicted and the structure optimized. The main modal testing procedures and modal parameter identification methods are reviewed, including both input/output and output-only approaches. The current critical elements in system identification for these applications are outlined and discussed in the broader context of the changing role of testing in the product engineering process. New trends in modal analysis that specifically address these problems are reviewed and illustrated with case studies. This includes the issues of instrumentation, test definition, measurement principles, and parameter estimation.

3:50

2pSA5. Noncontact modal testing of hard-drive suspensions using ultrasound radiation force. Thomas M. Huber (Dept. of Phys., Gustavus Adolphus College, 800 College Ave., St. Peter, MN 56082, huber@gac.edu), Daniel Calhoun (Hutchinson Technol., Hutchinson, MN 55350), Mostafa Fatemi, Randall R. Kinnick, and James F. Greenleaf (Mayo Clinic College of Medicine, Rochester, MN 55905)

The head-gimbal assembly suspension is a cantilever-like structure that holds the heads on a hard drive. We will discuss a noncontact method for modal testing, of suspensions in air, that utilizes the radiation force at the difference frequency generated by two intersecting ultrasound beams. The resulting low-frequency excitations were measured using a scanning vibrometer. This excitation technique has been demonstrated for MEMS and other small devices. There are several unique advantages of the ultrasound radiation force relative to mechanical shakers. Since the ultrasound radiation force is noncontact, a specialized test fixture was not needed; the technique was relatively insensitive to distracting resonances of fixtures and support structures. Another advantage is broadband excitation; a 550-kHz confocal ultrasound transducer excited suspension resonance frequencies from under 1 kHz to over 50 kHz. Other advantages include the ability to selectively excite different modes. For example, the amplitude of the suspension's 2.28-kHz transverse mode was suppressed by an order of magnitude by shifting the modulation phase between the two ultrasound beams by 90. In another test, the amplitude of the 6.01-kHz torsional mode was doubled by moving the ultrasound focus point from near the center to near the edge of the suspension.

Contributed Papers

4:20

2pSA6. Electronic speckle pattern interferometry for modal analysis. Thomas Moore (Dept. of Phys., Rollins College, Winter Park, FL 32789, tmoore@rollins.edu)

Electronic speckle pattern interferometry is a common method for determining resonant frequencies and visualizing mode shapes of relatively small objects; however, the apparatus required for this type of interferometry is typically expensive and extremely sensitive to ambient vibrations. A design is presented for a low-cost electronic speckle pattern interferometer, and methods to alleviate the strict requirements on limiting ambient vibrations are discussed.

4:35

2pSA7. The influence of modeling assumptions on the modal analysis of vocal fold structures. Douglas D. Cook and Luc Mongeau (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

Finite-element modal analysis studies of several vocal fold structural models were performed *in vacuo* under a variety of assumptions. The sensitivity of the vocal folds to changes in geometry and tissue properties was investigated. The Ritz method and the finite-element code ADINA were used to perform linear modal analysis of two-dimensional and three-dimensional models of the human vocal folds. The results allow for comparison and evaluation of various modeling approaches. In particular, the influence of geometrical constraints was shown to introduce errors of varying magnitude depending on the degree of anisotropy of the vocal fold tissue. [Research supported by the National Institute for Deafness and Other Communication Disorders.]

TUESDAY AFTERNOON, 18 OCTOBER 2005

SALON G, 1:30 TO 4:30 P.M.

Session 2pSC

Speech Communication: Speech Perception, Speech Intelligibility and Speech Processing (Poster Session)

Margaritis S. Fourakis, Chair
Univ. of Wisconsin, Communicative Disorders, Goodnight Hall 479, 1975 Willow Dr., Madison, WI 53706

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pSC1. Speech formant frequency estimation method based on Hilbert−Huang transform. Hai Huang and Jiaqiang Pan (Dept. of Instrumental Eng., Zhejiang Univ., 38 Zhedalu, Hangzhou 310027, China)

A speech formant frequency estimation method based on Hilbert–Huang transform (HHT) is proposed in this study. After filtering with bandpass filters with the center frequencies obtained by using the FFT analysis, speed data are decomposed into a set of intrinsic mode function (IMFs) by using the HHT analysis method. The IMFs containing formant frequencies are then identified according to the energy maximum criteria, their instantaneous frequencies and Hilbert spectra are calculated, and finally the formant frequencies of speech data are efficiently determined. The results in this study show that, compared with conventional formant estimation methods, the method based on HHT not only can give clearer descriptions of the nonlinear and nonstationary characteristics of speech signals, but also the speech formant frequencies and their variations with high time–frequency resolution and veracity. [This work was supported by Grant No. 60275004 from the National Natural Science Foundation of China (NNSFC).]

2pSC2. Effect of multiple compression channels on vowel spectra. Stephanie Bor (Dept. of Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105), Richard Wright, and Pamela Souza (Univ. of Washington, Seattle, WA 98105)

Multichannel wide-dynamic range compression (MCC) is a common amplification strategy, but whether using more channels provides an advantage for speech recognition is disputed. Some researchers have hypoth-

esized that MCC with a large number of compression channels can reduce spectral contrast and degrade recognition, but this has not been empirically tested. A detailed acoustic analysis was undertaken to determine (1) whether vowel spectra are altered by multichannel compression relative to single-channel compression, and (2) to quantify changes to vowel spectra as a function of the number of compression channels. Stimuli were eight naturally spoken vowels produced by five males and five females in an /hVd/ context. A series of compression conditions (1 through 16 channels) was created using clinically representative parameters. After compression processing, $F1$, $F2$, and $F3$ were measured using spectrograms, LPC, and FFT analysis. A quantitative measure was developed to represent spectral changes after compression. Preliminary data show spectral smoothing with more than four channels. Smaller numbers of compression channels have minimal effects on vowel spectra. These data will be related to vowel recognition scores.

2pSC3. Thresholds for vowel formant discrimination using a sentence classification task. Diane Kewley-Port, Eric Oglesbee, and Jae Hee Lee (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405, kewley@indiana.edu)

Accurate classification of vowels in sentences is challenging because American English has many acoustically similar vowels. Using a 2AFC paradigm, our previous research estimated thresholds for vowel formant discrimination in sentences that were two times smaller than the measured formant distance between close vowels. A new paradigm has been developed to estimate the ability to detect formant differences in a sentence classification task. A seven-token continuum of changes in either $F1$ or

$F2$ was generated from natural productions of ''bid'' and ''bed'' using the synthesizer STRAIGHT. These tokens were spliced into a nine-word sentence at different positions that also contained two other test words, one each from pairs ''cot/cut'' and ''hack/hawk.'' Listeners were asked to identify the three words they heard in the sentence. Listeners also identified whether ''bid'' or ''bed'' was heard when only the isolated tokens were presented. Thresholds to detect a change from ''bid'' were obtained from psychometric functions fit to the data. Thresholds were similar for the sentence and word-only tasks. Overall, thresholds in both classification tasks were worse than those from the 2AFC tasks. Results will be discussed in terms of the relation between these discrimination thresholds, vowel identification, and vowel spaces. [Work supported by NIHDCD-02229.]

2pSC4. Speech enhancement based on modified phase-opponency detectors. Om D. Deshmukh and Carol Y. Espy-Wilson (Dept. of Elec. and Computer Eng. and Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742)

A speech enhancement algorithm based on a neural model was presented by Deshmukh *et al.*, [149th meeting of the Acoustical Society America, 2005]. The algorithm consists of a bank of Modified Phase Opponency (MPO) filter pairs tuned to different center frequencies. This algorithm is able to enhance salient spectral features in speech signals even at low signal-to-noise ratios. However, the algorithm introduces musical noise and sometimes misses a spectral peak that is close in frequency to a stronger spectral peak. Refinement in the design of the MPO filters was recently made that takes advantage of the falling spectrum of the speech signal in sonorant regions. The modified set of filters leads to better separation of the noise and speech signals, and more accurate enhancement of spectral peaks. The improvements also lead to a significant reduction in musical noise. Continuity algorithms based on the properties of speech signals are used to further reduce the musical noise effect. The efficiency of the proposed method in enhancing the speech signal when the level of the background noise is fluctuating will be demonstrated. The performance of the improved speech enhancement method will be compared with various spectral subtraction-based methods. [Work supported by NSF BCS0236707.]

2pSC5. Knowledge-based formant tracking with confidence measure using dynamic programming. Sandeep Manocha and Carol Y. Espy-Wilson (Inst. for Systems Res. and Elec. and Computer Eng. Dept., Univ. of Maryland, College Park, MD 20742, smanocha@umd.edu)

In this study, we report on refinements to a formant tracking technique originally reported in [Xia *et al.*, ICSLP (2000)]. The formant tracker operates in two phases. First, it finds optimal formant track estimates in oral sonorant regions by imposing frequency continuity constraints using dynamic programming. Second, post-processing is performed to make the estimates more robust and accurate, and to extend formant tracks in nasal and obstruent regions. In recent work, we have improved on our initial estimates of the formants by combining the outputs from a 12th-order LPC analysis and a 16th-order LPC analysis. Additionally, we have added a confidence measure for each formant track in each frame. The confidence measure is based on formant continuity, competing formants, short-time energy, context information, and formant information over the entire utterance. The experiments show that most of the tracking errors are associated with a low confidence value, while the correct formants have high confidence values. The performance of the algorithm in the sonorant regions was tested using randomly selected male and female speech from the TIMIT database. [Work supported by NSF Grant No. BCS0236707.]

2pSC6. Investigating an optimum suppression rate of steady-state portions of speech that improves intelligibility the most as a pre-processing approach in reverberant environments. Nao Hodoshima and Takayuki Arai (Dept. of Elec. and Electron. Eng., Sophia Univ. 7-1 Kioi-cho, Chiyoda-ku, Tokyo 102-8554 Japan, n-hodosh@sophia.ac.jp)

Steady-state suppression has been proposed as a pre-processing approach to improve speech intelligibility in reverberant environments [Arai *et al.*, Acoust. Sci. Technol. **23**, 229–232 (2002)]. The goal of this work is to find the suppression rate that improves intelligibility the most. This is done by exploring the relationship between the suppression rates of steady-state portions of speech signals in a reverberant environment and syllable identification. This needs to be investigated because the optimum suppression rate might depend on the amount of overlap-masking, which in turn is determined by the reverberation. A syllable identification test was conducted with 21 normal-hearing listeners in two reverberant environments and six suppression rates. Results show a significant improvement in intelligibility for a 40% and a 50% suppression rate for a reverberation time of 1.1 s and a 60% suppression rate for a reverberation time of 1.3 s. Findings confirm that intelligibility is affected by both suppression rate and reverberation time. Results also show that an optimal suppression rate might not be derived from reverberation time alone. The spectral balance of the room response should also be taken into consideration. [Research supported by JSPS (16203041 and 176911).]

2pSC7. The relationship between the articulation index spectrogram and off-diagonal confusion matrix performance-intensity functions. Bryce Lobdell and Jont Allen (Beckman Inst., MC 251, 405 North Mathews, Urbana, IL 61801)

Confusion matrices (CM), similar to those of Miller Nicely 1955, as a function of the speech-to-noise ratio (SNR), have been collected for 16 consonants and 4 vowels, from 18 talkers and 14 listeners. The goal is to quantify perceptual features (denoted events) in the acoustic waveforms, from the articulation index (AI) spectrogram, and the off-diagonal ''PI-functions'' $P_{h,s}(\text{SNR})$, indices $0 \leq s, h \leq 64$ ($h = \text{hear} \neq s = \text{spoken}$). Near ''threshold'' SNRs (i.e., where the off-diagonal PI functions show confusion slightly greater than 1/64), the AI spectrograms show interesting transitions. The AI spectrogram is closely related to an information-theory measure called the channel capacity (and to jnd's). Examples of events in the ''CCgrams'' will be given for a variety of speech sounds at threshold SNRs. A cool sound-morphing demonstration will be presented, where the dominant confusion depends on SNR.

2pSC8. Speech intelligibility criteria for PA system qualification in large aircraft. Gopal Mathur (The Boeing Co., MC: H013-B308, Huntington Beach, CA 92647), Manolis Tsangarakis, David Lotts, Kenneth Barry (The Boeing Co., MC: C052-0066, Long Beach, CA 90807), and Naval Agarwal (The Boeing Co., MC: 6M6-31, Renton, WA)

The military aircraft test and evaluation community uses the Modified Rhyme Test (MRT) test for evaluating speech intelligibility of a communication system in an aircraft. An MRT test is conducted for the flight member/mission areas specified in the MIL/ANSI/ISO specifications. These standards specify that the MRT shall be conducted in a noise field that simulates the actual aircraft noise at the member's position. Since the method is based on the perception of words by listeners, there are no limitations in respect of the characteristics of the sound system or those of the environment. Modern communication systems, however, incorporate several new electronic designs, e.g., automatic gain control, analog-to-digital (ADC) conversions which can introduce non-linearity in the system. This paper will present results of investigations conducted in a large military aircraft where such features in conjunction with high background noise levels and moving speaker were considered and evaluated. Investigations were focused on ''relative'' measurements between different system configurations. The MRT was found to be the most reliable method for final verification of PA systems if performed as specified in the ANSI

S3.2-1989. Use of an alternative analytical method to measure speech intelligibility as a replacement for MRT will also be discussed.

2pSC9. On the number of channels needed to classify vowels: Implications for cochlear implants. Marios Fourakis (Dept. of Communicative Disord., Univ. of Wisconsin–Madison, Madison, WI 53706, fourakis.1@osu.edu), John W. Hawks (Kent State Univ., Kent, OH 44242), and Erin Davis (Ohio State Univ., Columbus, OH 43210)

In cochlear implants the incoming signal is analyzed by a bank of filters. Each filter is associated with an electrode to constitute a channel. The present research seeks to determine the number of channels needed for optimal vowel classification. Formant measurements of vowels produced by men and women [Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 (1995)] were converted to channel assignments. The number of channels varied from 4 to 20 over two frequency ranges (180–4000 and 180–6000 Hz) in equal bark steps. Channel assignments were submitted to linear discriminant analysis (LDA). Classification accuracy increased with the number of channels, ranging from 30% with 4 channels to 98% with 20 channels, both for the female voice. To determine asymptotic performance, LDA classification scores were plotted against the number of channels and fitted with quadratic equations. The number of channels at which no further improvement occurred was determined, averaging 19 across all conditions with little variation. This number of channels seems to resolve the frequency range spanned by the first three formants finely enough to maximize vowel classification. This resolution may not be achieved using six or eight channels as previously proposed. [Work supported by NIH.]

2pSC10. Auditory and visual biasing for sentence recognition in noise. Jennifer J. Main and Arlene E. Carney (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455)

Previous research [A. E. Carney *et al.*, J. Acoust. Soc. Am. **106**, 2270 (1999)] showed that listeners have a range of visual biasing when presented with stimuli that have conflicting auditory and visual cues (e.g., visual /gi/ and /auditory /bi/). Further, individuals with strong visual bias had substantially changed labeling for synthetic continua with visual cues added in noise conditions, while those with auditory biases did not [B. R. Clement and A. E. Carney, J. Acoust. Soc. Am. **106**, 2271 (1999)]. The purpose of the current study was to determine if listeners with strong auditory or visual biases had different patterns of auditory sentence recognition in noise with and without added visual cues. Two groups of ten listeners (one with auditory biases and one with visual biases on a screening task) were presented with lists of CID sentences at varying signal-to-noise ratios in auditory-only (AO) and auditory + visual (AV) conditions. Speech reception thresholds for both groups improved in the AV condition; visually biased listeners did show somewhat greater improvement when visual cues were added. In both groups, some listeners were much more adversely affected by the addition of noise than others. Results will be discussed in terms of individual differences among listeners. [For Speech Technical Committee Best Student Paper Award.]

2pSC11. Consonant confusions for temporally smeared envelope of syllables in noise. Yang-soo Yoon, David Gooler, and Jont Allen (Univ. of Illinios at Urbana-Champaign, Speech & Hearing, Champaign, IL 61820)

The purpose of this study was to better understand the contribution of speech temporal envelope cues to consonant recognition in noise (-8, 0, 8, 16, 24, and quiet dB SNR). Four tasks were performed for hearing-impaired listeners: Speech-in-Noise test for the measure of sentence repeat ability in noise, gap and intensity detection for the measure of ability to detect changes in sound over time, and consonant identification for confusion matrices. For consonant identification, the temporal envelope of 16 consonants with a vowel /a/ (CV) was extracted from each of 26 critical bands. Each processed CV was temporally smeared by low-pass filtering (4, 8, and 16 Hz). The results show that the mean P_c increases gradually as SNR increases SNR < 16 dB for all conditions, but decreases at SNR > 16 dB for no-smearing and 8-Hz smearing conditions. Consonant confusions result from certain sets of consonants, but the degree of confusion is a function of SNR. The results also show that measures of Speech-in-Noise and gap detection are more related to pure-tone threshold, while intensity detection is less related to pure-tone threshold. However, the effect of interstimulus interval is consistent regardless of degree and configuration of hearing loss.

2pSC12. Continuous multiword recognition performance of young and elderly listeners in ambient noise. Hiroshi Sato (Inst. for Human Sci. & Biomed. Eng., Natl. Inst. of Adv. Industrial Sci. and Technol., 1-1-1 Higashi, Tsukuba, Japan, sato.hiro@aist.go.jp)

Hearing threshold shift due to aging is known as a dominant factor to degrade speech recognition performance in noisy conditions. On the other hand, cognitive factors of aging-relating speech recognition performance in various speech-to-noise conditions are not well established. In this study, two kinds of speech test were performed to examine how working memory load relates to speech recognition performance. One is word recognition test with high-familiarity, four-syllable Japanese words (single-word test). In this test, each word was presented to listeners; the listeners were asked to write the word down on paper with enough time to answer. In the other test, five continuous word were presented to listeners and listeners were asked to write the word down after just five words were presented (multiword test). Both tests were done in various speech-to-noise ratios under 50-dBA Hoth spectrum noise with more than 50 young and elderly subjects. The results of two experiments suggest that (1) Hearing level is related to scores of both tests. (2) Scores of single-word test are well correlated with those of multiword test. (3) Scores of multiword test are not improved as speech-to-noise ratio improves in the condition where scores of single-word test reach their ceiling.

2pSC13. Variability in the clear speech intelligibility advantage. Kenneth Konopka, Rajka Smiljanic, and Ann Bradlow (Linguist. Dept., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, k-konopka@northwestern.edu)

The overall intelligibility advantage for sentences produced in clear versus conversational speech is well-documented. This study looked at recognition accuracy across words in early and late positions in semantically anomalous and meaningful sentences spoken in clear versus conversational speaking styles. For both sentence types, the results showed the expected overall intelligibility advantage for clear speech over conversational speech. For the semantically anomalous sentences, in both speaking styles, a decline in keyword identification rate was observed with words earlier in the sentence being more accurately recognized than words later in the sentence. Furthermore, the intelligibility advantage for clear over conversational speech remained relatively constant across word positions. For the meaningful sentences, the decline in keyword identification rate across word positions was observed for conversational speech only. Meaningful sentences spoken in clear speech yielded a high, relatively stable word identification rate across position-in-sentence, resulting in a larger clear speech intelligibility benefit for words late in the sentence than for words early in the sentence. These results suggest that for typical meaningful sentences, the acoustic-phonetic enhancements of clear speech and the availability of semantic-contextual information combine to "boost" the intelligibility of words in late sentence positions.

2pSC14. Landmark analysis of clear and conversational speaking styles. Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45267-0379), Ann Bradlow (Northwestern Univ., Evanston, IL 60208-4090), and Joel MacAuslan (Speech Technol. and Appl. Res. Corp., Bedford, MA 01730)

Speakers appear to adjust speech production according to tradeoffs between intelligibility and economy of effort [Lindblom (1992)]. Recently, there has been much interest in investigation of differences between the clear style of speech addressed to disadvantaged listeners (non-native speakers, hearing impaired listeners, etc.) and ordinary, or conversational speaking style. Clear speech has been shown to be more intelligible across a wide range of listener types [Bradlow *et al.*, (2002); Bradlow *et al.*, (2003)], but the full range of parameters of variation remain undetermined. Recently also, the use of abrupt changes in the speech signal, i.e., acoustic landmarks, as an organizing principle for speech recognition has garnered attention [Espy-Wilson (2005)]. Using a landmark analysis procedure based on that of Stevens (1991) and Liu (1995), we present evidence that clear and conversational speaking styles can be distinguished in terms of the distribution of particular clusters of landmarks, corresponding very roughly to syllable-sized units. The implications of this differential distribution of landmarks across speaking styles for the organization of speech production will be discussed.

2pSC15. Testing theories of vowel inherent spectral change. Geoffrey Stewart Morrison and Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada, gsm2@ualberta.ca)

Models which include information about vowel inherent spectral change (VISC) have higher correct classification rates, and greater correlation with human response patterns, compared to models which only include static spectral information. Three hypotheses have been proposed as to the aspect of VISC which is relevant for human perception. All three agree that the initial target (e.g., $F1$ and $F2$ values at the beginning of the vowel) is important. The dual-target hypothesis states that the final target (e.g., $F1$ and $F2$ values at the end of the vowel) is the other relevant factor. The target-plus-slope hypothesis states that it is the rate of change in spectral values over time that is relevant. The target-plus-direction hypothesis states that only the direction of formant movement (e.g., in an $F1$–$F2$ plane) is relevant. Synthetic /e/−/ɪ/−/ɛ/ stimuli with a fixed initial target but varying final target, slope, direction, and duration were identified by monolingual Western Canadian English listeners. Models of the listeners perception were parametrized according to the three hypotheses. Results indicate that the final−target model outperformed the slope modal. Neither the linear−final−target model nor the direction model clearly outperformed the other, but a final−target model including a cubic term was superior.

2pSC16. Intelligibility of foreign-accented speech: Effects of listening condition, listener age, and listener hearing status. Sarah Hargus Ferguson (Dept. of Speech-Lang.-Hearing, Univ. of Kansas, Dole Ctr., 1000 Sunnyside Ave., Rm. 3001, Lawrence, KS 66045, safergus@ku.edu), Allard Jongman and Joan A. Sereno (Univ. of Kansas, Lawrence, KS 66045)

It is well known that, for listeners with normal hearing, speech produced by non-native speakers of the listener's first language is less intelligible than speech produced by native speakers. Intelligibility is well correlated with listener's ratings of talker comprehensibility and accentedness, which have been shown to be related to several talker factors, including age of second language acquisition and level of similarity between the talker's native and second language phoneme inventories. Relatively few studies have focused on factors extrinsic to the talker. The current project explored the effects of listener and environmental factors on the intelligibility of foreign-accented speech. Specifically, monosyllabic English words previously recorded from two talkers, one a native speaker of American English and the other a native speaker of Spanish, were presented to three groups of listeners (young listeners with normal hearing,

elderly listeners with normal hearing, and elderly listeners with hearing impairment; $n = 20$ each) in three different listening conditions (undistorted words in quiet, undistorted words in 12-talker babble, and filtered words in quiet). Data analysis will focus on interactions between talker accent, listener age, listener hearing status, and listening condition. [Project supported by American Speech-Language-Hearing Association AARC Award.]

2pSC17. The perceptual relevance of locus equations. Ariane E. Rhone (Dept. of Linguist., Univ. of Kansas, 1541 Liliac Ln., Lawrence, KS 66044, arhone@ku.edu)

Locus equations reflect the linear relationship between the second formant ($F2$) at vowel onset and midvowel nucleus as a function of place of articulation for voiced stop consonants. Locus equations have been shown to provide a robust stop-place categorization metric across a variety of languages. However, there is a discrepancy between traditional locus equation calculations and data from human perception studies. While locus equations are derived from two points in the speech signal that are separated by 60–110 ms, listeners are capable of categorizing English voiced stops when presented with as little as the initial 10–30 ms of stop-vowel syllables. In order to investigate the perceptual relevance of the locus equation, 20 English speakers read 150 /CVt/ syllables (/b/, /d/, or /g/ in one of ten vowel contexts repeated five times) in a carrier phrase. $F2$ measurements were made at vowel onset and at each glottal pulse up to and including the traditional midvowel nucleus. In addition to expected replication of previous results, this research will establish at which point in the vowel stable and distinct locus equations emerge and the relation of that point to place categorization by human listeners.

2pSC18. The grain size of auditory mismatch response in speech perception. Yang Zhang, Patricia Kuhl, Toshiaki Imada (Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98195), Toshiaki Imada, and Makoto Kotani (Tokyo Denki Univ., Tokyo 101-8457, Japan)

This phonetic study examined neural encoding of within-and cross-category information as a function of language experience. Behavioral and magnetoencephalography (MEG) measures for synthetic /ba–wa/ and /ra–la/ stimuli were obtained from ten American and ten Japanese subjects. The MEG experiments employed the oddball paradigm in two conditions. One condition used single exemplars to represent the phonetic categories, and the other introduced within-category variations for both the standard and deviant stimuli. Behavioral results showed three major findings: (a) a robust phonetic boundary effect was observed only in the native listeners; (b) all listeners were able to detect within-category differences on an acoustic basis; and (c) both within- and cross- category discriminations were strongly influenced by language experience. Consistent with behavioral findings, American listeners had larger mismatch field (MMF) responses for /ra–la/ in both conditions but not for /ba–wa/ in either. Moreover, American listeners showed a significant MMF reduction in encoding within-category variations for /ba–wa/ but not for /ra–la/, and Japanese listeners had MMF reductions for both. These results strongly suggest that the grain size of auditory mismatch response is determined not only by experience-dependent phonetic knowledge, but also by the specific characteristics of speech stimuli. [Work supported by NIH.]

2pSC19. The role of fundamental frequency and formants in the perception of speaker sex. James M. Hillenbrand (Speech and Hearing Ctr., Western Michigan Univ., Kalamazoo, MI 49008)

The purpose of this study was to determine the relative contributions of fundamental frequency ($F0$) and formants in controlling the speaker-sex percept. A source-filter synthesizer was used to create four versions of 25 sentences spoken by men: (1) unmodified synthesis; (2) $F0$ only shifted up toward values typical of women; (3) formants only shifted up

toward values typical of women; and (4) both $F0$ and formants shifted up. Identical methods were used to generate four comparable versions of 25 sentences spoken by women (e.g., unmodified synthesis, $F0$ only shifted down toward values typical of men, etc.). Listening tests showed: (1) perceived talker sex for the unmodified synthesis conditions was nearly always correct; (2) shifting both $F0$ and formants was *usually* effective ($\sim$82%) in changing the perceived sex of the utterance; (3) shifting either $F0$ or formants alone was usually *ineffective* in changing the perceived sex of the utterance. Both $F0$ and formants are apparently needed to specify speaker sex, though even together these cues are not entirely effective. Results also suggested that $F0$ is just slightly more important than formants, despite the fact that the male–female difference in $F0$ is proportionally much larger than the difference in formants. [Work supported by NIH.]

2pSC20. Examining explanations for fundamental frequency's contribution to speech intelligibility in noise.

Robert S. Schlauch, Sharon E. Miller, and Peter J. Watson (Dept. of Speech-Lang.-Hearing Sci., 115 Shevlin Hall, 164 Pillsbury Dr. SE, Univ. of Minnesota, Minneapolis, MN 55455, schla001@umn.edu)

Laures and Weismer [JSLHR, **42**, 1148 (1999)] reported that speech with natural variation in fundamental frequency ($F0$) is more intelligible in noise than speech with a flattened $F0$ contour. Cognitive-linguistic based explanations have been offered to account for this drop in intelligibility for the flattened condition, but a lower-level mechanism related to auditory streaming may be responsible. Numerous psychoacoustic studies have demonstrated that modulating a tone enables a listener to segregate it from background sounds. To test these rival hypotheses, speech recognition in noise was measured for sentences with six different $F0$ contours: unmodified, flattened at the mean, natural but exaggerated, reversed, and frequency modulated (rates of 2.5 and 5.0 Hz). The 180 stimulus sentences were produced by five talkers (30 sentences per condition). Speech recognition for fifteen listeners replicate earlier findings showing that flattening the $F0$ contour results in a roughly 10% reduction in recognition of key words compared with the natural condition. Although the exaggerated condition produced results comparable to those of the flattened condition, the other conditions with unnatural $F0$ contours all yielded significantly poorer performance than the flattened condition. These results support the cognitive, linguistic-based explanations for the reduction in performance.

2pSC21. Backward masking for speech and tones.

Allison L. Paumen, Robert S. Schlauch, and Jennifer Windsor (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, paum0004@umn)

The purpose of this research was to determine if listeners who have difficulty with backward masking for tones in noise will also have difficulty recognizing speech sounds in noise that is presented in a backward-masking paradigm. To achieve this goal, backward masking for tones (1.0 kHz) and for speech (vowel-constant syllables) was measured in a sample of children and adults. Participants included four adults, and five children ranging in age from 8 to 13 years. Results show a significant correlation between tonal thresholds and speech recognition performance in the presence of a backward masker. There was large individual variability, with some of this variance accounted for by age. [Work supported by Research Experience for Undergraduates (REU) Program, University of Minnesota.]

2pSC22. Contributions of gross spectral properties and duration of spectral change to perception of stop consonants.

Joshua Alexander and Kluender Keith (1202 West Johnson St., Madison, WI 53706, jmalexa1@wisc.edu)

All speech contrasts are multiply specified. For example, in addition to onsets and trajectories of formant transitions, gross spectral properties such as tilt, and duration of spectral change (both local and global) con-

tribute to perception of contrasts between stops such as /b,d,g/. It is likely that listeners resort to different acoustic characteristics under different listening conditions. Hearing-impaired listeners, for whom spectral details are compromised, may be more likely to use short-term gross spectral characteristics as well as durational information. Here, contributions of broad spectral onset properties as well as duration of spectral change are investigated in perception experiments with normal-hearing listeners. Two series of synthesized CVs, each varying perceptually from /b/ to /d/, were synthesized. Onset frequency of $F2$, duration of formant transitions, and gross spectral tilts were manipulated parametrically. Perception of /b/ was encouraged by shorter formant transition durations and by more negative spectral tilt at onset independent of the rate of change in spectral tilt. Effects of spectral tilt at onset were contextual and depended on the tilt of the following vowel. Parallel studies with listeners with hearing impairment are ongoing. [Work supported by NIDCD.]

2pSC23. Production and perception of whispered vowels.

Michael Kiefte (Dalhousie Univ., Sch. of Human Commun. Disorders, Halifax, NS B3H 1R2, Canada)

Information normally associated with pitch, such as intonation, can still be conveyed in whispered speech despite the absence of voicing. For example, it is possible to whisper the question ''You are going today?'' without any syntactic information to distinguish this sentence from a simple declarative. It has been shown that pitch change in whispered speech is correlated with the simultaneous raising or lowering of several formants [e.g., M. Kiefte, J. Acoust. Soc. Am. **116**, 2546 (2004)]. However, spectral peak frequencies associated with formants have been identified as important correlates to vowel identity. Spectral peak frequencies may serve two roles in the perception of whispered speech: to indicate both vowel identity and intended pitch. Data will be presented to examine the relative importance of several acoustic properties including spectral peak frequencies and spectral shape parameters in both the production and perception of whispered vowels. Speakers were asked to phonate and whisper vowels at three different pitches across a range of roughly a musical fifth. It will be shown that relative spectral change is preserved within vowels across intended pitches in whispered speech. In addition, several models of vowel identification by listeners will be presented. [Work supported by SSHRC.]

2pSC24. The interaction of acoustic and linguistic grouping cues in auditory object formation.

Kathy Shapley (Harcourt Assessment, San Antonio, TX 78259, kathy_shapley@harcourt.com) and Thomas Carrell (Univ. of Nebraska-Lincoln, Lincoln, NE 68583)

One of the earliest explanations for good speech intelligibility in poor listening situations was context [Miller *et al.*, J. Exp. Psychol. **41** (1951)]. Context presumably allows listeners to group and predict speech appropriately and is known as a top-down listening strategy. Amplitude comodulation is another mechanism that has been shown to improve sentence intelligibility. Amplitude comodulation provides acoustic grouping information without changing the linguistic content of the desired signal [Carrell and Opie, Percept. Psychophys. **52** (1992); Hu and Wang, Proceedings of ICASSP-02 (2002)] and is considered a bottom-up process. The present experiment investigated how amplitude comodulation and semantic information combined to improve speech intelligibility. Sentences with high- and low-predictability word sequences [Boothroyd and Nittrouer, J. Acoust. Soc. Am. **84** (1988)] were constructed in two different formats: time-varying sinusoidal sentences (TVS) and reduced-channel sentences (RC). The stimuli were chosen because they minimally represent the traditionally defined speech cues and therefore emphasized the importance of the high-level context effects and low-level acoustic grouping cues. Results indicated that semantic information did not influence intelligibility levels of TVS and RC sentences. In addition amplitude modulation aided listeners' intelligibility scores in the TVS condition but hindered listeners' intelligibility scores in the RC condition.

Session 2pUWa

Underwater Acoustics, Signal Processing in Acoustics and Engineering Acoustics: Sonar Performance and Signal Processing in Uncertain Environments II

Lisa M. Zurk, Cochair

Portland State Univ., Electrical and Computer Engineering Dept., 1900 SW Fourth Ave., Portland, OR 97207

David R. Dowling, Cochair

Univ. of Michigan, Dept. of Mechanical Engineering, 2019 Automotive Lab 2133, 1231 Beal Ave., Ann Arbor, MI 48109-2133

Chair's Introduction—1:15

Contributed Papers

1:20

2pUWa1. Focalization in environments with stochastic sediment uncertainties. Ralph N. Baer and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Focalization has recently been generalized to problems involving internal waves [R. N. Baer and M. D. Collins, J. Acoust. Soc. Am. **115**, 2550 (2004)]. Although there are basic differences between this stochastic focalization problem and the deterministic focalization problems that were considered previously, the essential concept of a parameter hierarchy applies to both problems. Stochastic focalization is currently being implemented and tested for problems involving uncertainties in the sediment. This source localization problem is of practical importance since there are often significant uncertainties in sediments. Like internal waves, ocean sediments can be highly complex. Although it is often desirable to invert for ocean bottom parameters, this is not always practical when source localization is the main goal. Due to the parameter hierarchy, it is often possible to determine the correct source position without determining the correct sediment parameters. [Work supported by the ONR.]

1:35

2pUWa2. Field-calibrated source tracking using channel measurements from high-frequency mobile platform transmissions. Paul Hursky and Michael B. Porter (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130)

Modern signal processing methods for acoustic arrays in a waveguide use acoustic propagation models to construct signal replicas that capture channel effects such as multipath and refraction. However, such model-based techniques require a detailed map of the propagation environment (often unavailable) and impose the significant computational burden of modeling the channel. We propose an alternative approach to this problem, in which mobile platforms such as AUV's and gliders act as acoustic beacons for calibrating the field, replacing modeled signal replicas with directly measured channel impulse response functions. Key issues are whether channel estimates made in a (typically) high-frequency communications band and at a limited set of points can be extrapolated to lower frequency and to the entire field. Furthermore, the channel response may change with time, depending on the variability in the ocean waveguide, requiring periodic refresh of the calilbration. We will demonstrate this approach in simulation and with experiment data from two high-frequency experiments, KauaiEx and the Makai Experiment, in which a variety of acoustic communications waveforms were transmitted from towed and fixed systems.

1:50

2pUWa3. Optimal sonar tactics over uncertain sediments. Donald R. DelBalzo, William J. Powers, and Bernie F. Cole (Planning Systems, Inc., 40201 Hwy. 190 East, Slidell, LA 70461, ddelbalzo@plansys.com)

Tactical patterns for monostatic sensors were developed during the Cold War for deep, uniform underwater environments, where a simple median detection range defined a fixed spacing between search ladder legs. Acoustic conditions in littoral environments are so complex that spatial variability of bottom sediment properties destroys the simple homogeneous assumption associated with standard tactical search concepts. Genetic algorithms (GAs) have been applied to this problem to produce near-optimal, non-standard search tracks for monostatic mobile sensors that maximize probability of detection in such inhomogeneous environments. The present work describes a new capability called SPEAR (search planning with environmentally adaptive response) that adds tactical adaptation to search paths in a complex, littoral environment, as new *in situ* backscattering and bottom loss information becomes available. This presentation reviews the GA approach and discusses tactical adaptation to uncertain bottom sediment properties. The results show that easily implemented dynamic changes in active pulse depression angles and frequencies can produce significant improvement in detection performance in a complex littoral area. [Work supported by NAVSEA.]

2:05

2pUWa4. Guide source depth and range translation for robust matched field processing (MFP). Lisa M. Zurk (ECE Dept., Portland State Univ., P.O. Box 751, Portland, OR 97207)

The solutions to the modal wave equation have the well-known property of orthonormality. Previous work has suggested that mode functions also satisfy an approximate orthonormality condition formed when the product of the mode functions and their spatial derivatives are integrated over the water column. In the new orthogonality expression, unity is obtained when the integrand contains eigenfunctions from sequentially propagating modes (i.e., $m = n + 1$), not alike modes ($m = n$). This relationship can be exploited in matched field processing (MFP) with a guide source to derive the replica vectors. In standard MFP formulation, the guide source measurement only provides a transfer function at the guide location. Using the new formulation, it is shown that the MFP peak shifts to a frequency-dependent depth that differs from the guide depth. This depth-shifting operation enables the generation of replica vectors for a range of target depths from a single broadband guide source. The operation does not require environmental information, and is thus robust to uncertainty. In this talk the theoretical background is developed and simulation results are presented for various geometries. In addition, recent extensions of this work to include combined range and depth shifting using the invariance principle are presented and discussed.

1934 J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005 **1934**

2:20

2pUWa5. Mode-based adaptive processing in uncertain environments.
Manish R. Velankar and Lisa M. Zurk (Portland State Univ., P.O. Box 751, Portland, OR 97207)

Previous work explored the advantage of decomposing an acoustic field into normal modes and selecting appropriate modal subspaces. One primary motivation was the ability to separate surface sources from submerged sources for subsequent matched field processing. However, it was found that for many shallow environments, the modal subspaces are not sufficiently separated for adequate interference removal through filtering. This problem was further exacerbated in uncertain environments when details of the channel are unknown. In this paper, the addition of adaptive beam-forming techniques will be considered for the ability to detect and localize targets in the presence of surface interferers. In particular, robustness of the adaptive processing to environmental mismatch will be quantified. Schemes for providing robustness to adaptive self-nulling through modal filtering will be discussed and presented.

2:35

2pUWa6. Confidence intervals for low probability for false alarms and probability for a miss at low values based small sample space support.
Arthur Baggeroer (MIT, Rm. 5-204, Cambridge, MA 02139)

When using ROCs (receiver operating characteristics) based upon experimental data derived from either experimental or simulated data, there is a tendency to concentrate on the high values of PFA and PM (miss probability) in the middle of the ROC figure. Nevertheless, most good systems want to have a performance near the upper left of an ROC (low PFA, low PM) and this is where results based upon finite sample support have the largest proportional errors. While this may now be an artifact that "probability paper" is no longer routinely used, fitting a curve through these regions is almost always misleading unless the sample support is very large. The problem is fundamentally one of fitting the "tail" of a distribution with little data. Various methods from probability such as "Chi squared" and Kolmogorov-Smirnov test can be used, but they almost always lead to rejection being tested in this low probability region. Here, we use an approach based upon "Chernov tilted densities" to capture the uncertainty of predictions in these regions when there is sparse, or small, sample support.

2:50

2pUWa7. A Bayesian approach to incorporating maximum entropy-derived signal parameter statistics into the receiver operating characteristic (ROC) curves.
R. Lee Culver, Leon H. Sibul, David L. Bradley, Jeffrey A. Ballard, and H. John Camin (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804)

Our goal is to develop a probabilistic sonar performance prediction methodology that can make use of limited knowledge of random or uncertain environment, target, and sonar system parameters, but does not make unwarranted assumptions. The maximum entropy method (MEM) can be used to construct probability density functions (pdfs) for relevant environmental and source parameters, and an ocean acoustic propagation model can use those pdfs to predict the variability of received signal parameter. At this point, the MEM can be used once again to produce signal parameter pdfs. A Bayesian framework allows these pdfs to be incorporated into the signal processor to produce ROC curves in which, for example, the signal-to-noise ratio (SNR) is a random variable for which a pdf has been calculated. One output of such a processor could be a range-dependent probability of detection for fixed probability of false alarm, which would be more useful than the conventional range of the day that is still in use in some areas. [Work supported by ONR Code 321US.]

TUESDAY AFTERNOON, 18 OCTOBER 2005

MARQUETTE ROOM, 3:15 TO 4:35 P.M.

Session 2pUWb

Underwater Acoustics: Propagation, Reverberation and Scattering

Arthur B. Baggeroer, Chair
Massachusetts Inst. of Technology, Dept. of Ocean Engineering, 77 Massachusetts Ave., Cambridge, MA 02139

Chair's Introduction—3:15

Contributed Papers

3:20

2pUWb1. Spectral coupled mode approach to three-dimensional propagation and reverberation in two-dimensional, range-dependent environments.
Wenyu Luo and Henrik Schmidt (Ctr. for Ocean Eng., Dept. of Mech. Eng., MIT, Cambridge, MA 02139, lwy@mit.edu)

The use of spectral decomposition to 3D propagation and reverberation in 2D range-dependent environments is well established [J. Fawcett and T. Dawson, J. Acoust. Soc. Am. **88**, 1913 (1990)]. The resulting 2D Helmholtz equation for each value of the separation variable can be solved by any available two-way propagation model, e.g., wave number integration [H. Fan and H. Schmidt, J. Acoust. Soc. Am. **93**, 2322 (1993)] or the parabolic equation [G. J. Orris and M. D. Collins, J. Acoust. Soc. Am. **96**, 1725 (1994)]. However, these approaches are computationally intensive for realistic propagation problems due to the large number of solutions required to adequately sample the spectral integral kernel. However, if the two-way problems are formulated in terms of coupled modes, as has been suggested for cylindrical seamount problems by Taroudakis [Proceedings of the 14th IMACS World Congress (1994)], the resulting modal eigenvalue problems can be shown to be independent of the separation variable and therefore have to be solved only once, leaving only the mode coupling to be handled separately for each value of the separation variable. This approach has been efficiently implemented using a two-way, modified version of the CSNAP single-scatter coupled-mode code for the 2D problems. The code has been validated by comparison to exact 2D solutions for range-independent problems, and applied to a wide range of shallow-water propagation problems, and to modeling of diffraction around cylindrical seamounts. The paper will describe the modeling approach and its numerical implementation, and demonstrate and discuss the significance of 3D propagation and reverberation effects for such problems.

2pUWb2. Horizontal refraction of sound by random internal gravity waves. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305), Valery U. Zavorotny (NOAA/Environ. Technol. Lab., Boulder, CO 80305), Michael A. Wolfson (Univ. of Washington, Seattle, WA 98105), and Alexander G. Voronovich (NOAA/Environ. Technol. Lab., Boulder, CO 80305)

Internal gravity waves (IWs) previously have been shown to be the main cause of horizontal refraction of sound at long-range propagation in deep water. In this paper, statistical properties of IW-induced bearing perturbations and acoustic travel-time corrections associated with horizontal refraction are studied within the geometric acoustics approximation. Two versions of the Garrett–Munk spectrum describing random IWs are considered: a model based on summation of numerically calculated normal modes of the IWs and their dispersion relations, and a traditional model that relies on analytical approximations of the modal shape functions and dispersion relations. Frequency dependence of 3D acoustic effects is modeled by varying the minimal spatial scales of IWs contributing to refraction of sound. Predictions of a theory that allows propagation of statistical moments of acoustic observables are compared to results of large-scale Monte Carlo simulations of horizontal refraction. The comparison provides an insight into significance of various assumptions underlying the theory, and demonstrates the sensitivity of acoustic quantities to the accuracy of modeling the IW field. [Work supported in part by ONR.]

2pUWb3. Mode extraction from a radially accelerating cw source in shallow water. Shane C. Walker, Philippe Roux, and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9900 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238)

Recently a broadband method was introduced for extracting the depth dependence of the modes by filtering them according to their wave number spectrum. Here, a method is introduced for using a moving narrow-band source to extract the modes over a partial-water-column spanning vertical line array. Because the modes from a moving cw source are Doppler shifted in direct relation to their radial wave numbers and the radial velocity of the source, their depth dependence can be isolated directly in the frequency domain within the limits of the frequency resolution. Interestingly, the unambiguous relation between measured time and source range directly relates the frequency resolution to the wave number resolution. Thus, the range aperture covered by the source determines the extraction limits. Because a large aperture is required for low-order mode extraction, radial accelerations of non-end-fire sources subject to currents, wind, and waves tend to limit the usable aperture to an impractical extent. As a result, a technique is demonstrated to compensate for the time-dependent Doppler shift from a radially accelerating source. The technique is applied to simulation, laboratory, and ocean data.

2pUWb4. Iso-deviant 2D gridding with efficient adaptive gridder for littoral environments (EAGLE). Erik R. Rike and Donald R. DelBalzo (Planning Systems, Inc., 40201 Hwy 190 East, Slidell, LA 70461, ddelbalzo@plansys.com)

Transmission loss (TL) computations in littoral areas require a dense spatial and azimuthal grid to achieve acceptable accuracy and detail. The computational cost of accurate predictions led to a new concept, OGRES (Objective Grid/Radials using Environmentally-sensitive Selection), which produces sparse, irregular acoustic grids, with controlled accuracy. Recent work to further increase accuracy and efficiency with better metrics and interpolation led to EAGLE (Efficient Adaptive Gridder for Littoral Environments). On each iteration, EAGLE produces grids with approximately constant spatial uncertainty (hence, iso-deviance), yielding predictions with ever-increasing resolution and accuracy. The EAGLE point-selection mechanism is tested using the predictive error metric and 2D synthetic data sets created from combinations of simple signal functions (e.g., polynomials, sines, cosines, exponentials), along with white and chromatic noise. The speed, efficiency, fidelity, and iso-deviance of EAGLE are determined for each combination of signal, noise, and interpolator. The results show significant efficiency enhancements compared to uniform grids of the same accuracy. [Work sponsored by NAVAIR.]

2pUWb5. Comparison of experimental and theoretical forward scattering by the Kermit-Roosevelt Seamounts. Arthur B. Baggeroer, Henrik Schmidt, Joseph SikoraIII, Wenyu Luo (Rm. 5-204, MIT, Cambridge, MA 02139), and Edward K. Scheer (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

In the SPICEX-LOAPEX-BASSEX experiments conducted in Fall 2004 forward scattering from the Kermit Roosevelt Seamounts was measured. There were four sources insonifying the seamounts: Kauai, S1 (a moored source), S2 (another moored source), and a ship suspended source at ranges from hundreds of kilometers to 1600 km. The signals were M sequences and LFMs centered at 75 and 250 Hz. These data were presented in J. Acoust. Soc. Am. **117**, 2549, which indicated clear convergence zones and horizontal refraction by the seamounts as well as a gradual ''fill in'' of the scattering with increasing distance from the seamounts. Here, we compare the experimental results to a model of layered conic sections with both point excitations and a modal vertical line arrays. This Kauai, moored and ship suspended sources were deployed by Scripps Institute of Oceanography and the University of Washington, Applied Physics Laboratory as part of the joint experiment.

Session 3aAA

Architectural Acoustics, Noise and NOISE-CON: Safety of Acoustical Products

Nancy S. Timmerman, Chair
25 Upton St., Boston, Massachusetts 02118-1607

Chair's Introduction—9:30

Invited Papers

9:35

3aAA1. Achieving acoustical performance with fire safe products. Thomas Fritz (Armstrong World Industries, Inc., 2500 Columbia Ave., Lancaster, PA 17603, twfritz@armstrong.com)

Recent serious fires in North and South America have pointed out potential problems with attempts to improve acoustical performance in building spaces at the expense of using acoustical treatments that may have poor performance in fire situations. Foam plastic products, sometimes not designed for exposed use in buildings, can ignite quickly and spread fire rapidly throughout a building space, resulting in fire victims being trapped within the building or not being afforded the needed safe egress time. There are ways of achieving equivalent and even superior acoustical performance without sacrificing fire safety. Acoustical products are available which can add comparable or superior acoustical treatment without the fire hazard associated with exposed foam plastic materials. This presentation is a review of the U.S. code requirements of interior finish materials, the various types of fire tests that are applied to these products, and a discussion of the achievable fire and acoustical performance.

10:05

3aAA2. Fire safety of acoustical materials. Klaus Kleinschmidt (132 Mary Catherine Dr., Lancaster, MA 01523)

Recent deadly fires in overcrowded nightclubs used open cell foam material for sound absorption. The rapid progression of fire and release of opaque smoke contributed to the inability of some of the patrons to escape the fire. Suggestions for limiting the use of such materials in places of assembly are provided. Illogical use of such sound-absorbing materials to reduce sound transmission to adjoining spaces or outdoors are discussed.

10:35

3aAA3. The new science of fiber safety: Assuring the safety of fiber-based materials for acoustic and noise control applications. John Hadley (Owens Corning Sci. and Technol. Ctr., 2790 Columbus Rd., Rt. 16, Granville, OH 43023-1200, john.hadley@owenscorning.com)

In 2001, the International Agency for Research on Cancer (IARC) assembled a group of 18 international experts on the health effects of fibers. Their task was to review the available science on the health aspects of glass, rock, and slag wool fibers. The group of experts determined that glass, rock, and slag wool insulation fibers should be removed from the IARC list of possible carcinogens. That decision was based fundamentally on the development of new science that has led to an understanding of those properties of fibers which affect their potential biological activity. This presentation will provide both an overview of this new science and provide guidelines to assure the safety of fiber-based materials used in acoustic and noise-control applications.

11:05

3aAA4. The environmental, health, and safety issues of acoustical materials: A strategy for finding, using, and evaluating information effectively. Marsha S. Bischel (Armstrong World Industries, Inc., 2500 Columbia Ave., Lancaster, PA 17603)

Concern over the safety of our indoor environments has increased in recent years. The definition of safety has also evolved to include not just life safety issues such as fire, but issues such as mold growth, toxins, the emission of volatile organic compounds, seismic concerns, and ergonomic issues. Consequently, the understanding of product safety has become increasingly more complex. Simultaneously, there has been an explosion in the number of products available to specifiers, due largely to access to the World Wide Web by international manufacturers of all sizes. Some of these manufacturers may be unable to test all aspects of product safety, or simply may be unaware of safety regulations. Specifiers can no longer assume a product is inherently safe and must do their own evaluations of product safety attributes. This paper will lay out a basic methodology for finding, using, and evaluating environmental, health, and safety information on acoustical products in an effective manner.

1937   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1937**

3a WED. AM

3aAA5. Acoustical products used as interior finish materials: Product testing and requirements. Dwayne E. Sloan (Principal Engineer, Bldg. Mater., Roofing Mater. & Systems, Underwriter's Labs. 12 Laboratory Dr., Research Triangle Park, NC 27709)

This paper will focus on the test methods, requirements, and code specifications on acoustical products when used as interior finish materials. The content will discuss the test methods and the interpretation of results of key Standards such as ASTM E84, UL 723, ASTM E136, ASTM C423, and others. The content will also discuss the code differences between Class A, B, and C materials and other important aspects of the governing building codes.

WEDNESDAY MORNING, 19 OCTOBER 2005 HENNEPIN ROOM, 8:00 TO 11:40 A.M.

Session 3aAB

Animal Bioacoustics: Temporal Patterns of Sounds by Marine Mammals

Jeanette A. Thomas, Chair

Western Illinois Univ., Lab. of Sensory Biology, 3561 60th St., Moline, IL 61265

Chair's Introduction—8:00

Invited Papers

8:05

3aAB1. Seasonal and diurnal calling patterns of Ross and leopards. Tracey L. Rogers, Gayle A. Rowney, Michaela B. Ciaglia (Australian Marine Mammal Res. Ctr., Zoological Parks Board of NSW/Faculty of Veterinary Sci., Univ. of Sydney, Mosman, Australia), and Douglas H. Cato (Defence Sci. & Technol. Organisation, Pyrmont, Australia)

The temporal calling patterns of two Antarctic pack ice seals, the leopard and Ross seal, were examined. This included seasonal onset and decline of calling (coinciding with their breeding season) as well as diurnal changes. Understanding of calling behavior has important implications for acoustic surveying, since this allows the number of calls to be related to an index of the number of animals present and to estimate abundance. The monthly changes in diurnal calling and haul-out patterns (measured via satellite telemetry) were compared. Underwater acoustic recordings were made between 14 October 2003 and 10 January 2004 off Mawson, Eastern Antarctica (660 44.243S and 690 48.748E). Recordings were made using an Acoustics Recording Package (ARP by Dr. John Hildebrand, Scripps Institute of Oceanography, La Jolla, CA) which is designed to sit on the seafloor and passively record acoustic signals. The package was deployed at a depth of 1320.7 m. The sampling rate was 500 Hz and the effective bandwidth from 10 to 250 Hz, covering the bandwidth of only the low-frequency calls of the Ross and leopard seal.

8:25

3aAB2. Diel and seasonal patterns of underwater sounds by Weddell seals, leopard seals, and killer whales in the Antarctic: When it's adaptive to be quiet. Debrah Mindach and Jeanette Thomas (Biological Sci., Western Illinois Univ.-Quad Cities, 3561 60th St., Moline, IL 61265)

Automated underwater recordings taken during the austral breeding season of the Weddell seal (*Leptonychotes weddellii*) in Antarctica also provided data on the vocalizations of predators in the area; leopard seals (*Hydrurga leptonyx*) and killer whales (*Orcinus orca*). Weddell seals inhabit fast ice areas to give birth, mate, and molt. Near the end of the breeding season in December the fast ice often breaks out and the two pack ice predators are able to move near the Weddell seal colonies and prey on them, especially pups. Recordings were taken continuously for a 2.5-min period each hour from mid-October 1977 and late-January 1978 at Hutton Cliffs and South Turtle Rock Crack, in McMurdo Sound. The leopard seals increased their trill calls when killer whales came into the area as evidenced by an increase in their frequency-modulated squeak calls. Weddell seals decreased their vocalization rate dramatically ($\sim$10 sounds/min) compared to during the peak of the breeding season ($\sim$75 sounds/min). Perhaps by being quiet, Weddell seals do not attract predators to their area.

8:45

3aAB3. Stereotyped calling patterns of a male Weddell seal. Jack M. Terhune (Dept. of Biol., Univ. of New Brunswick, P.O. Box 5050, Saint John, NB E2L 4L5 Canada, terhune@unbsj.ca) and Andrea Dell'Apa (Ancona Univ., Ancona, Italy)

Underwater calling behavior between breathing bouts of a (presumed) single adult male Weddell seal (*Leptonychotes weddellii*) was examined at Davis Station, Antarctica. Underwater calls and in-air breathing sounds were recorded on 1 and 8 December. Thirty-six postdive and 37 presurfacing sets of calls were analyzed. Dives were 461 ± 259 s (mean $\pm$ S.D.). The seal called every 22.7 ± 21.7 s. The first two calls after diving were at 15.6 ± 9.3 and 33.3 ± 9.7 s. The last two calls occurred at 38.8 ± 32.0 and 22.7 ± 20.1 s before surfacing. The first call after a dive was usually (29/36) a low-frequency growl. One of three patterns of 3–5 call-type sequences was made following 26 of 36 dives. A common four-call sequence was also recorded earlier in the breeding season at this

location and was absent at four other distant breeding sites. The timing and call-type sequences of underwater vocalizations after a dive were highly stereotyped, while those before surfacing were not. Territorial males may call throughout each dive and use stereotyped call patterns to identify themselves. [Work supported by Australian Antarctic Division and NSERC.]

9:05

3aAB4. Diel patterns in underwater sounds by Weddell seals during two breeding seasons. Charisse Coulombe and Jeanette Thomas (Biological Sci., Western Illinois Univ.-Quad Cities, 3561 60th St., Moline, IL 61265)

Automated underwater recordings of Weddell seal (*Leptonychotes weddellii*) vocalizations were made continuously once per hour for 2.5 min during the austral breeding season. Recordings were made during two breeding seasons: October 1976 through December 1976 and October 1977 through January 1978 at Hutton Cliffs, Antarctica. From radio-telemetry studies this species is known to exhibit a diel pattern in haulout on the ice. From census studies of tagged seals it is known they exhibit seasonal fluctuations in the number of seals hauled out on the ice. We studied how variations in the rate of different underwater call types reflected the haulout pattern and the changes in the colony associated with territorial defense, underwater breeding, and dispersal from the colony. The 34-call repertoire was evaluated for hourly, weekly, and annual variations in usage. The trill vocalization and the serial chug vocalization had the most significant variation for both the weekly and hourly analysis. The trill call type is thought to be important in mate attraction by underwater territorial males. The chug call is thought to be important in aggressive encounters between males. An increase rate of these call types might be a useful predictor of the onset of the mating period.

9:25

3aAB5. Diurnal vocal activity of gray whales in Laguna San Ignacio, BCS, Mexico. Melania Guerra, Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., MC 0238, La Jolla, CA 92093-0238), Sheyna Wisdom (URS Corp., San Diego, CA 92109), Sergio Gonzalez, Jorge Urban (Universidad Autonoma de Baja California Sur, La Paz, Mexico), and James Sumich (Grossmont College, San Diego, CA 92101)

Three sets of portable horizontal acoustic arrays were deployed during a week in February 2005 to gather acoustic recordings of gray whales (*Eschrichtius robustus*) in Laguna San Ignacio, one of the three major breeding/calving lagoons in Baja California, Mexico. These arrays, which were constructed by attaching a pair of autonomous flash-memory acoustic sensors to a rope, were deployed for 36 consecutive hours on two occasions, spatially covering the narrowest point of the lagoon near Punta Piedra, the area of the highest concentration of whales. Additionally a single hydrophone was deployed off a small boat to record during friendly encounters with single whales and cow/calf pairs. Each recorder's time series was analyzed for Type 1 gray whale sounds (called pops), which are pulsive, broadband, and have substantial acoustic energy between 100 and 600 Hz. The number of automated acoustic detections per hour can be compared with population sizes estimated by two visual surveys conducted by scientists of the Autonomous University of Baja California Sur, La Paz. The results of several automated analyses of both the bottom-mounted and boat-deployed recordings will be presented, with a focus on potential diurnal patterns in the vocal activity.

9:45–10:00   Break

10:00

3aAB6. Diel patterns in underwater sounds produced by beluga whales and Pacific white-sided dolphins at John G. Shedd Aquarium. Jon Brickman, Erin Tanchez, and Jeanette Thomas (Western Illinois Univ., 3561 60th St., Moline, IL 61265, J-Thomas@wiu.edu)

Diel patterns in underwater sounds from five beluga whales (*Delphinapterus leucas*) and five Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) housed at John G. Shedd Aquarium in Chicago, IL were studied. Underwater sounds were sampled systematically over 24-h periods by using a battery-operated cassette recorder and an Ithaco 605C hydrophone controlled by a digital timer, which activated every hour and then shut off after 2.5 min. Belugas had 14 sounds and Pacific white-sided dolphins produced 5 sounds. For each species, the use of some sounds was correlated with other sounds. The diel pattern for both species was similar and mostly affected by the presence of humans. Sounds gradually increased after the staff and visitors arrived, peaked during the midday, gradually decreased as closing of the aquarium approached, and was minimal overnight. These data can help identify the best time of day to make recordings and perhaps could be used to examine social, reproductive, or health changes in these captive cetaceans.

10:20

3aAB7. Temporal patterns in marine mammal sounds from long-term broadband recordings. John A. Hildebrand, Sean Wiggins, Erin Oleson, Ana Sirovic, Lisa Munger, Melissa Soldevilla, and Jessica Burtenshaw (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093, jhildebrand@ucsd.edu)

Recent advances in the technology for long-term underwater acoustic recording provide new data on the temporal patterns of marine mammal sounds. Autonomous acoustic recordings are now being made with broad frequency bandwidth up to 200-kHz sampling rates. These data allow sound recording from most marine mammal species, including, for instance, the echolocation clicks of odontocetes. Large data storage capacity up to 1280 Gbytes allow these recordings to be conducted over long time periods for study of diel and seasonal calling patterns. Examples will be presented of temporal patterns from long-term recordings collected in four regions: the Bering Sea, offshore southern California, the Gulf of California, and the Southern Ocean. These data provide new insight on marine mammal distribution, seasonality, and behavior.

10:40

3aAB8. Automatic detection of regularly repeating vocalizations.
David Mellinger (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 SE Marine Sci. Dr., Newport, OR 97365, David.Mellinger@oregonstate.edu)

Many animal species produce repetitive sounds at regular intervals. This regularity can be used for automatic recognition of the sounds, providing improved detection at a given signal-to-noise ratio. Here, the detection of sperm whale sounds is examined. Sperm whales produce highly repetitive ''regular clicks'' at periods of about 0.2–2 s, and faster click trains in certain behavioral contexts. The following detection procedure was tested: a spectrogram was computed; values within a certain frequency band were summed; time windowing was applied; each windowed segment was autocorrelated; and the maximum of the autocorrelation within a certain periodicity range was chosen. This procedure was tested on sets of recordings containing sperm whale sounds and interfering sounds, both low-frequency recordings from autonomous hydrophones and high-frequency ones from towed hydrophone arrays. An optimization procedure iteratively varies detection parameters (spectrogram frame length and frequency range, window length, periodicity range, etc.). Performance of various sets of parameters was measured by setting a standard level of allowable missed calls, and the resulting optimium parameters are described. Performance is also compared to that of a neural network trained using the data sets. The method is also demonstrated for sounds of blue whales, minke whales, and seismic airguns. [Funding from ONR.]

10:55

3aAB9. The temporal characteristics of humpback whale songs.
Whitlow W. L. Au, Marc O. Lammers, Alison Stimpert (Hawaii Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744), and Michiel Schotten (Univ. of Groningen, The Netherlands)

Songs sung by male humpback whales consist of distinct, pulsed sounds that are designated as units. Units are produced in some sequence to form a phrase, a repeated set of phrases forms a theme, and repeated themes form a song. A song can last from minutes to hours. The songs of eight humpback whales were recorded with a vertical array of five hydrophones spaced 7 m apart with the array located within 100 m of the whales. At least seven distinct units were identified aurally from this data set obtained during the 2002 winter humpback whale session in Hawaii. Four distinct recurring themes were found in the songs, and for each whale at least two themes were recorded. The average duration of each unit sampled and the silent interval following the unit were determined in order to describe the temporal characteristics of the songs. From the data the temporal consistency and cadence control of unit production by each humpback whale and between whales were determined. Understanding the temporal and spectral characteristics of units within songs and how these units vary between whales could ultimately help in the design of computer algorithms to automatically identify individual whales.

11:10

3aAB10. Diel and spatial patterns in the singing behavior of humpback whales off Oahu, Hawaii. Marc O. Lammers, Alison K. Stimpert, Whitlow W. L. Au, and T. Aran Mooney (Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734)

The singing behavior of male humpback whales on the winter breeding grounds is still a poorly understood phenomenon. Previous work indicates that the chorusing levels of singing whales off west Maui are higher at night than during the day. However, the cause of this variation is not known. To investigate whether more whales sing at night or whether the same number simply move closer to near-shore recorders following sunset, the abundance and location of singing whales off Kaena Point, Oahu was examined. A bottom-moored recording system was used to establish that the same diel pattern observed off Maui also occurs off Oahu. The location of singing whales was determined by localizing singers along a preset transect track using a towed hydrophone array. More whales were found singing along the coastline examined at night than during the day. There was no indication of a shoreward migration. These results suggest that, at night, singing displays may be a more effective behavioral tactic than direct competition in a pod. Also, more singers were found along the northern part of the coastline, which is dominated by a shallow bank, indicating that singers were selective with respect to where they chose to sing.

11:25

3aAB11. Sound production patterns from humpback whales in a high latitude foraging area. Alison K. Stimpert (MMRP/Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, stimpert@hawaii.edu), David N. Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066), Kira L. Barton, Mark P. Johnson (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Marc O. Lammers, and Whitlow W. L. Au (MMRP/Hawaii Inst. of Marine Biol., Kailua, HI 96734)

Numerous studies have been conducted on humpback whale song, but substantially fewer have focused on the acoustic properties of non-song sound production (i.e., feeding and social sounds). Non-invasive digital acoustic recording tags (DTAGS) were attached to humpback whales on the western North Atlantics Great South Channel feeding grounds during July 2004. Acoustic records totaling 48.4 data hours from four of these attachments were aurally analyzed for temporal trends in whale signal production. A custom automatic detection function was also used to identify occurrences of specific signals and evaluate their temporal consistency. Patterns in sound usage varied by stage of foraging dive and by time of day. Amount of time with signals present was greater at the bottom of dives than during surface periods, indicating that sounds are probably related to foraging at depth. For the two tags that recorded at night, signals were present during a greater proportion of daylight hours than night hours. These results will be compared with previously published trends describing diel patterns in male humpback whale song chorusing on the breeding grounds. Data from the continuation of this research during the summer of 2005 will also be included.

Session 3aBBa

Biomedical Ultrasound/Bioresponse to Vibration, Physical Acoustics and Signal Processing in Acoustics: Medical Applications of Time Reversal Acoustics

Armen Sarvazyan, Chair

Artann Laboratories, Inc., 1457 Lower Ferry Rd., West Trenton, NJ 08618-1414

Chair's Introduction—8:05

Invited Papers

8:10

3aBBa1. Time reversal mirrors and invariants of the time reversal operator: Application to detection and imaging in biological media. Claire Prada and Mathias Fink (LOA, ESCPI, 10 rue Vauquelin, 75321 Paris Cedex 5, France, claire.prada-julia@espci.fr)

Acoustic time reversal mirrors have been developed in the Laboratoire Ondes et Acoustique for several years. In inhomogeneous media and especially in the presence of multiple scattering, these systems provide outstanding focusing properties. While used in pulse echo mode, the time reversal process can be iterated, leading to selective focusing on the most reflective point among a set of scatterers. This iteration is also used to enhance the echo from a defect like a kidney stone or a microcalcification in the breast. The iterative time reversal process is described by a matrix call the Time Reversal Operator (TRO). The eigenvectors of this matrix provide the invariants of the iterative time reversal process, their analysis is the base of the D.O.R.T. method. It allows the separation of the echoes from several scatterers in an inhomogeneous medium and the selective focusing on any of them. Furthermore, when an estimate of the medium's Green function is available, high-resolution and low-noise images can be achieved using the backpropagation of the dominant eigenvectors. The principle of acoustic time reversal mirrors and the D.O.R.T. method will be described through several examples.

8:35

3aBBa2. Advantages of time reversal acoustic focusing system in biomedical applications. Alexander Sutin (Davidson Labs., Stevens Inst. of Technol., 711 Hudson St., Hoboken, NJ 07030/Artann Labs., Lambertville, NJ 08530, asutin@artannlabs.com) and Armen Sarvazyan (Artann Labs., Lambertville, NJ 08530)

The development and biomedical applications of time reversal acoustics (TRA) systems for focusing and manipulating ultrasound beams are reviewed. The TRA focusing system (TRA FS) is capable to deliver ultrasound energy to the chosen region in highly inhomogeneous medium (including soft tissues and bones) with focusing efficacy hardly achievable using conventional phased array transmitters. TRA FS is able to focus and stir ultrasound beams in a 3-D volume using just a few piezoceramic transducers glued to the facets an aluminum block. Another advantage of TRA FS is its ability to produce pulses with arbitrary waveforms in a wide frequency band. A custom-designed compact multichannel TRA system operating in a wide frequency range from 0.01 to 10 MHz has been developed. Measurements of TRA field structure were conducted in a large variety of inhomogeneous tissue phantoms and *ex vivo* bones and soft tissues. Principles of TRA focusing optimization based on acoustical properties of the resonator material, parameters of the sonicated medium, and the coupling of the TRA resonator with the medium were developed and applied in the tested TRA systems. [Work was supported by NIH.]

9:00

3aBBa3. Time reversal for ultrasonic transcranial surgery and echographic imaging. Mickael Tanter, Jean-Francois Aubry, Francois Vignon, and Mathias Fink (Laboratoire Ondes et Acoustique, ESCPI, CNRS UMR 7587 10 rue Vauquelin 75005 Paris, France)

High-intensity focused ultrasound (HIFU) is able to induce non-invasively controlled and selective destruction of tissues by focusing ultrasonic beams within organs, analogous to a magnifying glass that concentrates enough sunlight to burn a hole in paper. The brain is an attractive organ in which to perform ultrasonic tissue ablation, but such an application has been hampered by the strong defocusing effect of the skull bone. Our group has been involved in this topic for several years, providing proofs of concept and proposing technological solutions to this problem. Thanks to a high-power time-reversal mirror, presented here are *in vivo* thermal lesions induced through the skull of 12 sheep. Thermal lesions were confirmed by T2-weighted magnetic resonance post-treatment images and histological examination. These results provide striking evidence that noninvasive ultrasound brain surgery is feasible. A recent approach for high-resolution brain ultrasonic imaging will also be discussed with a skull aberration correction technique based on twin arrays technology. The correction of transcranial ultrasonic images is implemented on a new generation of time-reversal mirrors relying on a fully programmable transmit and receive beamformer.

3a WED. AM

3aBBa4. Adaptive refocusing of dual-mode arrays. Emad Ebbini (Univ. of Minnesota Twin Cities, Minneapolis, MN 55455)

Advances in piezocomposite transducer technologies have allowed for the development of dual-mode ultrasound phased arrays (DMUAs) for noninvasive surgery. The imaging and therapeutic capabilities have been experimentally demonstrated. One major advantage of DMUAs is the use of image feedback for optimal refocusing to maximize the power deposition at the target while minimizing it to any critical structures in the intervening and surrounding tissues. The use of the same array elements in imaging and therapy allows for the development of image-based refocusing algorithm that employs a propagation operator from the array to the target point(s). In addition, a number of critical points can be identified from the DMUA image of the treatment region. The optimal complex driving signals to the array elements are obtained from a weighted solution employing an inverse spatial filter. The inverse is a cascade of a decorrelation step and a backpropagation step with a weighting matrix obtained from an auxiliary propagation operator from the array to the critical points. The theoretical formulation of the image-based optimal refocusing of DMUAs will be presented. Experimental results from refocusing experiments in the presence of ribs and the application to liver cancer treatment are presented and discussed.

Contributed Paper

9:50

3aBBa5. Acoustically induced tissue displacement for shear wave elasticity imaging using MRI. Kevin Haworth, Oliver Kripfgans, Derek Steele, Scott Swanson (Dept. of Radiol., 200 Zina Pitcher Pl., Kresge III R3322, Ann Arbor, MI 48109-0553), Alexander Sutin, and Armen Sarvazyan (ARTANN Labs. Inc., West Trenton, NJ 08618-1414)

Palpitation detects tissue abnormalities by exploiting the vast range of elastic properties found *in vivo*. The method is limited by tactile sensitivity and the inability to probe tissues at depth. Recent efforts seek to remove these limitation by developing a medical imaging modality based on ra- diation force shear wave excitation. Our approach uses an acoustic source to launch a shear wave in a tissue-mimicking phantom and MRI to record microscopic displacements. Gelatin (10% wt/vol) was used for the tissue-mimicking phantom. Results for *in situ* elasticity were obtained using an air-backed 10-cm-diam piezoelectric crystal. To correct for future *in vivo* beam aberrations, we also employ a high-pressure 1-bit time-reversal cav- ity. Frequency and pulse duration were selected to optimize the TRA sys- tem for acoustic output pressure. Shear wave displacements were recorded by MRI in 1-ms time increments in a complete basis that allowed for 3-D reconstruction and analysis. The Lamé coefficients are then derived from the shear wave velocity and attenuation.

WEDNESDAY MORNING, 19 OCTOBER 2005 CARVER ROOM, 10:30 TO 11:45 A.M.

Session 3aBBb

Biomedical Ultrasound/Bioresponse to Vibration: Ultrasound Contrast Agents

John S. Allen III, Chair
Univ. of Hawaii, Dept. of Mechanical Engineering, 2540 Dole St., Honolulu, HI 96822

Contributed Papers

10:30

3aBBb1. Theoretical analysis of a cell's oscillations in an acoustic field. John S. AllenIII (Dept. of Mech. Eng., Univ. of Hawaii-Manoa, 2540 Dole St., Honolulu, HI 96822) and Pavel Zinin (Hawaii Institue of Geophys. and Planetology, Univ. of Hawaii-Manoa, Honolulu, HI 96822)

The analysis and deformation of an individual cell in a high-frequency acoustic field is of fundamental interest for a variety of applications such as ultrasound cell separation and drug delivery. The oscillations of biologi- cal cells in a sound field are investigated using a shell model for the cell following an approach developed previously [Zinin, Ultrasonics, **30**, 26–34 (1992)]. The model accounts for the three components which com- prise the cell's motion: the internal fluid (cytoplasma), the cell membrane, and the surrounding fluid. The cell membrane whose thickness is small compared to the cell radius can be approximated as a thin elastic shell. The elastic properties of this shell together with the viscosities of the internal and external fluids determine the oscillations of the cell. The dipole oscil- lations of the cell depend on the surface area modulus and the maximum frequency for the relative change in cell area can be determined. More- over, the higher order oscillations starting with the quadrupole oscillations are governed by the shell's shear modulus. Induced stresses in bacteria cell membranes in the vicinity of an oscillating bubble are investigated and cell rupture with respect to these stresses is analyzed.

10:45

3aBBb2. Periodic nonlinear oscillations of a bubble between rigid parallel plates. Jianying Cui, Mark F. Hamilton, Preston S. Wilson, and Evgenia A. Zabolotskaya (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713-8029)

A linear analysis of forced bubble oscillations between rigid parallel plates that was presented at the previous ASA meeting is extended to include periodic nonlinear oscillations of the bubble. An equation of Rayleigh–Plesset form is obtained that satisfies the boundary conditions on the plates. Compressibility of the liquid is taken into account with radiation damping, and with time delays associated with acoustic reflec- tions from the plates. The equation is expanded to cubic order in the perturbation of the bubble radius, and the radius is expanded as a Fourier series in harmonics of the acoustic drive frequency. The result is a system of coupled nonlinear algebraic equations for the Fourier coefficients, the solutions of which describe the steady-state response of the bubble. Solu- tions were obtained numerically by iteration. An advantage of this frequency-domain approach is that the infinite sequence of pressure waves reflected from the plates can be summed analytically. Time histories and frequency responses are presented as functions of plate separation, the amplitude of the drive, and its frequency. This general approach can also be used to investigate moderate subharmonic generation. [Work supported by NIH Grant EB004047 and ARL IR&D funds.]

11:00

3aBBb3. Asymmetric oscillation of adherent targeted microbubbles.
Shukui Zhao, Katherine W. Ferrara, Dustin E. Kruse, and Paul A. Dayton
(Biomed. Eng. Dept., Univ. of California, 451 East Health Sci. Dr., Davis,
CA 95616, padayton@ucdavis.edu)

Recent development of molecular imaging with targeted ultrasound
contrast agents calls for a selective and sensitive detection method for
adherent targeted agents, which depends on our understanding of the re-
sponse of adherent agents to ultrasound interrogation. Here, asymmetric
oscillation of targeted microbubble contrast agents adherent to a cellulose
microtube was studied with a high-speed camera and compared to the
oscillation of free-floating microbubbles. The microbubbles were biotiny-
lated in order to bind to avidin on the inside of the microtube. A 1.5-MHz,
three-cycle ultrasound pulse was used to insonify the microbubbles. Ad-
herent microbubbles were observed to oscillate asymmetrically in the
plane normal to the boundary and nearly symmetrically in the plane par-
allel to the boundary, with a maximum expansion in each dimension sig-
nificantly smaller for bound than free microbubbles ($p<2E-5$ and $p
<0.015$, respectively). The observed oscillation period of adherent mi-
crobubbles was significantly different from that of free microbubbles ($p
<0.0015$) with the former closer to the period of the driving pulse. Asym-
metric oscillation led to shape instability and with sufficient acoustic pres-
sure a jet was observed traveling toward the boundary and fragmentation
of the adherent microbubbles was observed. A fraction of the resulting
fragments were observed to float freely in solution.

11:15

3aBBb4. Chirped-excitation contrast imaging.
Yang Sun, Dustin E.
Kruse, Paul A. Dayton, and Katherine W. Ferrara (Dept. of Biomed. Eng.,
Univ. of California, Davis, 451 East Health Sci. Dr., Davis, CA 95616,
kwferrara@ucdavis.edu)

Ultrasound contrast agent oscillation and echoes are investigated opti-
cally and acoustically following insonation with a high time-frequency
bandwidth pulse. Results of a numerical evaluation of the modified
Rayleigh–Plesset equation are used to predict the dynamics of a mi-
crobubble and compared with experimental observations. Two confocally
aligned transducers transmitting at 10 MHz and receiving at 5 MHz are

used, and a chirp waveform is transmitted with -6-dB bandwidth from
7.5 to 12.5 MHz. With a simple bandpass filter, subharmonic echo com-
ponents can be detected using the return from a single pulse, as the power
ratio of subharmonic components to fundamental components ranges up to
-7 dB. A high-speed camera system, which provides 2D frame images
and streak images, is used in optical experiments. From the optical obser-
vations of a single microbubble, nonlinear oscillation, destruction, and
radiation force are observed. For an initial diameter ranging from 1.5 to
5 μm, a maximum diameter below 8 μm is observed during insonation
with a 20-cycle 10-MHz linear chirp, with a peak negative pressure of 3.5
MPa. High-frequency pulse transmission may provide the opportunity to
detect contrast echoes resulting from a single pulse *in vivo*. [Work sup-
ported by NIH CA 76062.]

11:30

**3aBBb5. Destruction of contrast microbubbles used for ultrasound
imaging and drug delivery.**
Pankaj Jain, Kausik Sarkar (Mech. Eng.,
Univ. of Delaware, Newark, DE 19716), and Dhiman Chatterjee (IIT(M),
Chennai 600036, India)

Microbubble contrast agent destruction by ultrasound pulse is useful in
real-time blood-flow velocity measurement, stimulating arteriogenesis, or
targeted drug delivery. We investigated *in vitro* destruction of contrast
agent Definity (Bristol Meyer-Squibb Imaging, North Billerica, MA) by
measuring attenuation of ultrasound through it. The measurement is per-
formed with single-cycle bursts for varying pressure amplitudes at 50-,
100-, and 200-Hz pulse repetition frequencies (PRF). At low excitation
levels, the attenuation increases with time, indicating an increase in bubble
size due to ingress of dissolved air from the surrounding liquid. With
increased excitation levels, the attenuation level decreases with time, in-
dicating destruction of microbubbles. A critical pressure amplitude (1.2
MPa) was found for all three PRFs, below which there is no significant
bubble destruction. Above the critical excitation level, the rate of destruc-
tion depends on excitation levels. But, at high pressure amplitudes destruc-
tion becomes independent of excitation pressure amplitude. The results are
interpreted to identify two different mechanisms of bubble destruction by
its signature in attenuation, namely, slow dissolution by diffusion at inter-
mediate pressure amplitudes and catastrophic shell rupture at high ampli-
tudes.

WEDNESDAY MORNING, 19 OCTOBER 2005   CONRAD C, 8:55 A.M. TO 12:00 NOON

Session 3aEA

Engineering Acoustics: Acoustical Transducers, Arrays, Testing and Applications

Thomas R. Howarth, Chair
NAVSEA Newport, 1176 Howell St., Newport, RI 02841

Chair's Introduction—8:55

Contributed Papers

9:00

**3aEA1. Design and characterization of an anechoic aeroacoustic
facility.**
Jose Mathew, Chris Bahr, Bruce Carroll, Mark Sheplak, and Lou
Cattafesta (Univ. of Florida, Gainesville, FL 32611-6250)

The design and characterization of an anechoic wind tunnel facility at
the University of Florida are presented. A previously existing and ISO
3745 validated 100-Hz anechoic chamber is upgraded to incorporate an
open-jet anechoic wind tunnel facility suitable for airframe noise studies,
including swept-wing trailing edge studies. For suitable modeling of land-
ing conditions, a chord-based Reynolds number of 3 to 4 million is re-

quired. The wind tunnel is driven by a 224-kW centrifugal fan controlled
by a variable frequency drive. The test section measures 0.74 m (29) by
1.12 m (44) by 1.83 m (6 ft). The estimated maximum velocity attainable
in the test section is $\sim$ 76 m/s (250 ft/s). Preliminary measurements at 17
m/s indicate excellent flow uniformity and a turbulence intensity of 0.11%.
Background noise level measurements with an empty test section reveal an
overall SPL from 100 Hz 20 kHz of 49.9 dB, with a peak 1/3 octave-band
level of 46 dB at 100 Hz that decreases to 29.9 dB at 1 kHz. Facility
characterization experiments over a range of test section speeds are also
reported, along with the results of preliminary trailing edge noise experi-
ments.

3aEA2. A method to classify piezoelectric crystals. Changjiu Dang (GE Infrastructure Sensing, 1100 Technol. Park Dr., 1st Fl., Billerica, MA 01821, chang.dang@ge.com)

The properties of piezoelectric crystals used as the active elements in many ultrasonic transducers vary among the suppliers even if the crystals are claimed to be in the same category (say Navy I). A method of classifying the piezoelectric crystals is introduced. With this method, one can judge how similar to each other two types of crystals are by taking the crystal parameters as inputs. The parameters of a crystal include its density, velocity, mechanical Q, dissipation, frequency constants, piezoelectric constants, elastic constants, etc. They are transformed into the values around 1 to weaken the dominance of large parameter values. The transformed data form two attribute lists whose cross-correlation indicates their similarity. If two crystals are similar, their normalized cross-correlation coefficient is close to 1. Otherwise, it is close to 0. Based on this method, the available piezoelectric crystals can be further classified into different groups.

3aEA3. Grating lobe reduction in 1-D transducer arrays by driving a bar. Brian E. Anderson and W. Jack Hughes (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804, bea119@psu.edu)

This presentation investigates the effect of placing a finite-length bar between eight transducer elements and water in order to reduce undesirable grating lobes. The transducer elements are mounted to the bar, with a thin polyurethane layer between the elements and the bar, and force the bar to approximate the radiation of a continuous line source (i.e., no grating lobes). Reductions in the grating lobe with the insertion of a bar, compared to the unmodified array, have been found numerically and experimentally. In addition the transducer array can still be steered and amplitude shaded, and the bar geometry may be scaled according to the design frequency. Freely vibrating bending modes in the bar create undesirable additional radiation. The benefits and limitations of retrofitting existing array systems with a bar for grating lobe reduction, in frequency and steer angle, will be presented. The insertion of the bar provides a very inexpensive and easily implemented approach to reduce grating lobes without increasing the number of array elements. [Work sponsored by ONR Code 333, Dr. David Drumheller.]

3aEA4. A transfer matrix formulation of an electromechanical Helmholtz resonator. Fei Liu, Lou Cattafesta, Mark Sheplak (Dept. of Mech. and Aerosp. Eng., Univ. of Florida, Gainesville, FL 32611), Stephen Horowitz, and Toshi Nishida (Univ. of Florida, Gainesville, FL 32611)

An electromechanical Helmholtz resonator (EMHR) employing a compliant piezoelectric composite backplate is under development at the University of Florida. The acoustic impedance of the EMHR is made tunable and additional degrees of freedom (DOF) are added by coupling the piezoelectric composite diaphragm to a passive electrical shunt network. A formula of the transfer matrix is subsequently derived to predict the behavior of the EMHR. The theoretical aspects of electromechanical tuning are then reviewed and proof-of-concept experimental data are presented for a variety of resistive, capacitive, and inductive shunts in a normal-incidence plane wave tube (PWT). For resistive and capacitive shunt circuits, the resonator is shown to be a two DOF system, and short-circuit and open-circuit loads define the limits of the tuning range. There is approximately a 9% tuning limit under these conditions for the resonator configuration studied. An inductive shunt load results in a three DOF system with an enhanced tuning range of over 20% that is not restricted by the short-circuit and open-circuit limits.

3aEA5. Dual-backplate capacitive aeroacoustic microphone. David T. Martin, Jian Liu, Karthik Kadirvel, Robert M. Fox, Mark Sheplak, and Toshikazu Nishida (Univ. of Florida, Gainesville, FL 32611)

High-performance instrumentation grade microphones are a crucial component for aeroacoustic measurements. This paper presents the design, fabrication, and characterization of a MEMS dual-backplate capacitive microphone designed for aeroacoustic applications where small size, high dynamic range, and high bandwidth are key requirements. The microphone features a 230-micron radius, 2-micron air gaps, and a 118-dB dynamic range. Experimental results indicate a sensitivity of 0.282 mV/Pa, a linear response up to 160 dB, a noise floor of 42 dB at 1 kHz in a 1-Hz bin, and a measured resonant frequency of 230 kHz. A flat frequency response was measured up to 20 kHz in an $8.5 \times 8.5-$mm square rigid-walled acoustic plane-wave tube. The microphone was designed using an equivalent lumped-element model, simulated using finite-element methods, and fabricated using the SUMMiT V process at Sandia National Laboratories. Details on the acoustic and electrical characterization, fabrication process, and packaging will be described in the presentation.

3aEA6. High-intensity ultrasound radiation at over 10 MHz using hydrothermal polycrystalline lead zirconate titanate (PZT) thick film transducer. Mutsuo Ishikawa, Minoru Kurosawa (Interdisciplinary Grad. School of Sci. and Eng., Tokyo Inst. of Technol., G2-32, 4259 Nagatutamati, Midori-Ku, Yokohama, Kanagawa 226-8502, Japan), Tomohito Hasegawa, Akito Endoh, and Shinichi Takeuchi (Toin Univ. of Yokohama, Kanagawa 225-8502, Japan)

High intensity at over 10-MHz ultrasonic transducer was fabricated by hydrothermal deposition technique of lead zirconate titanate polycrystalline thick film (HPTF-PZT). The HPTF-PZT ultrasonic transducer had a 0.05-mm-thick PZT layer on a 0.05-mm-thick titanium substrate in an active area of 8 mm ~ 8 mm. The PZT thick film was deposited on a titanium substrate in an autoclave by solution including ions of Pb, Zr and Ti at 160 °C. A sound-pressure generation at thickness mode of the HPTF-PZT ultrasonic transducer was investigated in degassed water. This HPTF-PZT ultrasonic transducer could radiate sound pressures of 110 and 30 kPa at 20 and 10 MHz of ten sine waves, respectively. Cavitations and an acoustic streaming were generated when driving signal was 20-MHz continuous sine waves. Additionally, this ultrasonic transducer has the wide frequency response. Consequently, when driving signal was 10 MHz of 10-cycle rectangular waves from an odd order harmonics, the generated sound pressure was 54 kPa higher than 30 kPa at 10 MHz of 10-cycle sine waves. In fact, the HPTF-PZT ultrasonic transducer allows one to radiate high-intensity ultrasound at varied wave form of over 10 MHz.

3aEA7. Free-field reciprocity calibration of microphones in ultrasonic frequency range. Nourreddine Bouaoua (Dept. of Sound, Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany), Thomas Fedtke (Physikalisch-Technische Bundesanstalt, 38116 Braunschweig, Germany), and Volker Mellert (Oldenburg Univ., 26111 Oldenburg, Germany)

Exposure to airborne low-frequency ultrasound occurs in many industrial applications such as cleaning, welding plastics, measuring distances in buildings, and by using consumer devices such as camera rangefinders, automatic door openers, parametric ultrasound loudspeakers, etc. Exposure to ultrasound in air may be dangerous to the hearing and may have negative bio-effects on humans. Care should be taken in its use. In order to establish appropriate limits and to measure the output of ultrasound devices, there is a need to develop a sound-pressure standard in the frequency range from 20 kHz to about 160 kHz. A calibration project for quarter-inch microphones by the reciprocity method has been started in the PTB, and a new automated measurement setup for free-field reciprocity

calibration of quarter-inch microphones has been established. A procedure for a free-field reciprocity calibration of quarter-inch microphones in the frequency range 20 to 160 kHz is described. First results of free-field calibration of quarter-inch microphones are presented, giving the sensitivities and the repeatability of results. The effects of reflections and cross talk on the accuracy of the measurements will be explained.

11:00

3aEA8. Development of an acoustical energy harvester. Stephen B. Horowitz, Toshikazu Nishida, Louis N. CattafestaIII, and Mark Sheplak (231 MAE-A, Univ. of Florida, Gainesville, FL 32611)

This paper presents the development of one component of a miniature acoustical energy harvesting system that is designed to convert acoustical energy into electrical energy. This component consists of a composite circular membrane containing a circular piezoelectric ring near the clamped boundary. An acoustic pressure fluctuation leads to a deflection of the membrane. This deflection generates a strain concentrated near the clamped boundary. The piezoelectric material, strategically placed in this region of high strain, converts the strain into an electrical voltage that can then be utilized in an electrical circuit or stored in a battery. The energy harvester was designed by combining a linear piezoelectric composite plate model with a lumped-element, electroacoustic model. Characterization was performed via acoustic excitation in a plane-wave tube. Preliminary results indicate a maximum generated output voltage of 18.2 mV across a $78.3-k\Omega$ resistive load, resulting in 4.2 μW of generated power for an incident sound pressure level of 153.4 dB.

11:15

3aEA9. Sonic gas analyzer for microbiological metabolic measurements. Miguel A. Horta and Steven Garrett (Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804)

A differential open-pipe resonator was built to track changes in gas-mixture concentration. A single miniature loudspeaker simultaneously drives two adjacent ducts at resonance and 180 deg out of phase. The resonant frequency is tracked with a phase-locked loop, using the difference signal from two electret microphones whose sensitivities are balanced by adjustment of the preamplifier gains to provide common-mode rejection of extraneous noise sources (for example, a magnetic stirrer) within the bioreactor. A small change of the gas concentration produces a proportional change of the driving frequency for a given binary mix of gases. This sensor is designed to measure the production of hydrogen or methane from metabolic processes of anaerobic bacteria. Results from an initial set of experiments using helium injection and hydrogen release from a HCl+Zn reaction will be presented. [For Engineering Acoustics Best Student Paper Award.]

11:30

3aEA10. Zero-g acoustic fire suppression system. Dmitriy Plaks, Elizabeth Nelson, Nesha Hyatt, James Espinosa (Phys. Dept., Univ. of West Georgia, 1601 Maple St., Carrollton, GA 30118), Zade Coley, Cathy Tran, and Ben de Mayo (Univ. of West Georgia, Carrollton, GA 30118)

The objective is to study the effects of acoustics on a flame in microgravity. This research is meant to provide a new approach to reducing and extinguishing a combustion reaction in space (where a conventional fire extinguisher is hazardous). The group proposes to test the following hypotheses: (1) as sound intensity increases, so does the magnitude of the effect on the flame; (2) there is one optimal frequency for maximizing the effects of sound waves on a flame; (3) homogeneous flames (found only in microgravity) can be affected in a single area separate from others; (4) a sustained pulse of sound, rather than a single, brief pulse, can be used to extinguish a flame. The group will describe the experimental apparatus in detail, which was flown aboard a NASA C-9B Aircraft through their Reduced Gravity Student Flight Opportunities Program (RGSFOP), and present the findings and compare data obtained in a 1-g environment with data collected in a microgravity environment. [Thanks to the following organizations for their support: Kicker Audio, Georgia Space Grant Consortium, Siemens, FLIR Systems, PCB Piezotronics, UWG Honors College, and NASAs Reduced Gravity Student Flight Opportunities Program.]

11:45

3aEA11. A noncontact transportation in water using ultrasonic traveling waves. Shinfuku Nomura (Dept. Mech., Ehime Univ, Matsuyama, 790, Japan, nomu@eng.ehime-u.ac.jp), Thomas J. Matula (Univ. of Washington, Seattle, WA 98105), Jun Satonobu (Hiroshima Inst. of Tech., 731, Hiroshima, Japan), and Lawrence A. Crum (Univ. of Washington, Seattle, WA 98105)

A noncontact transport experiment in water using a traveling-wave-type linear motor was investigated. Acrylic disks were used as the floating objects. A vibrating plate was enclosed in acrylic plates, and a water tank was made that would allow the vibrating plate to be placed on the bottom. In order to propagate the traveling wave, two ultrasonic transducers were attached at both ends to the bottom of the vibrating plate. One side was used as the wave-sending side and the other side was used as the wave-receiving side. Comparing the transport experiments conducted in water with those conducted in air, the transport velocity becomes faster for floating transport in water than for floating transport in air. The occurrence of cavitation bubbles acts as a resistive force on the movement of the object being transported, and causes the transport velocity to be reduced. Transport velocity depends on the height of the water. If the height of the water surface is too shallow, the water surface freely deforms and the water surface forms bumpy standing waves, making it difficult for the object to be transported.

Session 3aED

Education in Acoustics: Acoustics Demonstrations II

Thomas D. Rossing, Chair
Northern Illinois Univ., Physics Dept., DeKalb, IL 60115

Invited Papers

9:00

3aED1. The ping-pong cannon demonstration: Optical studies and numerical simulation. Richard Peterson and Keith Stein (Dept. of Phys., Bethel Univ., St. Paul, MN 55112, petric@bethel.edu)

This presentation describes the use of laser pulse photography, optical timing, pulsed Schlieren, and heterodyne interferometry to look more closely at the fluid dynamics of a recently popular lecture demonstration—the so-called "ping-pong cannon." Optical diagnostic techniques have been applied to two types of these cannons, and led to greater knowledge of the kinematics of the accelerating ball, along with some insight into the exit mechanism and subsequent target interactions. Also will be described how a 1D numerical simulation allows visualization of shock wave formation within the tube during ball acceleration. Solutions of the Euler equation are obtained using the method of space-time finite elements, while the ball is tracked as an interface in the compressible fluid mesh. [Work supported in part by the MN NASA Space Grant and the Carlsen-Lewis Endowment of Bethel University.]

9:20

3aED2. Acoustic levitation. Uwe J. Hansen (Indiana State Univ., Terre Haute, IN 47809)

A speaker, driven by an amplified audio signal is used to set up a standing wave in a 3b-ft-long, 4-in.-diam transparent tube. Initially the tube is oriented horizontally, and Styrofoam packing peanuts accumulate near the pressure nodes. When the tube is turned to a position with the axis oriented vertically, the peanuts drop slightly, until the gravitational force on the peanuts is balanced by the force due to the sound pressure, at which point levitation is observed. Sound-pressure level measurements are used to map the air column normal mode pattern. Similarly, standing waves are established between an ultrasonic horn and a metal reflector and millimeter size Styrofoam balls are levitated.

9:40

3aED3. Three demonstrations of degeneracy lifting. Andrew Morrison (Dept. of Phys., Northern Illinois Univ., DeKalb, IL 60115, morrison@physics.niu.edu)

Two normal modes of vibration of a single object having exactly the same frequency are said to be degenerate modes. Certain simple systems, such as a circular membrane, have predictable degenerate modes. A lack of isotropy in the material or a geometric asymmetry can separate the frequencies and lift the degeneracy. Demonstration of this effect is easily accomplished in the classroom. Three methods of showing the effect are presented using a handbell, a short metal rod, and a coffee mug.

10:00

3aED4. Auditory demonstrations simulating Mayan architecture. David Lubman (Acoust. Consultant, 14301 Middletown Ln., Westminster, CA 92683)

Fascination with the ancient temples and ball court at Chichen Itza provide rich opportunities for science education. Children of all ages are delighted to learn that the sound of handclaps scattered from long temple staircases are transformed into bird chirps. Their engagement in such seemingly magical phenomena provides magic moments for teaching acoustical principals, including the picket-fence effect (PFE). PFE transforms impulsive sounds scattered from spatially periodic structures into tonal sounds. PFE is demonstrated with a computer possessing a sound card and a simple sound editing program. The inverse relationship between tonal frequency and the time interval between periodic impulses is easily demonstrated. The number of impulses needed to produce an audible tone is easily demonstrated and compared with the number of steps on the staircase. Transformation of audible tones into downward-gliding chirps is simulated by monotonically increasing the time between impulses. The Great Ball Court also provides opportunities for acoustical demonstration. Observers clapping their hands while standing between the long, tall, and parallel walls of the playing field marvel at the profound flutter echo heard for about 1.5 s. The flutter echo sonogram demonstrates the speed of sound and frequency-selective atmospheric attenuation.

10:20

3aED5. Demonstration of nonlinear effects in acoustic landmine experiments using a clamped-plate soil oscillator. Murray S. Korman and Emilia Bond (Dept. of Phys., U. S. Naval Acad., Annapolis, MD 21402)

Current nonlinear experiments involving the detection of plastic landmines using acoustic-to-seismic coupling have been developed from Sabatier's (linear) and Donskoy's (nonlinear) earlier methods. A laboratory apparatus called the soil-plate oscillator has been developed at the National Center for Physical Acoustics, and later at the U.S. Naval Academy, to model acoustic mine detection.

The apparatus consists of a thick-walled cylinder filled with sifted homogeneous soil resting on a thin elastic plate that is clamped to the bottom of the column. It represents a good simplified physical model for VS 1.6 and VS 2.2 inert anti-tank plastic buried landmines. Using a loudspeaker (located over the soil) that is driven by a swept sinusoid, tuning curve experiments are performed. The vibration amplitude versus frequency is measured on a swept spectrum analyzer using an accelerometer located on the soil-air interface or under the plate. The backbone curve shows a linear decrease in peak frequency versus increasing amplitude. A two-tone test experiment is performed using two loudspeakers generating acoustic frequencies (closely spaced on either side of resonance, typically ~100 Hz). A rich vibration spectrum of combination frequency tones (along with the primaries) is observed which is characteristic of actual nonlinear detection schemes.

10:40

3aED6. Report on a workshop for construction of acoustic demonstration apparatus. Matthew E. Poese (Penn State Appl. Res. Lab., P.O. Box 30, State College, PA 16804, poese@psu.edu and Dept. of Psych., Boston Univ., Boston, MA 02115)

A workshop where over 20 participants built two acoustical demonstration devices was organized by the author and sponsored by the Central Pennsylvania Chapter of the Acoustical Society of America. This talk will show construction details of the cavitation stick and the thermoacoustic engine (''acoustic laser'') that graduate and undergraduate students, faculty and community members built at the workshop as well as demonstrate the finished products. In order that participants can finish the workshop in an evening, the level of preparation required for such an event will be discussed as will the shop facility used in this case.

11:00

3aED7. Hey kid! Wanna build a loudspeaker? The first one's free. Steven Garrett (Grad. Program in Acoust., Penn State Univ., PO Box 30, State College, PA 16804-0030)

In 2000, Penn State University instituted a First Year Seminar (FYS) requirement for every entering undergraduate student. This paper describes a hands-on FYS on audio engineering that has each freshman, in a class of 20, construct and test a two-way loudspeaker system during eight 2-hour meetings. Time and resource constraints dictated that the speaker system must be assembled using only hand tools and characterized using only an oscillator and digital multimeters. Due to limitations on the funds made available for each FYS by the College of Engineering, the total cost of entire system could not exceed $65/side. Each student is provided with a woofer, tweeter, crossover components, and enclosure parts to build one speaker system. The students are offered the option of purchasing a second set of parts for $65 so that they can complete the course with a stereo pair. Ninety percent of the students exercise the stereo option. This presentation will describe the speaker system, using an enclosure made primarily from PVC plumbing parts, and the four laboratory exercises that the students perform and write up that are designed to introduce basic engineering concepts including graphing, electrical impedance, resonance, transfer functions, mechanical and gas stiffness, and nondestructive parameter measurement.

WEDNESDAY MORNING, 19 OCTOBER 2005　　　　　　　　　　　CONRAD D, 8:00 TO 11:00 A.M.

Session 3aMU

Musical Acoustics: General Topics in Musical Acoustics

Ian M. Lindevald, Chair
Truman State Univ., Div. of Science/Physics, Kirksville, MO 63501

Contributed Papers

8:00

3aMU1. Violin bridge-hill to radiation: The energy trail. George Bissinger (Phys. Dept., East Carolina Univ., Greenville, NC 27858)

Simultaneous experimental modal and acoustical analysis in an anechoic chamber was used to track energy flow through quality-rated violins to clarify the link between the mechanical motions at a driving point on the G-corner of the bridge, and subsequent radiation from the violin. The 0–4-kHz range covered included a broad hump near 2.4 kHz in driving point mobility generally labeled the bridge-hill and thought to be related to violin quality and radiativity (and sound). The bridge-hill enhancement showed up successively in the averaged mobility of the bridge, at the bridge feet, and over the corpus; a corresponding peak was also seen in the radiativity, indicating that enhanced radiation accompanied enhanced bridge motion. All bridge motions except low-frequency rigid-body motions were complex, and generally showed some semblance of rocking motion at higher frequencies. No evidence was observed for any bridge resonance separate from string or corpus resonances, although most bridges showed a complex squat mode near 0.8 kHz (not accompa-

nied by enhanced radiation) that is likely an out-of-plane bending mode falling far below the rocking mode. Bridge-hill magnitude was not found to correlate well with violin quality, or averaged corpus mobility, or averaged radiativity.

8:15

3aMU2. Guitar radiation—resonant and nonresonant. Richard Lyon (RH Lyon Corp, 60 Prentiss Ln., Belmont, MA 02478)

The sound radiation from a guitar contains about 50 fundamental semi-tones and about 200 harmonics. The resonant modes of the guitar body are insufficient in number and of insufficient bandwidth to support all of these frequencies, so some frequencies are coincident with resonances and some are not. It is the purpose of this paper to discuss the role of resonances as they occur in the guitar top, sides, and back in affecting the radiation. Conjectures about the role of the nonresonant radiation in the attack of the sounds and the hearing mechanism on the perception of the sound are presented in a tentative manner.

8:30

3aMU3. Sound decay of notes from acoustic guitars. Erika Galazen, Joni Nordberg, and Thomas M. Huber (Phys. Dept., Gustavus Adolphus College, St. Peter, MN 56082, huber@gac.edu)

The acoustic guitar produces tones by transferring energy from the strings, through the bridge to the top plate, back, and air cavity of the guitar. The vibrations are ultimately radiated into the air as sound. The air-cavity and body resonances of the guitar play an important role in both the tone and the sustain (the time it takes notes to decay) produced by the guitar. To study the relationship between resonances of the guitar and the sustain of notes, the resonance frequencies were measured using a mechanical shaker attached to the body of the guitar and laser Doppler vibrometer to measure its vibration. A string was tuned to different frequencies and plucked. The decay of the note was measured with an electromagnetic pickup that measured the vibration of the string, a vibrometer to measure vibration of the top plate, and microphones located inside and outside the guitar. As expected, when the fundamental frequency of the string was near one of the resonances of the guitar, the decay rate was faster (shorter sustain) than when the string was between resonances. The relationship between the decay rates of the different parts of the system will also be discussed.

8:45

3aMU4. The low-frequency resonances of harp soundboxes. Chris Waltham and Andrzej Kotlicki (Dept. of Phys. & Astron., Univ. of British Columbia, Vancouver BC, V6T 1Z1, Canada, cew@phas.ubc.ca)

The harp soundbox is a thin shell coupled directly to the strings. The sound board at the front is usually solid, but at the back are holes for access to the tethering points of the strings. As with other plucked instruments, the most important resonances for sound production tend to be those at low frequency, and these often involve the coupling of the whole soundbox to the air inside, and also to other parts of the harp structure. The low-frequency resonant structure of small folk harps and a full-size concert harp has been investigated, and results will be presented here. The instruments used, a tap hammer and accelerometer, were made in the laboratory from inexpensive materials, and their manufacture and calibration will also be discussed.

9:00

3aMU5. Input impedance of brass instruments from velocity measurement. Daniel O. Ludwigsen (Kettering Univ., 1700 W. Third Ave., Flint, MI 48504)

A velocity sensor known as the Microflown measures particle velocity from a difference in temperature between two MEMS-scale wires. With a small precision microphone in a package the size of a matchstick, simultaneous measurement of particle velocity and pressure can be accomplished in a tiny space such as the mouthpiece of a brass instrument. Traditional measurements of input impedance rely on a constant flow provided by a capillary tube or feedback loop control of the driver. This velocity sensor eliminates these technical requirements. The apparatus and calibration procedures will be described, and results of measurements of several instruments will be presented. In an easily used device, this approach could benefit instrument designers, makers, and repair technicians.

9:15

3aMU6. An air-coupled actuator array for active modal control of timpani. Douglas Rollow (Digidesign, Inc., 2001 Junipero Blvd., Daly City, CA 94014), Victor W. Sparrow and David C. Swanson (The Penn State Univ., University Park, PA 16801)

The timbral characteristics of kettledrums can be described by a modal formulation of the vibration of a thin, air-loaded membrane. Modification of the modal time history can be brought about with the use of a control system which has independent influence on each structural mode. By replacing the usual kettle with a shallow chamber and a planar array of piston sources, a modal controller is created when driving the sources in

appropriate linear combinations. A theoretical formulation of active control of structural vibration by means of fluid-coupled actuators is expressed, and a Boundary Element simulation provides insight to the coupled modes, independence of control, and constraints due to the geometry of the chamber. Advantages and limitations of this type of control source to general problems in actively controlled musical instruments are explored.

9:30–9:45   Break

9:45

3aMU7. Airborne ultrasound enters the ear through the eyes. Martin Lenhardt (Depts. of Biomed. Eng., Otolaryngol., Emergency Medicine, Virginia Commonwealth Univ., Richmond, VA 23298-0168)

Musical spectrum above 20 000 Hz has been demonstrated to influence human judgments and physiology. Moreover airborne ultrasonic noise has been implicated in hearing loss, tinnitus, and other subjective effects such as headaches and fullness in the ear. Contact ultrasound, i.e., with a transducer affixed to the skin of the head/neck, is audible; assumed by bone conduction. However, lightly touching the soft tissues of the head, avoiding bone, can also produce audibility. When contact ultrasound is applied to the head, energy from 25 to ~60 kHz can be recorded from the closed eyelid, with care to avoid sensor contact with the orbit. If the same frequency band of noise is passed through a transducer in from of the eye, with just air coupling, the same response is again recordable on the head. An acrylic barrier between the eye and the transducer eliminates the response. Once airborne ultrasound exceeds the impedance mismatch of the eye it readily propagates through the soft tissues of the eye and brain via one of the fluid windows (end lymphatic, perilymphatic or vascular) to the cochlea. The eye fenestration explains how people can detect airborne ultrasonic components in music and develop ear effects from airborne ultrasonic noise.

10:00

3aMU8. Separation of piano tones into tonal and broadband components. Martin Keane (Acoust. Res. Ctr., Sch. of Architecture, Univ. of Auckland, Private Bag 92019, Auckland, New Zealand)

Few studies have proposed simple objective methods to evaluate the quality of piano tones. Recordings of both sound and vibration of six pianos (two grands and four uprights) were analyzed. The results suggest an objective measure for piano tone quality. A method is given for splitting recorded tones into tonal and broadband components. The broadband vibration was found to be more intense in the upright pianos than in the grand pianos, this is due to differences in the construction of the keybed. There was no such difference found in the tonal component. These results suggest the hypothesis that a smaller broadband component correlates with higher quality in piano tones.

10:15

3aMU9. Polyphonic pitch detection and instrument separation. Mert Bay and James W. Beauchamp (School of Music and Dept. of Elec. and Computer Eng., Univ. of Illinois at Urbana-Champaign, 2136 Music Bldg., 1114 W. Nevada, Urbana, IL 61801, mertbay@uiuc.edu)

An algorithm for polyphonic pitch detection and musical instrument separation is presented. Each instrument is represented as a time-varying harmonic series. Spectral information is obtained from a monaural input signal using a spectral peak tracking method. Fundamental frequencies (F0s) for each time frame are estimated from the spectral data using an Expectation Maximization (EM) algorithm with a Gaussian mixture model representing the harmonic series. The method first estimates the most predominant F0, suppresses its series in the input, and then the EM algorithm is run iteratively to estimate each next F0. Collisions between instrument harmonics, which frequently occur, are predicted from the estimated F0s, and the resulting corrupted harmonics are ignored. The amplitudes of these corrupted harmonics are replaced by harmonics taken from a library of spectral envelopes for different instruments, where the spectrum which

most closely matches the important characteristics of each extracted spectrum is chosen. Finally, each voice is separately resynthesized by additive synthesis. This algorithm is demonstrated for a trio piece that consists of 3 different instruments.

10:30

3aMU10. Pitch circularity produced by varying the amplitudes of odd and even harmonics. Diana Deutsch, Kevin Dooley (Dept. of Psych., UCSD, La Jolla, CA 92093), Shlomo Dubnov, Trevor Henthorn (UCSD, La Jolla, CA 92093), and Alex Wurden (UCSD, La Jolla, CA 92093)

Pitch circularities have been produced using tones whose components stand in octave relation (Shepard, 1964; Risset, 1969). This paper describes two circular pitch scales produced by a new algorithm. Each scale consists of a bank of harmonic complex tones, each tone 500 ms in duration. As the scale descends in semitone steps, the relative amplitudes of the odd-numbered harmonics decrease in 5-dB steps, so that at the bottom of the scale they are 55 dB lower than the even-numbered harmonics. The entire bank of tones is then low-pass filtered. The lowest tone of such a scale is therefore heard as though its fundamental frequency were displaced up an octave. For each scale all possible (i.e., 132) ordered tone pairs were presented, and 15 subjects judged for each pair whether the second tone was higher or lower than the first. The data derived from these pairwise comparisons were subjected to Kruskal's nonmetric multidimensional scaling (MDS). For both scales, two-dimensional plots yielded approximately circular configurations; Stress1 values for the two scales were 0.016 and 0.034. When the tones were presented in semitone steps, the impression of infinitely ascending and descending scales was obtained. These circular scales are demonstrated with sound examples.

10:45

3aMU11. Seeing the B-A-C-H motif. Palmyra Catravas (ECE Dept., Union College, Schenectady, NY 12308, catravap@union.edu)

Musical compositions can be thought of as complex, multidimensional data sets. Compositions based on the B-A-C-H motif (a four-note motif of the pitches of the last name of Johann Sebastian Bach) span several centuries of evolving compositional styles and provide an intriguing set for analysis since they contain a common feature, the motif, buried in dissimilar contexts. We will present analyses which highlight the content of this unusual set of pieces, with emphasis on visual display of information.

WEDNESDAY MORNING, 19 OCTOBER 2005

SALONS B AND C, 8:15 TO 9:15 A.M.

Session 3aNCa

NOISE-CON: Plenary: What Do We Know About Noise in Hospitals

Patricia Davies, Chair

Purdue Univ., School Mechanical Engineering, Ray W. Herrick Labs., 140 S. Intramural Dr., West Lafayette, IN 47907-2031

Chair's Introduction—8:15

8:20

3aNCa1. What do we know about noise in hospitals. James E. West and Ilene Busch-Vishniac (Elec. and Computer Engineer., Johns Hopkins Univ., Barton 105, 3400 N. Charles, St., Baltimore, MD 21218)

Little is known about noise levels inside hospitals and its effect on healing, safety, staff, and doctors. Many independent studies of noise levels in a wide range of hospital venues in many countries have been reported in the literature authored mainly by physicians and nurses. The World Health Organization (WHO) has set guidelines for noise levels in hospitals, but none of the reported studies meet WHO guidelines. In most cases noise levels are more then 15 dB A-weighted higher then those specified by WHO guidelines. Since 1960 the average noise levels in hospitals has increased an average of 0.38 dB per year during daytime hours and 0.42 dB during the night. This talk reviews the state of the art on interior hospital noise control and the remaining challenging issues. Equivalent sound pressure levels as a function of location, frequency and time of day were measured in five different venues at the Johns Hopkins Hospital in Baltimore Maryland. Results of our measurements, which confirm the trends seen in prior studies, will be used to describe patterns of hospital interior noise and avenues ripe for further investigation.

Session 3aNCb

NOISE-CON and Noise: Active Noise Control

Scott D. Sommerfeldt, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, N281A ESC, Provo, UT 84602-4673

David C. Copley, Cochair
Caterpillar Inc., Technical Ctr., P.O. Box 1875, Peoria, IL 61656

9:45

3aNCb1. Decentralized active feedback control approach for vibration and noise reduction through an aircraft fuselage. Gopal Mathur (Boeing, Huntington Beach, CA) Christopher R. Fuller and J. Carneal (Virginia Polytechnic and State Univ., Blacksburg, VA 24061)

Active noise control has been commonly implemented with fully coupled, feed-forward approach, which employs reference sensors on the fuselage/source and error microphones in the radiated field. Implementation limitations include: complexity, delay/causality, and use of microphone error sensors. An alternative approach of feedback control utilizing sensors located on each panel may overcome most of these problems. The sensors on each panel should estimate the sound radiation from the panel and also should de-couple the control from the neighbor actuator signals by primarily observing the control signals on the panel to which it is attached. With this sensing arrangement, multiple, independent single input single output feedback loops applied to each panel in conjunction with an actuator and sound radiation sensor can be utilized. Causality problems are overcome by utilizing feedback control approach with sensors located on the panel as feedback sensors. This paper will present results from laboratory tests, conducted on an aircraft fuselage, which showed that a multiple SISO feedback control approach can provide global reductions of broadband interior noise. These tests demonstrated that the actuator, sensor, and control hardware technology is mature and can be integrated into a reliable, compact system.

10:05

3aNCb2. Stability conditions in decentralized active vibration control of a flexible plate. P. Micheau and A. Berry (G.A.U.S, Mech. Eng. Dept., Universite de Sherbrooke, Sherbrooke, PQ, J1K 2R1, Canada)

This paper presents the analysis of decentralized active control of periodic panel vibration using multiple pairs combining PZT actuators and PVDF sensors distributed (not necessarily collocated) on the panel. Frequency-domain gradient and Newton-Raphson adaptation of decentralized control were analyzed, both in terms of performance and stability conditions. The stabilization of the closed loop by adjusting the convergence coefficient was especially investigated through an adjustable control effort term in the quadratic cost function. Various necessary or sufficient stability conditions derived from the transfer matrix were obtained and ranked from the most restrictive to the less restrictive condition. One of them, the strictly positive real condition of the transfer matrix, appears to be very useful. This stability condition is obtained when the open-loop system is dissipative. Simulation and experimental results are presented for a simply supported panel with 4 PZT-PVDF pairs distributed on the panel. The experimental results were obtained with a decentralized harmonic controller demonstrated equivalent to any narrow-band disturbance control methods. Decentralized vibration control is shown to be highly dependent of the actuator and sensor configurations, but can be as effective as a fully centralized control even when the transfer matrix is not diagonal-dominant.

10:25

3aNCb3. Advantages and challenges of distributed active vibro-acoustic control. Kenneth D. Frampton (Dept. of Mech. Eng., Vanderbilt Univ., VU Station B 351592, Nashville, TN 37235)

As active control technologies reach their performance limits in large scale systems, many investigators have looked toward decentralized control as a means of expanding the application horizons. Decentralized control is defined here as numerous independent controllers operating on a single system. These decentralized approaches have been shown to be effective, but not as effective as traditional centralized control. In an effort to achieve control performance approaching centralized control while maintaining the scalability benefits of decentralized control, the use of distributed control is considered. Distributed control consists of numerous control processors operating on a system that are capable of communicating with each other over a network. This work discusses the application of distributed active control to vibroacoustic problems. Key elements in distributed control systems will be presented along with the state of each of the key technologies involved. Of particular interest are the limitations in enabling technologies that limit the application of distributed control: namely real-time network communications; distributed control algorithms and design tools; software infrastructure for system management; and other factors. Results will demonstrate that distributed control can perform nearly as well as centralized control, but that it can also be "scaled up" for use in large complex systems.

10:45

3aNCb4. Development of transparent thin film speaker and its application for active sound transmission control through windows. X. Yu, R. Rajamani, K. A. Stelson, and T. Cui (Dept. of Mech. Eng., Univ. of Minnesota, 111 Church St. SE, Minneapolis, MN 55455)

Homes close to airports and highways in major metropolitan areas suffer from serious noise problems. Typically, windows constitute the primary path through which noise enters a home. As an effective method for attenuating low frequency noise, active noise cancellation approach has drawn increasing interest. A small voice coil actuator can be used to vibrate a glass panel and achieve significant reduction in noise transmission at its location. However, global noise cancellation over the entire panel can not be achieved with such a single point actuator. Using multiple voice coil actuators is also not practical, since several actuators on a window-pane would destroy the aesthetics of the window. This necessitates the development of thin film actuators that can provide distributed canceling sound over the entire surface of large sized glass panels. The need of transparence for the windows application poses a great challenge to the development of such thin film actuators. In this paper, transparent thin film speakers are developed, based on the use of carbon nanotubes for conductive films that are coated on polzerized PVDF films. Preliminary experimental results show that the transparent thin film speaker has the potential to be a promising solution for the ANC windows.

1950   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1950**

11:05

3aNCb5. Energy density active noise control in an earthmoving machine cab. David C. Copley (Caterpillar Inc., Tech. Ctr. Bldg. E, P.O. Box 1875, Peoria, IL 61656), Ben Faber, and Scott Sommerfeldt (Brigham Young Univ., Provo, UT 584602)

Conventional active noise control (ANC) systems attempt to minimize squared pressure at an error sensor (or sensors) to achieve control, and often result in a relatively small zone of control. Considering a practical application in a vehicle, this approach is acceptable provided the operator's head remains within the control zone. In the case of an earthmoving machine, an operator's head regularly moves about the cab, either because of the motion of the machine during operation or as the operator moves about to perform various tasks. By minimizing energy density (ED), instead of squared-pressure, a potentially larger control zone may be achieved. Active control of the sound field within an earthmoving machine cab was attempted using the ED technique. The scope was limited to controlling the engine firing frequency tonal within the cab with the machine running, but in a static condition (engine idle). Frequency response and spatial results show the method achieved nearly global control within the cab and effectively reduced the A-weighted sound-pressure level by a few decibels, despite only reducing a single low frequency tone. In this sense, active control using ED has the potential to be more suitable to earthmoving machine applications than squared-pressure techniques.

11:25

3aNCb6. Stability analysis of the active control of self-sustained pressure fluctuations due to flow over a cavity. Philippe Micheau, Ludovic Chatellier, Janick Laumonier, and Yves Gervais (LEA, Bat. K, 40 av Recteur Pineau, 86000 Poitiers, France, Philippe.Micheau@USherbrooke.ca)

The addressed problem is to actively control the self-sustained pressure fluctuations due to low Mach number turbulent flow ($M < 0.1$) over a large, shallow cavity in a wing-shaped model. For the studied configuration, the sound pressure is generated by the unstability of the mixing layer [L. Chatellier, J. Laumonier, and Y. Gervais, ''Theoretical and experimental investigations of low Mach number turbulent cavity flows,'' Exp. Fluids **36**, 728–740 (2004)]. An original active device, a vibrating surface located at the trailing edge, and an adjustable narrow-band controller are used to control the self-oscillation: to synchronize and amplify it in open loop, and to attenuate it in closed loop. Experimental data were obtained in the test section of an Eiffel-type wind tunnel. For a free stream at 31 m/s, the SPL measured at 397 Hz in the cavity is attenuated from 118 dB to the ground level at 105 dB with the active control. A complete analysis of the system based on the describing function analysis is presented. The experimental results favorably compare with the prediction of the model; moreover, they allows identification of some parameters, such as the equivalent volume velocity of the unstable mixing layer and its coupling gain with the pressure fluctuation.

WEDNESDAY MORNING, 19 OCTOBER 2005   SALON B, 9:45 TO 11:25 A.M.

Session 3aNCc

NOISE-CON and Noise: Product Noise and Vibration Control—Case Studies II

Raymond Cheng, Chair

Fleetguard Inc., Engineering and Technology, 1801 Highway 51 138, Stoughton, WI 53589-0428

9:45

3aNCc1. Hydrofoil vibration and noise reduction with leading edge isolation. Timothy A. Brungart, Eric C. Myer, Dean E. Capone, and Robert L. Campbell (Appl. Res. Lab., The Penn. State Univ., P.O. Box 30, State College, PA 16804-0030)

A technique for reducing the vibration and noise from hydrofoils subject to unsteady hydrodynamic loads was demonstrated experimentally. Unsteady loads are generated when hydrofoils interact with approach-flow disturbances such as hull boundary layer turbulence or wakes from upstream rotors. Since the unsteady loads are known to be concentrated in the vicinity of the leading edge, a single stage vibration isolation mount was incorporated into a hydrofoil at its 20% chord location to inhibit the leading edge-generated unsteady loads from being transmitted to the remainder of the foil and any structures coupled to it. The hydrofoil was tested in a water tunnel facility with a wake generator placed upstream to produce the approach-flow disturbances. The reduction in the loading on the portion of the hydrofoil isolated from the leading edge was inferred from vibration measurements. Reductions of 5 to 10 dB were demonstrated.

10:05

3aNCc2. Measurement of underwater noise of the new National Oceanic and Atmospheric Administration (NOAA) Ship OSCAR DYSON. Michael Bahtiarian (Noise Control Eng., 799 Middlesex Tpke., Billerica, MA 01821)

The NOAA Ship OSCAR DYSON is the first-in-class fisheries research vessel (FRV) designed and built for National Oceanic and Atmospheric Administration (NOAA) for fish stock assessment. One of its most important contractual requirements was to meet the underwater noise limit of 1995 International Council for the Exploration of the Seas (ICES) report. This paper describes the efforts to conduct the underwater noise tests of the vessel. A commercial underwater measurement system was developed which was set-up off the Gulf of Mexico. The system consists of a small hydrophone array, signal conditioning, and data acquisition hardware. Field calibration was performed. Requirements for mounting the hydrophones and operation of a support vessel are discussed. In addition to measuring underwater sound, the distance between the source and receiver was also measured. Results are compared to the specification requirements and subsequent measurements performed by the U.S. Navy.

10:25

3aNCc3. Product audit for heavy duty diesel engines in production environment. Sanghoon Suh (Cummins Tech. Ctr., Cummins Inc., Columbus, IN 47202) and Jim Beresford (Jamestown Engine Plant, Cummins Inc., Lakewood, NY 14750)

A product audit at manufacturing plants has become more important due to the customer's requirements on product quality. Noise and vibration performance have been a primary concern for gas engines and small size diesel engines. Lately, more interest has been shown by truck manufacturers about engine noise for heavy duty diesel application. It has been regarded that acoustic measurements requires dedicated measurement environment for detailed study. This case study shows that acoustic measurements can be performed at performance cell without any dedicated acoustic treatment at the manufacturing plant to identify some of the noise characteristics with proper preparation. Order tracking and loudness were used to identify two different characteristics related to front gear train in heavy duty diesel engines. In addition, the coordination between technical organization and manufacturing plant for the data acquisition and analysis is discussed.

10:45

3aNCc4. Hand transmitted vibration measurement using experimentally simulated hand-arm rig. Rahul Kadam, Marty Johnson, Jorge Muract, and Ricardo Burdisso (Vib. and Acoust. Labs, Virginia Tech, 143 Durham Hall, VA 24061-0238)

Hand transmitted vibration (HTV) is one of the most common hazards faced by workers in the construction industry. A major source of HTV is hand held percussion tools, such as pneumatically driven chipping hammers and rock drills. This paper presents a new approach to measuring the vibration from these tools using a mechanically simulated hand arm model to which the tools are attached. The simulated hand-arm model has been designed to have similar dynamic characteristics to that of a human hand-arm system. This approach addresses the issue of repeatability as HTV measurements suffer from variability between cases and from user to user. Different percussion tools ranging from a small chipping hammer to a rock drill, from two major construction equipment manufacturing companies, were used to conduct the vibration tests. The goals of these tests were to measure the vibration levels and to determine the most dominant modes of vibration. Comprehensive data analyses of the tests results are presented. Further, the paper discusses potential vibration control methods and expected performance levels.

11:05

3aNCc5. Vibro-acoustic prediction of low-range planetary gear noise of a automotive transfer case. Ravi Narayanaswamy and Charles D. Glynn (BorgWarner TorqTransfer Systems, 3800 Automation Ave., Auburn Hills, MI 48326)

This paper discusses a practical approach to predict low range planetary gear noise in an automotive transfer case using commercial numerical codes. Dynamic responses of the planetary gear set are calculated using a 2D finite element/contact mechanics model. These responses are used as boundary conditions to calculate surface velocities on the transfer case housing using model frequency response analysis. Subsequently, the surface velocities are used in a vibro-acoustic model to predict acoustic responses of the transfer case. The predicted acoustic responses are compared to experimental measurements conducted in a hemianechoic chamber. It is shown that the predictions are in reasonable agreement with the experiments and the approach enables designers to obtain required information on acoustic responses of the transfer case in a timely and cost-efficient manner.

WEDNESDAY MORNING, 19 OCTOBER 2005

SALON C, 9:45 TO 11:05 A.M.

Session 3aNCd

NOISE-CON and Noise: Measurement of Product Noise Emissions I

Matthew A. Nobile, Chair

IBM Hudson Valley Acoustics Lab., Boardman Rd. Site, 2455 South Rd, Poughkeepsie, NY 12601

9:45

3aNCd1. Use of existing standards to measure sound power levels of powered hand tools—necessary revisions. Charles S. Hayden (Natl. Inst. of Occupational Safety and Health, 4676 Columbia Pkwy. C27, Cincinnati, OH 45226) and Edward Zechmann (Costella Group, Cincinnati, OH 45226)

At recent NOISE-CON and Acoustical Society of America meetings, noise rating labeling was discussed as a way of manufacturers providing full disclosure information for their noise emitting products. The first step is to gather sound power level data from these products. Sound power level data should be gathered in accordance with existing ANSI and/or ISO standards. Some standards, such as ANSI 12.15, may not define true operational noise emissions[r1] and thus may provide inaccurate information when that information is used to choose a hearing protection device or used to make a purchasing decision. A number of standards were systematically combined by NIOSH researchers to provide the most accurate information on sound power levels of powered hand tools used in the construction industry. This presentation will detail some of the challenges of existing ANSI 12.15 (and draft ANSI 12.41) to measure sound power levels of electric (and pneumatic) powered hand tools.

10:05

3aNCd2. Consumer oriented product noise testing. Les Blomberg (Noise Pollution Clearinghouse, P.O. Box 1137, Montpelier, VT 05601-1137)

This paper explores the need for product noise measurements and how best to meet that need in the near future. Currently there is only a small market place for quieter consumer products. This is not because of lack of interest. No one really wants to announce to everyone in their house that they just flushed the toilet, few really want the entire neighborhood to know they are mowing their yard, etc. The small market place is primarily due to a lack of regulations on product noise, a lack of information easily available to consumers about which products are quieter, and market consolidation resulting in fewer manufacturers, most of whom are unwilling to emphasize their quieter products at the risk of eroding sales of their noisier ones (that currently have greater market share). In the absence of the EPA fulfilling its statutory requirement to regulate and label product noise under the Noise Control Act of 1972, and with the unwillingness of most industries to voluntarily publish accurate product noise data, there is a significant role for "Consumer Oriented Product Noise Testing." This

paper explores the Noise Pollution Clearinghouse's ongoing and planned product noise testing, evaluating its advantages, disadvantages, and limitations.

repeatability. Additionally, a microphone was placed in the nominal hearing zone of the tool operator to acquire a time series to assess other sound metrics.

10:25

3aNCd3. Sound power measurement techniques for powered hand tools. Edward L. Zechmann (Constella Group) and Charles Hayden (Natl. Inst. of Occupational Safety and Health, 4676 Columbia Parkway C27, Cincinnati, OH 45226)

As part of a project to reduce noise induced hearing loss in the construction industry, NIOSH developed a database of sound power level measurements of electric powered hand tools typically used in the construction industry. The tool testing jigs and setups specified and illustrated in ANSI S12.15 were modified to accommodate the higher precision ten-microphone arrangement used in ISO 3744. ANSI S12.15 is sometimes vague regarding the tool testing jig design, so test jigs were designed to supplement existing specifications in the standard. In the course of the project, test jigs were designed, techniques were devised to improve repeatability of measurements, to reduce waste materials, to reduce measurement setup time, and to reduce data acquisition time. Several types of tools were tested including circular saws, grinders, screw drivers, drills, jig saws, reciprocating saws, miter saws, hammer drills, belt sanders, and impact wrenches. The test jig designs and measurement techniques may help others to save time, reduce waste material, and improve measurement

10:45

3aNCd4. Measurement of noise from toys. Robert Altkorn, Scott M. Milkovich, and Gene Rider (RAM Consulting, 2107 Swift Dr., Ste. 200, Oak Brook, IL 60523)

Noise from toys is an issue receiving increasing attention in the toy and consumer product safety communities. Concern over loud toys is motivated both by reports of increasing hearing loss among children (the U.S. CDC estimated in 2001 that 12.5% of U.S. children 6 to 19 years old have permanent or temporary noise induced threshold shift in one or both ears) and by technological advances enabling sound and noise producing toys of increased play value at lower and lower cost. Consumer watchdog groups such as PIRG routinely identify excessively loud toys in their yearly lists of most dangerous toys. In 2003 ASTM revised its toy safety standard (F963-03) to include A and C weighted sound pressure level measurements and specific play or use dependent measurement geometries. RAM Consulting measures noise from toys as part of a comprehensive product safety program. Sound measurement equipment, geometries, and procedures used at RAM for different types of toys will be discussed. Unusual problems in noise measurement will be considered, as will the appropriateness of A and C weighting for the youngest age groups.

WEDNESDAY MORNING, 19 OCTOBER 2005 ⟶ SALON E, 9:45 A.M. TO 12:05 P.M.

Session 3aNCe

NOISE-CON and Noise: Predicting and Assessing Community Responses to Noise I

Jim Barnes, Cochair
Acentech, 29 Isaac Davis Rd., Concord, MA 01742

Lawrence S. Finegold, Cochair
1167 Bournemouth Ct., Centerville, OH 45459-2647

William A. Russell, Cochair
U.S. Army Center for Health Promotion and Preventive Medicine, Army Operational Noise, 5158 Blackhawk Rd., Aberdeen Proving Ground, MD 21010–5403

9:45

3aNCe1. Case study: The environmental process at Logan International Airport. Nancy Timmerman (Consultant in Acoust. and Noise Control, 25 Upton St., Boston, MA 02118-1609)

For a facility like Logan International Airport in Boston, the Environmental Impact Analysis is performed annually; it is an ongoing activity. The effort is not performed by any one person or any one company. This paper will present the analysis, reporting, and review process for Logan as far as noise is concerned. Examples will be drawn from the 2003 Environmental Data Report (submitted to the Commonwealth of Massachusetts) and from the author's personal experience.

10:05

3aNCe2. Demonstrating adequacy of railroad warnings: A case study. John Erdreich (Ostergaard Acoust. Assoc., 200 Executive Dr., West Orange, NJ 07052) and Victor DiFrancesco (New Jersey Div. of Law, Trenton, NJ 08625-0116)

Frequently a jury is asked to evaluate a claim that a device provided adequate warning of a hazardous condition. Although there is extensive research on which basis an expert may evaluate such claim, the layman is unprepared to integrate such information into a decision. Previous attempts to produce courtroom demonstrations of warnings have been easily challenged by opposing experts. In a recent case involving a pedestrian accident on a railroad right-of-way we produced a jury demonstration that

was accepted by the court, was not challenged effectively by opposing counsel, and communicated effectively with the jury. This paper describes a demonstration that incorporated binaural recording and presentation with a synchronized video orientation and simulation. We will present an excerpt of the demonstration.

10:25

3aNCe3. Using transportation demand models to assess regional noise exposure. Kenneth Kaliski (Resource Systems Group, Inc., 55 Railroad Row, White River Junction, VT 05001)

In the United States, most metropolitan areas run some type of transportation demand model to estimate regional travel patterns, and, to some extent, air pollution. The more advanced of these models accurately represent the geographic contours of the roadways (in contrast to the older straight-line node and link models). This allows an almost seamless integration of these new transportation demand models into noise prediction models. Combined with the locations of individual homes from a separate E911 database, we can readily make estimates of the noise exposure of populations over large areas. In this paper, the regional traffic noise exposure of residences of Chittenden County, VT is estimated and mapped. It was found that 30% of the residences are exposed to noise levels exceeding the WHO sleep disturbance level of 45 dB LAeq(8) and 20% of residences are exposed to levels exceeding the WHO ''serious annoyance'' level of 55 dB LAeq(16). Maps show noise contours as well as individual homes color coded based on relative day and night noise exposure levels. Measured sound level data are given for particular locations to validate the predictions.

10:45

3aNCe4. Development of method for assessing traffic noise in certain typical conditions. Weixiong Wu (AKRF Inc., 440 Park Ave., New York, NY 10016)

Assessing traffic noise is difficult in certain typical conditions in New York City due to changed street geometries, challenges of collection of non-traffic noise components, and levels of existing noise affected by heavy traffic at adjacent streets, among other variables. In general, a proportional model, i.e., a logarithmic equation to compute total passenger car equivalents (PCEs), is employed to assess traffic noise impacts based upon the noise methodology and the noise criteria under the City Environmental Quality Review (CEQR) guidelines. However, in some typical conditions, such as significant changes in roadway or street geometry, roadways that currently carry no or very low traffic volumes, and existing noise levels that are the result of multiple sources, the FHWA Traffic Noise Model (TNM) can be used to better compute project-generated traffic components. This paper presents a development of noise analysis method dealing with these conditions. Once a proportional model identifies any potential noise impacts for screening purposes, TNM computations can be conducted for more thorough and detailed noise analyses. The results demonstrate that while a proportional model provides a practical and convenient noise analysis for most situations, TNM can provide more accurate noise assessments for the conditions listed above.

11:05

3aNCe5. Heavy construction equipment noise study using dosimetry and time-motion studies. Ellsworth R. Spencer and David S. Yantek (NIOSH, Pittsburgh Res. Lab., 626 Cochrans Mill Rd., Pittsburgh, PA 15236)

Noise-induced hearing loss continues to afflict workers in many occupational settings despite longstanding recognition of the problems and well-known methods of prevention and regulations. Sound levels associated with heavy construction equipment range from 80 to 120 dB(A) and power tools commonly used in construction produce sound levels up to 115 dB(A). The focus of the research was to determine the noise exposures of heavy construction equipment operators while documenting the workers' tasks, (i.e., hauling, moving, and/or pushing construction material). Time-motion studies were performed at the construction sites and were used to correlate the noise dosage with the work performed by equipment operators. The cumulative dose for the operator was then plotted with references to work tasks, to identify the tasks that caused the greatest noise exposure. Three construction sites were examined and located in the western Pennsylvania and eastern Ohio areas. The types of construction equipment studied included asphalt pavers, backhoes, bulldozers, compaction equipment, excavators, haul trucks, telehandlers, and wheeled loaders. The results showed that bulldozer operators consistently had the highest noise exposures, ranging from a NIOSH REL (Recommended Exposure Limit) dose of 844% to 25 836% and an OSHA PEL (Permissible Exposure Limit) dose of 139% to 1397%.

11:25

3aNCe6. Federal Highway Administration (FHWA) Roadway Construction Noise Model (RCNM). Judith L. Rochat and Clay N. Reherman (U.S. Dept. of Transportation, Volpe Natl. Transportation Systems Ctr., 55 Broadway, Cambridge, MA 02142)

Roadway construction is often conducted in close proximity to residences and businesses and should be controlled and monitored in order to avoid excessive noise impacts. To aid in this process, the Volpe Center Acoustics Facility, in support of the Federal Highway Administration (FHWA), has developed a construction noise screening tool. The FHWA Roadway Construction Noise Model (RCNM) is a newly developed national model for the prediction of construction noise. The model is based on the construction noise prediction spreadsheet developed for the Central Artery/Tunnel Project in Boston, MA (CA/T Project or ''Big Dig'') by Erich Thalheimer of Parsons Brinckerhoff Quade & Douglas, Inc. The CA/T Project is the largest urban construction project ever conducted in the United States and has the most comprehensive noise control specification ever developed in the United States. RCNM incorporates the CA/T Project's noise limit criteria and extensive construction equipment noise database, where these parameters can be modified according to each user's needs. Users can also activate and analyze multiple pieces of equipment simultaneously and define multiple receptor locations, including land-use type and baseline noise levels, where RCNM will calculate sound level results for multiple metrics.

11:45–12:05

Design Software Demonstration

Soren Pedersen and Judith Rochat

Session 3aNS

Noise, NOISE-CON, Engineering Acoustics and Archiectural Acoustics: Hospital Interior Noise Control

Ilene J. Busch-Vishniac, Cochair

Johns Hopkins Univ., Dept. of Mechanical Engineering, 3400 N. Charles St., Baltimore, MD 21218-2681

James E. West, Cochair

Johns Hopkins Univ., Dept. of Electrical and Computer Engineering, 3400 N. Charles St., Baltimore, MD 21218-2686

Chair's Introduction—9:40

Invited Papers

9:45

3aNS1. Noise control considerations for patient rooms. Benjamin Davenny (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

The patient room envelope is a path between outside noise sources and the patient receiver. Within the patient room there are several sources including televisions, clinical monitor alarms, medical pumps, etc. Noise control in patient rooms relies on a combination of the sound transmission loss of the patient room envelope and the level of background sound at the patient's head. Guidelines published by the American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), American Institute of Architects (AIA), and the U.S. Department of Defense for background noise and sound transmission loss in patient rooms will be discussed. Appropriate levels, spectra, and temporal characteristics of background sound at the patient head location may be helpful in raising the threshold of annoying sounds. Various means of personal hearing protection for patients will be discussed. Sound-pressure levels in patient rooms reported in previous literature will also be discussed.

10:05

3aNS2. Noise and vibration issues in contemporary facilities. Jack B. Evans, (JEAcoust., 1705 West Koenig Ln., Austin, TX 78756)

Contemporary medical facilities, whether for diagnosis, treatment or research, must accommodate the needs of humans and medical instruments or equipment. With regard to acoustics, the disparate requirements vary from acoustic privacy to freedom from intrusive noise, to vibration control. Medical facilities also grow and change, subjecting occupants to demolition and construction disturbances. Acoustical criteria employed in building designs are often inadequate to control the full spectral range of requirements, although criteria for many conditions do exist. This paper presents interesting problems and conditions that occur in healthcare and research facilities with brief case studies to illustrate them, including construction disturbances, tonal noise affects on patients, privacy, and vibration and low frequency noise affects on medical equipment. Measurement data will contrast ambient versus disturbed conditions. Solution concepts to example problems will be presented with photo or drawing illustrations. Discussion will include typical conflicts, constraints, limitations and compromises. Concluding remarks will suggest application of appropriate and comprehensive acoustical, noise and vibration criteria.

10:25

3aNS3. Frequency and neighborhood effects on auditory perception of drug names in noise. B. L. Lambert, S.-J. Lin, S. Toh (Dept. of Pharmacy Administration, Univ. of Illinois at Chicago, 833 S. Wood St., MC871, Rm. 271, Chicago, IL 60612), P. A. Luce, C. T. McLennan, R. La Vigne (Univ. at Buffalo, Buffalo, NY 19260), W. M. Fisher (William M. Fisher Consulting, Gaithersburg, MD), L. W. Dickey (Northwestern Univ., Evanston, IL 60208), and J. W. Senders (Toronto, ON, Canada)

Drug name confusions occur when drugs whose names look or sound alike (e.g., Celebrex/Cerebyx) are confused by clinicians and patients. Our project examines effects of frequency and similarity on auditory perception of drug names. Luce's Neighborhood Activation Model is the theoretical framework. Frequency information was taken from national prescribing frequency databases. Spoken drug names were transcribed to ARPAbet. Similarity neighborhoods were calculated based on phoneme edit distance using phoneme-to-phoneme confusion probabilities as substitution costs. The confusability of each name was estimated using Luce's frequency-weighted neighborhood probability rule (FWNPR). One hundred brand and one hundred generic drug names, stratified by FWNPR, were recorded. Low and high frequency sounds were filtered out to mimic telephone bandwidth. Names were used as stimuli in a perceptual identification experiment. Participants heard the drug names, at three different signal-to-noise ratios, against a background of 20-speaker babble. Participants were then shown the 200 names and asked to pronounce them and rate their familiarity. Data have been collected from 66 pharmacists. These data are being scored. Additional data will be collected from 50 physicians, 50 nurses, and 50 lay people. Results will be used to refine the theory and to build screening tools for drug companies and regulators.

3a WED. AM

10:45

3aNS4. Curing the noise epidemic. Susan Mazer (Healing HealthCare Systems, 100 W. Grove St., Ste. 245, Reno, NV 89509)

The argument is made that design does not stop when the fixed architectural and acoustical components are in place. Spaces live and breathe with the people who reside in them. Research and examples are presented that show that noise, auditory clutter, thrives on itself in hospitals. Application of the Lombard reflex studies fit into the hospital setting, but do not offer solutions as to how one might reduce the impact. In addition, the basis for looking at the noise component as a physical as well cultural dynamic will be addressed. Whether the result of the wrong conversation in the wrong place or the right conversation in an unfortunate place, talk mixed with sounds of technology is shown to cause its own symptoms. From heightened anxiety and stress to medical errors, staff burnout, or HIPAA violations, the case is made that noise is pandemic in hospitals and demands financial and operational investment. An explanation of how to reduce noise by design of the dynamic environment - equipment, technology, staff protocols is also provided.

11:05

3aNS5. Health Insurance Portability and Accountability Act (HIPAA) legislation and its implication on speech privacy design in health care facilities. Gregory C. Tocci and Christopher A. Storch (Cavanaugh Tocci Assoc., Inc., 327 F Boston Post Rd., Sudbury, MA 01776)

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 (104th Congress, H.R. 3103, January 3, 1986), among many things, individual patient records and information be protected from unnecessary issue. This responsibility is assigned to the U.S. Department of Health and Human Services (HHS) which has issued a Privacy Rule most recently dated August 2002 with a revision being proposed in 2005 to strengthen penalties for inappropriate breaches of patient privacy. Despite this, speech privacy, in many instances in health care facilities need not be guaranteed by the facility. Nevertheless, the regulation implies that due regard be given to speech privacy in both facility design and operation. This presentation will explore the practical aspects of implementing speech privacy in health care facilities and make recommendations for certain specific speech privacy situations.

11:25

3aNS6. Noise in neonatal intensive care units (NICUs) and its effect on high risk newborns. Robert E. Lasky, Amber L. Williams (Ctr. for Clinical Res. and Evidence-Based Medicine, Univ. of Texas-Houston Med. School, 6431 Fannin, MSB 2-106 Houston, TX 77030), Wim van Drongelen (Univ. of Chicago, Chicago, IL 60637), and Lincoln C. Gray (Univ. of Texas-Houston Med. Sch., Houston, TX 77030)

We conducted sound surveys in a large state of the art NICU with six separate rooms devoted to the sickest babies requiring the most intensive care (Level III) and six rooms devoted to babies requiring special but less intensive care (Level II). Each room was capable of caring for up to 8 babies. Additionally, there were 8 individual Isolation rooms. We used Larson Davis Spark® 703+ dosimeters to record 21 week long sound surveys, seven in each type of room. The American Academy of Pediatrics (1997) has recommended that sound levels in NICUs should never exceed 45 dB(A). That recommendation was exceeded 73.6% of the time in Level II, 92.1% of the time in Isolation, and 96.6% of the time in Level III. Sound levels were lowest in the Level II rooms especially for the softest sounds recorded (L_{90} and L_{70}). Level III rooms were noisiest except for the noisiest decile of sound (L_{10}). Isolation rooms were noisiest at the highest sound levels (probably because of their reverberant construction materials and enclosed space). Autocorrelation functions were calculated identifying periodic components in all three rooms at about 12 and 24 hrs. Periodic variations were very small compared to random sound variations.

WEDNESDAY MORNING, 19 OCTOBER 2005   CONRAD A, 8:00 A.M. TO 12:15 P.M.

Session 3aPA

Physical Acoustics: Scattering and Material Properties

Joel Mobley, Chair

Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

8:00

3aPA1. Fast calculations of transient pressure fields produced by spherically focused radiators. Robert McGough (Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824)

A fast method for calculating the pressure distribution generated by a spherical shell is demonstrated for a pulsed excitation. The integral expression derived for the fast method, which is analytically equivalent to the convolution integral defined for the impulse response, converges very quickly. This rapid convergence reduces the error and the computation time with respect to the impulse response approach. Furthermore, the resulting integral expressions are greatly simplified relative to those obtained with the impulse response approach. Results are computed for a pulsed excitation where the integrals are evaluated with the midpoint rule and Gauss quadrature. Simulation results demonstrate that these expressions achieve smaller errors in less time with Gauss quadrature, whereas the midpoint rule is appropriate for computing transient fields generated by very large spherical shells. [Work supported in part by NIH Grant R01-CA093669.]

3aPA2. Time-space decomposition of the transient pressure field generated by a rectangular piston. James F. Kelly and Robert J. McGough (Dept. of Elec. and Computer Eng., Michigan State Univ., East Lansing, MI 48824)

Exact solutions to the acoustic wave equation for canonical geometries are important for transducer evaluation and as a reference for large-scale simulations of wave propagation. An exact solution for the transient pressure field generated by a rectangular piston is derived. Consisting of a superposition of integrals, this solution avoids the edge singularities associated with earlier solutions. Computational overhead is significantly reduced by applying a novel time-space decomposition technique, whereby the spatial and temporal dependence of the integrand is analytically separated. By evaluating a set of integrals at each point in space, the transient pressure field is synthesized by weighting and summing against a set of time-dependent functions. This decomposition technique, which utilizes spatially band-limited integrands, avoids the artificially large sampling frequencies associated with the standard impulse response approach. Quantitative comparison with the spatial impulse response method is made in terms of maximum error and computation times. At 10% peak error, the time-space decomposition technique achieves a factor of 2 speed-up compared to spatial impulse response. Due to the rapid convergence of the integrals in the new technique, the speed-up increases to a factor of 6.8 at 1% peak error. [Work partially supported by NIH Grant 1R01 CA093669.]

3aPA3. Equivalence of expressions for the radiation force on cylinders and application to elliptical cylinders. Wei Wei and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

Using an appropriate grouping of terms, a radiation force expression for cylinders in a standing wave based on far-field scattering [W. Wei, D. B. Thiessen, and P. L. Marston, J. Acoust. Soc. Am. **116**, 202–208 (2004)] is transformed to an expression given elsewhere [F. G. Mitri, Eur. Phys. J. B **44**, 71–78 (2005)]. Mitri's result is from a near-field derivation for the specific case of a circular cylinder. In the usual case, in an ideal lossless media the far-field derivation is not an approximation. The far-field derivation also applies to noncircular objects having mirror symmetry about the incident wave vector. Some general and historical aspects of far-field derivations of optical and acoustical radiation force (going back to 1909) will be noted. Our formulation yields a simple low-frequency approximation for the radiation force on elliptical cylinders by introducing approximations for the partial-wave scattering coefficients of elliptical cylinders first derived by Rayleigh. [Work supported by NASA.]

3aPA4. Time domain evolution of diffuse fields in heterogeneous slabs. Joseph A. Turner (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588)

Fundamental studies of elastic wave scattering in heterogeneous media are applicable for problems at several length scales from ultrasonic to seismic waves. The intermediate scattering regime that lies between the single scattering and the diffusion limits is perhaps the least understood. Experiments of elastic wave scattering through a heterogeneous slab have been studied in the time domain using diffusion theory to fit the data. However, numerical solutions of the elastic wave radiative transfer equation (RTE) in the steady state have shown that the conditions for validity of the diffusion limit are only satisfied in the interior of the slab, many mean free paths away from the boundaries. Thus, an examination of the time domain multiple scattering in heterogeneous slabs is important to this class of experiments. The spatial distribution, temporal evolution, and partitioning of the diffuse longitudinal and shear energies are studied as a function of direction and frequency for several types of microstructure including polycrystalline metals and two-phase media using numerical solutions of the RTE. Finally, the ability of a diffusion-type solution to fit RTE solutions is also discussed with applications to inversion of experimental results. [Work supported by DOE.]

3aPA5. Monte Carlo simulations of elastic wave multiple scattering in polycrystalline media. Goutam Ghoshal and Joseph A. Turner (Dept. of Eng. Mech., W 317.4 Nebraska Hall, Univ. of Nebraska-Lincoln, Lincoln, NE 68588-052)

The scattering of elastic waves in polycrystalline media is primarily due to the orientation distribution of the crystal axes of the grains. This scattering may be used to extract microstructural parameters of the material such as grain size and grain texture. In particular, diffuse ultrasonic backscatter measurements have been especially useful for extracting microstructural information. Diffuse backscatter is often modeled analytically under a single scattering assumption. Beyond single scattering, the elastic radiative transfer equation (RTE) governs the evolution of diffuse energy and includes all multiple scattering effects. In this presentation, a rigorous connection between the RTE theory and the backscatter experiments is discussed. Specific solutions are obtained for a specimen excited by a normally incident longitudinal wave. Results are compared with previous backscatter theories. In addition, numerical results are presented using Monte Carlo simulations including various levels of scattering to observe differences between single and multiple scattering solutions. The Monte Carlo solutions are based on the analogy between the RTE and the Boltzmann theory of gas. Relevant applications for materials of common interest are discussed. These results are anticipated to impact ultrasonic nondestructive evaluation of polycrystalline media. [Work supported by U.S. DOE.]

3aPA6. Propagation of elastic waves in hexagonal crystals with fiber texture. Liyong Yang and Joseph A. Turner (Dept. of Eng. Mech., W317.4 Nebraska Hall, Univ. of Nebraska, Lincoln, NE 68588)

Many cold-working processes for polycrystalline metals cause alignment of the grains with a single symmetry axis called fiber texture. The existence of a preferred orientation of the grains has a big influence on the propagation and scattering of ultrasonic waves, which are often used for materials inspection. Knowledge of the wave attenuation of such textured materials is of both theoretical and practical interest to nondestructive testing and materials characterization. In this presentation, the quantitative relations between fiber texture and wave attenuations of hexagonal crystals are presented. The texture is characterized by a Gaussian distribution function that contains a single parameter that governs the transition of the texture from statistically isotropic to fiber texture. Under this assumption, the materials of interest have a varying degree of transverse isotropy representatives of processing conditions. Simple expressions for the attenuations of the three modes of waves are given in a concise representation. Finally, numerical results are presented and discussed in terms of the directional, frequency, and texture dependence. The results presented are expected to improve the understanding of the microstructure evolution during thermomechanical processing. [Work supported by DOE.]

3aPA7. Acoustic performance of a sound attenuating duct utilizing spatially periodic area changes with absorbing material. Donald Bliss (Mech. Eng., Duke Univ., Durham, NC 27708, dbb@duke.edu) and Pavel Danilov (Continuum Dynam., Inc., Ewing, NJ 08618)

Sound propagation of one-dimensional waves through a duct with spatially periodic area changes has been studied theoretically and experimentally in previous research. The resulting wave behavior is similar to that observed for periodic structures, with Bloch waves, passbands and stop bands. This configuration substantially reduces sound transmission from noise sources having harmonics that fall in the stop bands, while having little effect in passbands. The present study adapts this concept to configurations with dissipative sound absorbing material. The periodically spaced regions of the duct with larger cross-sectional area are partially filled with porous material, so that the open cross-sectional area remains constant throughout the duct. The duct is analyzed as a quasi-one-dimensional system using an approach based on periodic structure theory. The effect of the porous material is to make the propagation wave number in the porous/

open subsections complex, thereby changing the nature of the stop and passbands, and providing significant attenuation even in passbands. Results suggest that, for a given amount of porous material, the best attenuation strategy may be to distribute the porous material in periodically spaced subregions rather than in either a single larger volume, or in a uniform distribution all along the duct.

9:45

3aPA8. A novel ultrasonic method for characterizing suspensions of nanoparticles. Steven Africk (Prodyne Corp., 30 Fenwick Rd., Waban, MA 02468, safrick@att.net)

The Ultrasonic Pulsed Doppler (USPD) technique for characterization of nanoparticle suspensions is described. This method uses a single transducer as both source and receiver of pulsed ultrasound to measure backscattered energy from suspended and moving particles. Particle motion can be provided mechanically (e.g., by stirring or unidirectional flow) or by acoustic streaming generated by the interrogating acoustic field itself. The only requirement on sample volume is that it be larger than several cubic millimeters. It has been demonstrated that the system can measure particle concentration for Islets of Langerhans (on the order of 150 μm in diameter), cells (20 μm), emulsions (150 nm), carbon nanotubes (10 nm diameter), and polymer beads as small as 40 nm. Preliminary data suggest that the system can detect and measure dendrimers on the order of 4 nm in diameter. Other potential applications include measurement of particle size distribution, suspension viscosity, sample purity, and particle compressibility. The system concept is proposed as a research tool and is under development as a quality control device for nanomanufacturing in batch mode or as a monitor of a stream with particles. [Work supported by NIH.]

10:00–10:15   Break

10:15

3aPA9. Ultrasonic scattering in two-phase polycrystalline media. Dalie Liu and Joseph A. Turner (Dept. of Eng. Mech., W 317.4 Nebraska Hall, Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0526, dalieliu@bigred.unl.edu)

Successful processing of materials by powder sintering relies on the creation of strong interparticle bonds. During certain critical stages of the sintering process, the medium may be modeled as two phases consisting of the particles and a surrounding matrix. Ultrasonic methods have been proposed as a potential tool for monitoring such sintering processes. Thus, an understanding of the propagation and scattering of elastic waves in two-phase solids is of fundamental importance to these monitoring techniques. Here, a combined theoretical and numerical approach is used to address this problem. Ultrasonic attenuation and diffuse backscatter are studied theoretically using elastodynamic and stochastic wave theory, based on the spatial statistics of the density and Lamé parameters of the materials constituents under assumptions of statistical homogeneity and statistical isotropy. The numerical models are based on Voronoi polycrystals surrounded by a matrix of different material properties. Elastic wave simulations using the finite-element method (FEM) are then created to examine the influence of the grain size, wave type, and material statistics. The numerical results are compared with the attenuation theory proposed. The results presented are anticipated to impact the monitoring of materials processing of important engineering materials. [Work supported by U.S. DOE.]

10:30

3aPA10. Superluminal and negative group velocity of ultrasonic waves in microsphere suspensions. Joel Mobley (Dept. of Phys. and Astron., Univ. of Mississippi, 108 Lewis Hall, P.O. Box 1848, University, MS 38677-1848)

For acoustic propagation in a sufficiently dispersive medium, it is possible for the group velocity to exceed the vacuum speed of light and even become negative. In this talk a physically realizable suspension of elastic spheres in water is described that can support such abnormal (i.e., superluminal/negative) acoustic group velocities at moderate solids concentrations. The relative importance of single versus multiple scattering descriptions of these suspensions at volume fractions that can support abnormal velocities is discussed. The physical relevance of the group velocity concept to pulse propagation is also illustrated. Conditions under which superluminal/negative group velocities can be observed as the speed of a peak in the envelope of an acoustic pulse in the low-megahertz ultrasound band are identified. Waveforms derived from propagation simulations are used to demonstrate how superluminal/negative group velocities could be directly observed experimentally.

10:45

3aPA11. Finite-bandwidth Kramers–Kronig relations for the determination of acoustic group velocity and frequency derivative of attenuation. Joel Mobley (Dept. of Phys. and Astron., Univ. of Mississippi, 108 Lewis Hall, P.O. Box 1848, University, MS 38677-1848)

Kramers–Kronig (K–K) analyses of experimental data are complicated by the extrapolation problem, which is concerned with how the unexamined spectral bands impact K–K calculations. In this talk the causal linkages in resonant-type data are demonstrated without extrapolation through acoustic K–K relations for the group velocity and the derivative of the attenuation coefficient (components of the derivative of the complex wave number). Due to their shape-invariant predictions with respect to subtraction constants, these new relations provide stricter tests of causal consistency than previously established K–K relations for the phase velocity and attenuation coefficient. For both the group velocity and attenuation derivative, three forms of the relations are derived. These relations are equivalent for bandwidths covering the infinite spectrum, but differ when restricted to finite spectral windows. Using experimental data from microsphere suspensions, the accuracy of finite-bandwidth K–K predictions for group velocity, and the attenuation derivative are demonstrated. Of the multiple methods, the most accurate forms were found to be those whose integrals were expressed only in terms of the phase velocity, and attenuation coefficient themselves requiring no differentiated quantities.

11:00

3aPA12. Quantitative acoustic gas sensing: A two-frequency reconstructive algorithm based on molecular relaxation. Andi Petculescu and Richard Lueptow (Northwestern Univ., 2145 Sheridan Rd., Evanston, IL 60208)

Commonly, acoustics-based devices for the sensing and monitoring of fluid media have involved measuring changes in the phase velocity or sound speed. What has been left out is the acoustic attenuation. The non-classical or relaxational attenuation of sound in fluids arises from the inability of internal molecular degrees of freedom (vibration, rotation, isomerism, etc.) to follow the acoustic fluctuations. Thus, attenuation is directly related to the molecular relaxation times. In traditional laboratory experiments of acoustic relaxation in fluids, one needs to use several acoustic frequencies and vary the ambient pressure over a wide range in order to cover the main relaxation processes. In a sensing device, this would be highly impractical due to time constraints. A novel and unique algorithm is presented that uses only two acoustic frequencies at a single pressure to reconstruct the full frequency dependence of the primary relaxation processes via the effective specific heat, thus enabling a potential sensor to pinpoint the nature and concentration of contaminant molecular species in a host gaseous environment. It will be shown how the algorithm identifies a contaminant of methane in a simulated host atmosphere of nitrogen with accuracy better than 0.5%. [Work supported by NASA.]

3aPA13. Future trends in acoustic gas sensing. Andi G. Petculescu and Richard M. Lueptow (Northwestern Univ., 2145 Sheridan Rd., Evanston, IL 60208, a-petculescu@northwestern.edu)

Traditionally, acoustic devices for monitoring gaseous environments have involved measuring the sound speed in the analyzed medium. However, measurements of sound speed alone do not always allow a quantitative analysis of the gas species present in the sensed volume. What has been consistently omitted is the acoustic attenuation. The nonclassical or relaxational attenuation of sound in fluids arises from the inability of internal molecular degrees of freedom (vibration, rotation, isomerism, etc.) to follow the acoustic fluctuations. This feature relates it directly to relaxation times, which are characteristic of the molecular species that are present. Together with sound speed, attenuation can be used to obtain the effective specific heat of the fluid, which is the fundamental macroscopic footprint of all internal relaxation processes characteristic of the gases that form the sensed environment. The purpose of this presentation is to present several trends and challenges in the development of the next-generation ''smart'' acoustic gas sensors. The operational line of thought, based on physical principles that future intelligent sensors might employ to quantitatively detect and identify contaminant gases leaking into a host environment, is outlined. [Work was supported by NASA.]

3aPA14. Resonant tube spectroscopy of liquids. Bjorn A. Prenzlow, Jonathan Tapson (Dept. of Elec. Eng., Univ. of Cape Town, Private Bag, Rondebosch, 7701, South Africa, bjorn.prenzlow@gmail.com), and Jevon R. Davies (Cape Peninsula Univ. of Technol., Cape Town, 8000, South Africa)

The resonant behavior of a fluid-filled piezoelectric tube is studied in an attempt to design a sensor to measure the density, bulk modulus, and viscosity of the sample fluid. It is proposed that the impedance spectrum of the tube over several of its thickness resonances, and the effect of coupling the fluid's radial modes to these modes, will allow the measurands to be evaluated. The fluid produces several high-Q modes. These couple to the mechanical vibrations of the tube, producing a spectrum with rich modal behavior. In this work, a theoretical model describing the effective acoustic impedance of the fluid is mathematically coupled to that of the empty tube. The relevant governing properties of the piezoelectric tube and the fluid are obtained by optimizing the model with respect to a real spectrum, using an adaptive mutation breeder algorithm. The error between the theoretical and the real spectra for calibrated samples is shown to be small. As a result, the density, bulk modulus, and viscosity of the fluid may be calculated successfully from the features of the measured spectra.

3aPA15. A low-pressure technique for measuring the nonlinearity parameter *B/A* of small volume liquids. Jevon Davies (CIR, Dept. of Elec. Eng., Cape Peninsula Univ. of Technol., Tennant St., Cape Town, 8000, South Africa, daviesj@cput.ac.za), Bjorn Prenzlow, and Jonathan Tapson (Univ. of Cape Town, Rondebosch 7700, South Africa)

The acoustic nonlinearity parameter B/A is important for characterizing acoustic propagation in fluids. Recent studies have shown that B/A may be used to investigate biological tissue composition and provide insight into the molecular behavior of fluids. Such methods require a means of measuring B/A noninvasively using small sample volumes. This paper describes a variation of the isentropic phase method, which makes use of high-Q resonant modes found within a liquid-filled piezoelectric cavity resonator. B/A is measured indirectly by observing the resultant change in velocity occurring with a low-pressure (20 kPa) sinusoidal variation, performed under isothermal and steady-state conditions. It is shown that a high-frequency pressure oscillation is all that is needed for reliable B/A measurements, and these can be performed on sample volumes of less than 700 μl. Furthermore, measurements can be performed continuously once steady-state conditions have been achieved. Values for B/A are provided for three liquids, namely FC43, FC75, and water, with a consistency error of less than 1%.

3aPA16. New technology for investigation of photoelastic properties of nontransparent crystals. Farkhad R. Akhmedzhanov (Samarkand State Univ., 15, Univ. Blvd., Samarkand 703004, Uzbekistan, farkhad2@rambler.ru)

As is well known, in absolutely transparent crystals an acousto-optical interaction is lacking. In this connection's greatest interest acousto-optical investigations are presented in optical range, in which the crystal is not transparent. In order to carry out similar experiments, it is necessary to use the crystal with a small light absorption coefficient at applied wavelength. Evidently, the geometry of experiment will be different from that for transparent crystal because the few-millimeters-thick crystal will absorb the incident light completely. The main problem is selection of an appropriate buffer crystal for the investigated crystal that the light beam penetrated into sample by Bragg angle at required frequency. In the present work the measurements of effective photoelastic constant have been done for Si crystal, which is not transparent at wavelength 632.8 nm. The crystal of lithium niobate was used as the buffer crystal because its refractive index ($n=2.28$) is not far from that for Si crystal. The results of executed investigations have shown the possibility of determination of photoelastic constants in nontransparent crystals by Bragg scattering method. At that rate, the very high intensity of scattered light can be obtained in comparison with that in transparent crystals.

3a WED. AM

Session 3aSA

Structural Acoustics and Vibration, Noise and NOISE-CON: Noise and Vibration of Transit Systems

Lance Meister, Chair

Harris Miller Miller and Hanson, 15 New England Executive Park, Burlington, MA 01803-5221

Invited Papers

8:30

3aSA1. Vibration response of buildings to rail transit groundborne vibration. James Phillips (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618)

The FTA guidelines for detailed analysis and prediction of groundborne noise and vibration generated by rail transit systems are based on established empirical methods. The procedures for the measurement of vehicle/track system source strength and the attenuation of vibration as it propagates with distance through the ground are generally accepted practice at this time. However, characterization of the building response is open to debate, due in part to the wide array of building construction encountered adjacent to transit systems. Numerous measurements that have been obtained in a variety of building construction types are presented and preliminary conclusions are drawn regarding the responses of several common building types to rail transit groundborne vibration.

9:00

3aSA2. Latest developments of a tire-derived aggregate underlayment to reduce groundborne vibration from light rail transit track. Steven Wolfe (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 94618)

Groundborne vibration reduction for light rail transit (LRT) tracks usually takes the form of some type of soft or special track fastener, or a type of isolation system built under or incorporated as part of the track structure. Two well-known types of track isolation systems are the ballast mat and the floating slab trackbed. A more recent and less-expensive type of track isolation system involves tire-derived aggregate (TDA or shredded scrap tires) used as an underlayment beneath ballast and tie tracks. Field tests have been performed to determine the vibration attenuation and damping properties of TDA for use beneath rail lines and other possible applications. The most recent tests were conducted at an existing LRT system with the trains operating at speed on several sections of revenue ballast and tie track, and where the TDA was installed under ballast and tie rail transit track. Details of the field tests are presented with a focus on the most recent test results, which indicate the performance of the TDA underlayment in comparison with other track isolation methods.

9:30

3aSA3. Vibration transfer mobility measurements using maximum length sequences. Herbert L. Singleton, Jr. (Cross-Spectrum Labs., P.O. Box 90842, Springfield, MA 01139)

Vibration transfer mobility measurements are required under Federal Transit Administration guidelines when developing detailed predictions of ground-borne vibration for rail transit systems. These measurements typically use a large instrumented hammer to generate impulses in the soil. These impulses are measured by an array of accelerometers to characterize the transfer mobility of the ground in a localized area. While effective, these measurements often make use of heavy, custom-engineered equipment to produce the impulse signal. To obtain satisfactory signal-to-noise ratios, it is necessary to generate multiple impulses to generate an average value, but this process involves considerable physical labor in the field. To address these shortcomings, a transfer mobility measurement system utilizing a tactile transducer and maximum length sequences (MLS) was developed. This system uses lightweight off-the-shelf components to significantly reduce the weight and cost of the system. The use of MLS allows for adequate signal-to-noise ratio from the tactile transducer, while minimizing the length of the measurement. Tests of the MLS system show good agreement with the impulse-based method. The combination of the cost savings and reduced weight of this new system facilitates transfer mobility measurements that are less physically demanding, and more economical when compared with current methods.

10:00−10:15   Break

10:15

3aSA4. In-service tests of the effectiveness of vibration control measures on the BART rail transit system. Hugh Saurenman (ATS Consulting, LLC, 725 S. Figueroa St., Ste. 1580, Los Angeles, CA 90017, hsaurenman@ATSConsulting.com)

Controlling vibration from new rail transit systems can be quite expensive when the alignment passes through residential areas. However, there is relatively little documented information on how effective different vibration mitigation approaches perform under in-service conditions. This paper presents results of a number of vibration measurements of the different track forms used on the current San Francisco Bay Area Rapid Transit (BART) system including floating slab, resiliently supported half ties, and high-resilience direct fixation fasteners in subways and one section of floating slab used on at-grade tracks. The goal was to obtain data that would improve the predictions of future vibration levels and perhaps lead to more cost effective vibration mitigation strategies for the proposed BART extension to San Jose. The tests show that the floating slabs are performing much as designed, the resiliently

supported half ties are less effective than expected, and the high resilience track fasteners are probably performing as expected, although the results are clouded because of severe rail corrugation in the area where the new fasteners were installed. Some unanticipated results are the apparent interaction of the floating slab resonance, the wheel rotation frequency, the bogey dynamics, and vibration propagation characteristics of the ground.

Contributed Papers

10:45

3aSA5. An evaluation of the accuracy of the Federal Transit Administration approved, groundborne noise and vibration prediction model for rail transit systems; a case study for a new subway line using post construction, transit operation measurements. Richard A. Carman and Carlos H. Reyes (Wilson, Ihrig & Assoc., 5776 Broadway, Oakland, CA 95618)

The groundborne noise and vibration model developed by Nelson and Saurenman in 1984, now recognized by the Federal Transit Administration as the approved model for new transit system facilities, is entering its third decade of use by engineers and consultants in the transit industry. The accuracy of the model has been explored in the past (e.g., Carman and Wolfe). New data obtained for a recently completed extension to a major heavy rail transit system provides an opportunity to evaluate the accuracy of the model once more. During the engineering design phase of the project, noise and vibration predictions were performed for numerous buildings adjacent to the new subway line. The values predicted by the model were used to determine the need for and type of noise and/or vibration control measures. After the start of transit operations on the new line, noise and vibration measurements were made inside several of the buildings to determine whether the criteria were in fact achieved. The measurement results are compared with the values predicted by the model. The predicted and measured, overall noise and vibration levels show very good agreement, whereas the spectral comparisons indicate some differences. Possible reasons for these differences are offered.

11:00

3aSA6. Federal Transit Administration's noise and vibration Guidance Manual: Updates after 10 years of experience. Carl E. Hanson (Harris Miller Miller & Hanson, Inc., 15 New England Executive Park, Burlington, MA 01803)

In April 1995, the Federal Transit Administration of the U.S. Department of Transportation issued its guidance manual, "Transit Noise and Vibration Impact Assessment." The manual was written to provide direction for the preparation of noise and vibration sections of environmental documents for public transportation projects in the interest of promoting quality and uniformity in assessments. In the 10 years since its release, the guidance manual has been used as the basis of noise and vibration assessments in over 80 environmental documents and has been used worldwide as a standard method for performing transit noise analyses. An updated version has recently been prepared that takes account of improvements in tools, changes in the environmental assessment procedures, modifications of impact criteria, and experience with mitigation measures. These factors, and other background information, will be discussed to provide insight regarding the revisions in the new edition.

11:15

3aSA7. An investigation of the noise levels surrounding the Chicago Transit Authority (CTA) Brown Line during peak and off-peak hours of operation. Connie G. Lee, Philip Muzzy, Samuel Lalk, Michael Cappiello (Columbia College Chicago, 33 E. Congress Ave., Chicago, IL 60605, miss.connie.lee@gmail.com), Nicholas Bailey, Judson Brown, Megan Dareing, Edward Fitzgerald, Coleman Martin, Justin Nuner, and Yotam Schlezinger (Columbia College Chicago, Chicago, IL 60605)

The Chicago Transit Authority's (CTA) Brown Line train is part of the Chicago "El" system and passes mostly through residential neighborhoods. Currently at maximum capacity, the Brown Line will be undergoing expansion, increasing the train length from six to eight cars. By taking sound-level measurements during morning peak hours and evening off-peak hours, this study analyzes the contribution of the Brown Line trains to the existing neighborhood noise floor. Data analysis reveals a significant disparity between the L90 and higher L_{eq} readings. As the sole contributor to the L10 levels, this study finds that the Brown Line trains are highly influential to the elevated L_{eq}. It is interesting to note that, while exempt from the city's evening noise ordinance, the L_{eq} is in excess of the prescribed level of 55 dB.

11:30

3aSA8. Assessing of the solid-borne noise produced by the urban traffic. Case study: St. Giuseppe dei Teatini's Church in Palermo. S. Barbaro, G. Barrera (Dept. of Environ. and Energy Res., Univ. of Palermo, Sci. Ave., 90128 Palermo, Italy), and A. Sansone Santamaria (Regional Agency for the Environ. Protection, 90100 Palermo, Italy)

In past years, the continuously increasing road traffic, correlated with both economical and social development, has improved the vibration rate upon the structures. Anyway often the component vibration is ignored, for different reasons. Today in Italy a law about vibration impact upon structures does not exist; we refer to the International Standard Organization. The vibration waves hitting the structures may cause damages, increasing at particular frequencies. There are other kinds of nonstructural damages, that are called cosmetic damages, such as cracks in the plaster and damages on architectural elements. In this work a survey protocol has been set up, with application to a busy street in the historical center of Palermo: Vittorio Emanuele street. In this street a church has been chosen like a case study: the St. Giuseppe dei Teatini's Church, that is a heritage building. The case study includes at the same time measurements of both vehicular traffic and vibration on the ground floor and first floor.

11:45

3aSA9. A theoretical-experimental approach for assessing the vibration impact due to vehicular traffic upon the heritage building structures. S. Barbaro, G. Barrera (Dept. of Environ. and Energy Res., Univ. of Palermo, Sci. Ave., 90128 Palermo, Italy), and A. Sansone Santamaria (Regional Agency for the Environ. Protection, 90100 Palermo, Italy)

In past years, a new method of structural analysis has been developed for assessing the vibration impact due to vehicular traffic upon the heritage building structures. This method is based on on the conjuncture among the monitoring of the structure, and the structural identification. The satisfaction of such demands in the case of buildings already realized, above all if ancient, can be prevented by the impossibility to reach to a complete knowledge both of the structural organism and the strength state of maintenance of every component of the building. An exhaustive knowledge would be achievable only applying such methods of investigation to induce heavy troubles to the analyzed manufactured article. From these consideration the idea of the construction of an elementary system is born. An elementary PC model, that is an ideal receiver whose all the mechanical and geometric characteristics are known, has been designed for *in situ* analysis and excited applying the experimental vibrational levels in order to evaluate the response. A further development is to realize the physical model and execute local experimental campaigns. This operative system has to be transportable and positionable near the source without holding up the traffic nor damaging the measurement site.

Session 3aSC

Speech Communication: Perception and Production in Atypical Populations (Poster Session)

Rachael F. Holt, Chair

Indiana Univ. School of Medicine, Otolaryngology-Head and Neck Surgery, 699 West Dr., Indianapolis, IN 46202

Contributed Papers

All posters will be on display from 8:45 a.m. to 11:45 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:45 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 11:45 a.m.

3aSC1. Relative saliency of pitch versus phonetic cues in infancy. Gina Cardillo, Patricia Kuhl (Dept. of Speech and Hearing Sci., Box 354875, and Inst. for Learning and Brain Sci., Box 357988, Univ. of Washington, Seattle, WA 98105), and Megha Sundara (Univ. Washington, Seattle, WA 98015)

Infants in their first year are highly sensitive to different acoustic components of speech, including phonetic detail and pitch information. The present investigation examined whether relative sensitivity to these two dimensions changes during this period, as the infant acquires language-specific phonetic categories. If pitch and phonetic discrimination are hierarchical, then the relative salience of pitch and phonetic change may become reversed between 8 and 12 months of age. Thirty-two- and 47-week-old infants were tested using an auditory preference paradigm in which they first heard a recording of a person singing a 4-note song (i.e., "go-bi-la-tu") and were then presented with both the familiar and an unfamiliar, modified version of that song. Modifications were either a novel pitch order (keeping syllables constant) or a novel syllable order (keeping melody constant). Compared to the younger group, older infants were predicted to show greater relative sensitivity to syllable order than pitch order, in accordance with an increased tendency to attend to linguistically relevant information (phonetic patterns) as opposed to cues that are initially more salient (pitch patterns). Preliminary data show trends toward the predicted interaction, with preference patterns commensurate with previously reported data. [Work supported by the McDonnell Foundation and NIH.]

3aSC2. Lexical and context effects in children's audiovisual speech recognition. Rachael Holt, Karen Kirk (Indiana Univ. School of Medicine, Otolaryngol.-HNS, Riley Res. Wing 044, 699 West Dr., Indianapolis, IN 46202), David Pisoni (Indiana Univ., Bloomington, IN 47405), Lisa Burckhartzmeyer, and Anna Lin (Indiana Univ. School of Medicine, Indianapolis, IN 46202)

The Audiovisual Lexical Neighborhood Sentence Test (AVLNST), a new, recorded speech recognition test for children with sensory aids, was administered in multiple presentation modalities to children with normal hearing and vision. Each sentence consists of three key words whose lexical difficulty is controlled according to the Neighborhood Activation Model (NAM) of spoken word recognition. According to NAM, the recognition of spoken words is influenced by two lexical factors: the frequency of occurrence of individual words in a language, and how phonemically similar the target word is to other words in the listeners lexicon. These predictions are based on auditory similarity only, and thus do not take into account how visual information can influence the perception of speech. Data from the AVLNST, together with those from recorded audio-

visual versions of isolated word recognition measures, the Lexical Neighborhood, and the Multisyllabic Lexical Neighborhood Tests, were used to examine the influence of visual information on speech perception in children. Further, the influence of top-down processing on speech recognition was examined by evaluating performance on the recognition of words in isolation versus words in sentences. [Work supported by the American Speech-Language-Hearing Foundation, the American Hearing Research Foundation, and the NIDCD, T32 DC00012 to Indiana University.]

3aSC3. Intelligibility and clarity of clear and conversational speech by children with implants and hearing aids. Arlene E. Carney, Yingjiu Nie, Jennifer Main, Edward Carney (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455), and Maureen B. Higgins (Boystown Natl. Res. Hospital, Omaha, NE 68131)

Studies of clear versus conversational speech have shown perceptual and acoustic differences. The basic premise is that the speaker understands the direction to speak more clearly, and translates the direction to speech motor acts. Children with hearing losses receive intervention during which they are instructed to use their best speech. The purpose of this study was to determine: (1) whether hearing-impaired children's intelligibility changed with directions to use better speech, and (2) whether these children's speech was judged to be clearer when they had intended to produce clear speech. There were two groups of speakers: 14 deaf children with cochlear implants and 7 hard-of-hearing children with hearing aids. Each produced ten short sentences using typical speech, better speech, and best speech. All sentences were presented to a total of 189 adult listeners with normal hearing who wrote down what they heard. Hard-of-hearing children had average speech intelligibility of 98%; those with implants averaged 66%. Both groups had very small increases across conditions. All sentences in three speech conditions were presented in a paired-comparison task to ten additional listeners. Results of clarity judgments will be discussed in relation to the relatively small changes in speech intelligibility. [Research supported by NIH.]

3aSC4. Phonetic context effects in adult listeners with cochlear implants. Radhika Aravamudhan and Andrew J. Lotto (Boys Town Natl. Res. Hospital, 425 North 30th St., Omaha, NE 68131)

From previous studies it is known that normal-hearing (NH) listeners have the ability to compensate for the acoustic variability present in speech through context-dependent perception of speech sounds. One question of practical and theoretical interest is whether listeners with cochlear implants (CI) also show context-dependent speech perception. Because of the lack of spectral resolution in the input, the representation of speech for

1962   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1962**

CI listeners may differ from NH listeners, which may interfere with perceptual compensation. In a test of this prediction, adult postlingually deafened CI listeners did not demonstrate the contrastive context effects elicited from NH listeners for either /da/–/ga/ targets and /al/–/ar/ contexts or V targets and /b_b/–/d_d/ contexts. In contrast, as predicted by the good temporal resolution of the CI signal, CI listeners showed normal effects of vowel length on preceding glide-stop categorization. CI simulations with NH listeners were also performed for some of these context effects. The results support the view that spectral content of contexts largely determines their effect on target sounds, and that phonetic content does not play an essential role. In addition, the findings predict that CI listeners will have particular difficulty with heavily coarticulated speech. [Work sponsored by NIH and NSF.]

3aSC5. The effects of speechreading training on viseme categories for vowels. Carolyn Richie (Commun. Disord., Butler Univ., 4600 Sunset Ave., Indianapolis, IN 46208)

The status of visemes, groups of visually confusable speech sounds, for American English vowels has been disputed for some time. While some researchers claim that vowels are visually distinguishable, others claim that some vowels are visually confusable and comprise viseme categories. Data from our study on speechreading words and sentences were examined for evidence of vowel visemes [C. Richie and D. Kewley-Port, J. Acoust. Soc. Am. **117**, 2570 (2005)]. Normal-hearing listeners were tested in auditory-visual conditions in masking noise designed to simulate a hearing loss. They were trained on speechreading tasks emphasizing vowels, consonants, or vowels and consonants combined. Pre- and post-training speechreading tests included identification of 10 vowels in CVC context. Pre-training data (vowel-identification confusion matrices) were used to determine whether vowel visemes exist for untrained speechreaders. Post-training results were examined to determine whether the number of vowel response categories increased and whether the number of vowel identification errors decreased, for trained versus untrained participants. The impact of these training programs on speechreading performance is discussed in terms of vowel visemes. [Work supported by NIH-DCD02229.]

3aSC6. The relation between categorical perception of speech stimuli and reading skills in children. Joshua Breier, Jack Fletcher, Patricia Klaas, and Lincoln Gray (Univ. of Texas Med. School, Houston, TX 77030, joshua.i.breier@uth.tmc.edu)

Children ages 7 to 14 years listened to seven tokens, /ga/ to /ka/ synthesized in equal steps from 0 to 60 ms along the voice onset time (VOT) continuum, played in continuous rhythm. All possible changes (21) between the seven tokens were presented seven times at random intervals, maintaining the rhythm. Children were asked to press a button as soon as they detected a change. Maps of the seven tokens, constructed from multidimensional scaling of reaction times, indicated two salient dimensions: one phonological and the other acoustic/phonetic. Better reading, spelling, and phonological processing skills were associated with greater relative weighting of the phonological as compared to the acoustic dimension, suggesting that children with reading difficulty and associated deficits may underweight the phonological and/or overweight the acoustic information in speech signals. This task required no training and only momentary memory of the tokens. That an analysis of a simple task coincides with more complex reading tests suggests a low-level deficit (or shift in listening strategy). Compared to control children, children with reading disabilities may pay more attention to subtle details in these signals and less attention to the global pattern or attribute. [Supported by NIH Grant 1 RO1 HD35938 to JIB.]

3aSC7. Importance of low-frequency on understanding interrupted speech: Sentence recognition of hearing-impaired (HI) and normal-hearing (NH) listeners. Su-Hyun Jin (Univ. of Wyoming, 1000 East University Ave., Division 3311, Laramie, WY 82070) and Peggy Nelson (Univ. of Minnesota, Minneapolis, MN 55455)

In a recent study of HI listeners' speech perception in noise (Jin and Nelson, 2004), two main factors, low-frequency audibility and auditory filter bandwidths, were highly related to amplified sentence recognition in modulated noise. Nine young adult listeners with sensorineural hearing loss and eight young NH listeners as controls participated in the series of experiments. Amplified speech recognition of the HI listeners was equal to that of the NH listeners in quiet and in steady noise, but was significantly poorer in modulated noise. Thus, even with adequate amount of amplification, HI listeners experienced significantly less masking release from the modulated maskers. The results indicated that those listeners with greatest hearing losses in the low frequencies were poorest at speech recognition in modulated noise. Additional results from HI and NH listeners will be presented, in which we systematically vary the audibility of different frequency regions of speech by filtering. Sentences are interrupted by either speech-shaped noise or silence gap while measuring the percent of sentence recognition. The purpose of the current study is to examine the importance of low-frequency audibility for auditory segregation/integration of interrupted speech.

3aSC8. Voice quality of children with cochlear implants acquired at early and later ages. Melanie M. Campbell, Stefanie Hanstein, and Christina Ney (Univ. of Alberta, 2-70 Corbett Hall, Speech Pathol. and Audiol., Univ. of Alberta, Edmonton, AB T6G 2G4, Canada)

The speech gains of children with cochlear implants (CIs) are well documented, but the literature on voice quality is sparse. It has reported atypical measures/ratings of voice pitch, pleasantness, timing, and acoustic features [Higgins *et al.* (2003); Perrin *et al.* (1998)]. Is voice quality now improving in children implanted very early? This pilot study compared the voice quality of (a) children with early acquired CIs and children with normal hearing and (b) the voice quality of children implanted later and earlier in life. Children aged 6 to 10 years, with early acquired CIs, and participants with normal hearing, age-matched to them, audio recorded sentences, vowels, and conversation. PERCI pressure measures were also performed. PERCI Differential and Oral Pressure values and Computerized Speech Lab (CSL) and Visipitch measures of voice-onset time and fundamental frequency were analyzed comparing the values from the hearing and the early implanted children and values gleaned from the study of Higgins *et al.* of children with later-acquired implants. CSL and Visipitch measures of intonation contour, intensity, and jitter were analyzed to compare the hearing and the early implanted participants. Ratings on the Wilson Voice Scale were correlated with measures of jitter, fundamental frequency, and intonation contour.

3aSC9. Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users. Joseph S. Perkell, Margaret Denny, Harlan Lane, Melanie L. Matthies, Ellen Stockmann, Mark Tiede, Majid Zandipour (Speech Commun. Group, Res. Lab. of Electron., MIT, Rm. 36-511, 50 Vassar St., Cambridge, MA 02139), and Jennell Vick (Univ. of Washington, Seattle, WA 98105)

The role of auditory feedback in speech production was investigated by examining how speakers modify phonemic contrasts in response to increases in the noise-to-signal (N/S) level in their auditory feedback. Seven cochlear implant users and seven normal-hearing controls pronounced utterances containing the vowels IY, UW, EH, and AE and the sibilants s and SH while hearing their speech mixed with noise at seven equally spaced levels over their tolerated ranges (which were smaller for implant users than for controls). Average vowel duration and spl rose with increasing N/S, monotonically for controls and more irregularly for implant users. On the other hand, a measure of overall vowel contrast rose

3a WED. AM

initially then fell with increasing N/S. The initial rise is interpreted as part of speakers attempts to maintain clarity with degraded acoustic transmission conditions. As N/S increases further, speakers can no longer detect phoneme contrasts, and economy of effort becomes an increasingly predominant influence, leading to contrast decrements. The sibilant contrast was more vulnerable to noise: the s-SH contrast distance fell over the entire range of increasing N/S. Results from implant users were more variable than from controls, especially for the sibilants. [Research supported by the NIDCD.]

3aSC10. Effects of speaking condition and hearing status on vowel production in postlingually deaf adults with cochlear implant. Lucie Menard, Margaret Denny, Harlan Lane, Melanie L. Matthies, Joseph S. Perkell, Ellen Stockmann, Jennell Vick, Majid Zandipour (Speech Commun. Group, Res. Lab. of Electron., MIT, Cambridge, MA 02139, menard@speech.mit.edu), Thomas Balkany, Marek Polack (Univ. of Miami Med. School, Miami, FL 33136), and Mark K. Tiede (Speech Commun. Group, Res. Lab. of Electron., MIT, Cambridge, MA 02139)

This study investigates the effects of speaking condition and hearing status on vowel production by postlingually deafened adults. Thirteen cochlear implant users produced repetitions of nine American English vowels in three time samples (prior to implantation, one month, and one year after implantation); in three speaking conditions (clear, normal, and fast); and in two feedback conditions after implantation (implant processor turned on and off). Ten speakers with normal hearing were also recorded. Results show that vowel contrast, in the $F1$ versus $F2$ space, in mels, does not differ from the pre-implant stage to the 1-month stage. This finding indicates that shortly after implantation, speakers had not had enough experience with hearing from the implant to adequately retune their auditory feedback systems and use auditory feedback to improve feedforward commands. After 1 year of implant use, contrast distances had increased in both feedback conditions (processor on and off), indicating improvement in feedforward commands for phoneme production. Furthermore, after long-term auditory deprivation, speakers were producing differences in contrast between clear and fast conditions in the range of those found for normal-hearing speakers, leading to the inference that maintenance of these distinctions is not affected by hearing status. [Research supported by NIDCD.]

3aSC11. Perception and production of /r/ allophones improve with hearing from a cochlear implant. Melanie Matthies (Dept. of Speech, Lang. and Hearing Sci., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215), Frank Guenther (Dept. Cognit. and Neural Systems, Boston Univ., Boston, MA 02215 and MIT, Cambridge, MA), Margaret Denny, Joseph Perkell, Ellen Stockmann, Jennell Vick (MIT, Cambridge, MA 02139), Mark Tiede (Haskins Lab., New Haven, CT 06511), and Harlan Lane (Northeastern Univ., Boston, MA 02115)

While variations in tongue shape for /r/ occur in different contexts, they normally result in only small changes in the primary acoustic cue (lowered $F3$). It was hypothesized that deaf speakers would show larger allophonic variation than normal-hearing speakers and that auditory feedback from a cochlear implant would reduce that variability. In this study, acoustic variation of stop and stop+/r/ tokens was evaluated pre- and post-implant ($n = 8$) and once in normal-hearing controls ($n = 8$). Measures included (1) r-variation: distance from the center of the distribution defined by the $F3$ transition and the value of $F3$ minus $F2$ at the /r/-vowel boundary for wabrav, wadrav, and wagrav tokens (2) vowel variation: distance from the center of $F1$, $F2$ for the second /a/ in wabav, wadav, and wagav tokens and (3) pairwise r-variation normalized by vowel variation. Pre-implant, seven of eight speakers had significantly greater normalized /r/ variation than the hearing speakers mean. Post-implant, seven of eight speakers had normalized /r/ are variation that was not different from the control mean. Implant users perception and production of stop and stop+/r/ blend productions improved; however, their /b-br/ and /g-gr/ contrast distances were still smaller one-year post-implant than the controls. [Research supported by NIDCD.]

3aSC12. Speech production in experienced cochlear implant users undergoing short-term auditory deprivation. Geoffrey Greenman (Dept. of Communicative Disord. and Sci., Univ. at Buffalo, 3435 Main St., 122 Cary Hall, Buffalo, NY 14214, greenman@buffalo.edu), Kris Tjaden (Univ. at Buffalo, Buffalo, NY 14214), and Alexa T. Kozak (Univ. at Buffalo, Buffalo, NY 14214)

This study examined the effect of short-term auditory deprivation on the speech production of five postlingually deafened women, all of whom were experienced cochlear implant users. Each cochlear implant user, as well as age and gender matched control speakers, produced CVC target words embedded in a reading passage. Speech samples for the deafened adults were collected on two separate occasions. First, the speakers were recorded after wearing their speech processor consistently for at least two to three hours prior to recording (implant "ON"). The second recording occurred when the speakers had their speech processors turned off for approximately ten to twelve hours prior to recording (implant "OFF"). Acoustic measures, including fundamental frequency (F_0), the first (F_1) and second (F_2) formants of the vowels, vowel space area, vowel duration, spectral moments of the consonants, as well as utterance duration and sound pressure level (SPL) across the entire utterance were analyzed in both speaking conditions. For each implant speaker, acoustic measures will be compared across implant "ON" and implant "OFF" speaking conditions, and will also be compared to data obtained from normal hearing speakers.

3aSC13. The effect of acoustic context and scale resolution on judgments of voice quality. Jody Kreiman, Bruce R. Gerratt, Mika Ito, and Norma Antonanzas-Barroso (Div. of Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, jkreiman@ucla.edu)

Anchored protocols (in which listeners match test stimuli to fixed examples of different scale levels) control response variability in voice-quality judgment tasks, and may be clinically applicable because they are also easy to use. Successful application requires that two conditions hold. First, listeners must be able to isolate the target quality in the natural pathologic voice and match the level of that quality to levels present in a different acoustic context (the anchors). Second, the anchored scales resolution must be sufficient for listeners to make consistent judgments about stimuli that fall between anchors. To examine these issues, listeners judged noise levels for 40 test voices in four tasks. Two were anchored tasks (limited-scale resolution), in which the synthetic anchor voices matched the test voice (matched acoustic environment) in one condition and differed from the test voice environment in a second condition (traditional approach). Tasks three and four used the method of adjustment (fine-scale resolution). Listeners adjusted the noise level of a synthetic voice to match either that of an identical test voice, or that of a voice that differed from the test stimulus. Implications for theory and clinical practice will be discussed. [Research supported by NIDCD.]

3aSC14. Prediction of vowel intelligibility based on acoustic characteristics in speakers with Parkinson's disease. Kate Bunton (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721)

This study investigated the contribution of acoustic features produced by three vowel contrasts (front–back, high–low, tense–lax) to the intelligibility of the contrast and overall speech intelligibility in speakers with Parkinson's disease. Multiple regression analyses were used to determine which acoustic measures or combination of measures accounted significantly for the variance in the intelligibility measures. The predictor variables in the multivariate analysis included acoustic measures thought to represent essential perceptual cues to signal the vowel contrast. These findings are discussed in relation to specific areas of production deficiency that are consistent with Parkinson's disease. [Work supported by NIH R03 DC005902.]

3aSC15. Temporal and acoustic characteristics of Greek vowels produced by adults with cerebral palsy. Antonis Botinis, Ioanna Orfanidou, Marios Fourakis (Dept. of Linguist., Faculty of Philosophy, Univ. of Athens, Athens 15784, Greece, abotinis@phil.uoa.gr), and Marios Fourakis (Univ. of Wisconsin, Madison, WI 53706)

The present investigation examined the temporal and spectral characteristics of Greek vowels as produced by speakers with intact (NO) versus cerebral palsy affected (CP) neuromuscular systems. Six NO and six CP native speakers of Greek produced the Greek vowels [i, e, a, o, u] in the first syllable of CVCV nonsense words in a short carrier phrase. Stress could be on either the first or second syllable. There were three female and three male speakers in each group. In terms of temporal characteristics, the results showed that: vowels produced by CP speakers were longer than vowels produced by NO speakers; stressed vowels were longer than unstressed vowels; vowels produced by female speakers were longer than vowels produced by male speakers. In terms of spectral characteristics the results showed that the vowel space of the CP speakers was smaller than that of the NO speakers. This is similar to the results recently reported by Liu *et al.* [J. Acoust. Soc. Am. **117**, 3879-3889 (2005)] for CP speakers of Mandarin. There was also a reduction of the acoustic vowel space defined by unstressed vowels, but this reduction was much more pronounced in the vowel productions of CP speakers than NO speakers.

3aSC16. Formant trajectory characteristics in speakers with dysarthria and homogeneous speech intelligibility scores: Further data. Yunjung Kim, Gary Weismer, and Ray D. Kent (Waisman Ctr. and Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705)

In previous work [J. Acoust. Soc. Am. **117**, 2605 (2005)], we reported on formant trajectory characteristics of a relatively large number of speakers with dysarthria and near-normal speech intelligibility. The purpose of that analysis was to begin a documentation of the variability, within relatively homogeneous speech-severity groups, of acoustic measures commonly used to predict across-speaker variation in speech intelligibility. In that study we found that even with near-normal speech intelligibility (90%–100%), many speakers had reduced formant slopes for some words and distributional characteristics of acoustic measures that were different than values obtained from normal speakers. In the current report we extend those findings to a group of speakers with dysarthria with somewhat poorer speech intelligibility than the original group. Results are discussed in terms of the utility of certain acoustic measures as indices of speech intelligibility, and as explanatory data for theories of dysarthria. [Work supported by NIH Award R01 DC00319.]

3aSC17. Analysis of modal and creaky voice quality variations. Avanti S. Shetye and Carol Y. Espy-Wilson (Inst. for Systems Res. & Elec. and Computer Eng. Dept., Univ. of Maryland, College Park, MD 20742, ashetye@umd.edu)

Voice quality, as a major vehicle of information about physical, phonological, and social characteristics of the speaker, has a vital semiotic role to play in spoken interaction [Laver (1968), Laver and Trudgill (1979)]. In the past couple of years, our lab developed an Aperiodicity/Periodicity/Pitch (APP) detector that produces a spectro-temporal profile of the periodic and aperiodic regions of the speech waveform [Deshmukh *et al.* (in press)]. To do so, the speech signal is passed through a 60-channel gamma tone auditory filterbank. The distribution of the dips occurring in the average magnitude difference function (AMDF) computed from each channel envelope is analyzed to determine periodicity and aperiodicity. Presently, the APP detector classifies both turbulent noise and irregular vocal fold vibration (creakiness) as aperiodic. In this work, we are investigating the detailed characteristics of the AMDF waveform when speech is creaky. This information is presently being used to distinguish

aperiodicity due to turbulence from aperiodicity due to creakiness. We will present results from the refined APP detector using various male and female utterances from the TIMIT database.

3aSC18. Analysis by synthesis of electrolarynx speech. Yoko Saikachi (Speech and Hearing Bioscience and Technol., Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139, yokos@mit.edu), Robert Hillman (Ctr. for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Harvard Med. School, Boston, MA 02114), and Kenneth Stevens (MIT, Cambridge, MA 02139)

The ability to accurately synthesize electrolarynx (EL) speech may provide a basis for better understanding the acoustic deficits that contribute to its poor quality. Such information could also lead to the development of acoustic enhancement methods that would improve EL speech quality. This effort was initiated with an analysis-by-synthesis approach that used the Klatt formant synthesizer to study vowels at the end of utterances spoken by the same subjects both before and after laryngectomy (normal versus EL speech). The temporal and spectral features of the original speech waveforms were analyzed and the results were used to guide synthesis and to identify parameters for modification. EL speech consistently displayed utterance-final fixed (mono) pitch and normal-like falling amplitude across different vowels and subjects. Subsequent experiments demonstrated that it was possible to closely match the acoustic characteristics and perceptual quality of both normal and EL speech with synthesized replicas. It was also shown that the perceived quality of the synthesized EL speech could be improved by modification of pitch parameters to more closely resemble normal speech. Some potential approaches for modifying the pitch of EL speech in real time will be discussed. [Funded by NIH Grant R01 DC006449.]

3aSC19. Using on-line altered auditory feedback treating Parkinsonian speech. Emily Wang (Dept. of Commun. Disord. and Sci., Rush Univ. Medical Ctr., 1653 W. Congress Pkwy., 203 Senn Chicago, IL 60612), Leo Verhagen (Rush Univ. Medical Ctr., Chicago, IL 60612), and Meinou H. de Vries (Utrechat Univ., The Netherlands)

Patients with advanced Parkinson's disease tend to have dysarthric speech that is hesitant, accelerated, and repetitive, and that is often resistant to behavior speech therapy. In this pilot study, the speech disturbances were treated using on-line altered feedbacks (AF) provided by SpeechEasy (SE), an in-the-ear device registered with the FDA for use in humans to treat chronic stuttering. Eight PD patients participated in the study. All had moderate to severe speech disturbances. In addition, two patients had moderate recurring stuttering at the onset of PD after long remission since adolescence, two had bilateral STN DBS, and two bilateral pallidal DBS. An effective combination of delayed auditory feedback and frequency-altered feedback was selected for each subject and provided via SE worn in one ear. All subjects produced speech samples (structured-monologue and reading) under three conditions: baseline, with SE without, and with feedbacks. The speech samples were randomly presented and rated for speech intelligibility goodness using UPDRS-III item 18 and the speaking rate. The results indicted that SpeechEasy is well tolerated and AF can improve speech intelligibility in spontaneous speech. Further investigational use of this device for treating speech disorders in PD is warranted [Work partially supported by Janus Dev. Group, Inc.].

3aSC20. Imitation of contrastive lexical stress in children with speech delay. Jennell C. Vick and Christopher A. Moore (Dept. of Speech and Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Box 354875, Seattle, WA 98105)

This study examined the relationship between acoustic correlates of stress in trochaic (strong-weak), spondaic (strong-strong), and iambic (weak-strong) nonword bisyllables produced by children (3;0-5;0) with normal speech acquisition and children with speech delay. Ratios compar-

ing the acoustic measures (vowel duration, rms, and $f0$) of the first syllable to the second syllable were calculated to evaluate the extent to which each phonetic parameter was used to mark stress. In addition, a calculation of the variability of jaw movement in each bisyllable was made. Finally, perceptual judgments of accuracy of stress production were made. Analysis of perceptual judgments indicated a robust difference between groups: While both groups of children produced errors in imitating the contrastive lexical stress models ($\sim 40\%$), the children with normal speech acquisition tended to produce trochaic forms in substitution for other stress types, whereas children with speech delay showed no preference for trochees. The relationship between segmental acoustic parameters, kinematic variability, and the ratings of stress by trained listeners will be presented.

3aSC21. Coarticulatory evidence in stuttered disfluencies. Timothy Arbisi-Kelm (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095)

While the disfluencies produced in stuttered speech surface at a significantly higher rate than those found in normal speech, it is less clear from the previous stuttering literature how exactly these disfluency patterns might differ in kind [Wingate (1988)]. One tendency found in normal speech is for disfluencies to remove acoustic evidence of coarticulation patterns [Shriberg (1999)]. This appears attributable to lexical search errors which prevent a speaker from accessing a word's phonological form; that is, coarticulation between words will fail to occur when segmental material from the following word is not retrieved. Since stuttering is a disorder which displays evidence of phonological but not lexical impairment, it was predicted that stuttered disfluencies would differ from normal errors in that the former would reveal acoustic evidence of word transitions. Eight speakers four stutterers and four control subjects participated in a narrative-production task, spontaneously describing a picture book. Preliminary results suggest that while both stutterers and controls did produce similar rates of disfluencies occurring without coarticulatory evidence, only the stutterers regularly produced disfluencies reflecting this transitional evidence. These results support the argument that disfluencies proper to stuttering result from a phonological deficit, while normal disfluencies are generally lexically based.

WEDNESDAY MORNING, 19 OCTOBER 2005

RAMSEY ROOM, 9:00 TO 11:45 A.M.

Session 3aSP

Signal Processing in Acoustics and Underwater Acoustics: Classifying, Localizing, Partitioning, Distributing, and Controlling

Charles F. Gaumond, Chair

Naval Research Laboratory, Acoustics Div., 4555 Overlook Ave., SW, Washington, DC 20375-5320

Contributed Papers

9:00

3aSP1. Depth classification based solely on incoherent sonar information. Donald W. Miklovic (AETC Inc., 8910 University Ctr. Ln., San Diego CA 92122, dmiklovic@sd.aetc.com)

Most work on sonar contact depth estimation has been based on deterministic, coherent propagation modeling of the sound channel, e.g., matched-field processing. This has met with limited success due to the inability to precisely predict the sound-pressure field in realistic scenarios. This paper addresses the problem of using probabilistic, incoherent information from the sonar itself for depth classification with active sonar, without having to depend on precise and accurate propagation models and ancillary environmental measurements. In particular, the problem of deciding whether or not a given contact is on the bottom based solely on the sonar data is looked at, i.e., without resorting to the use of any additional environmental measurements or predictive models. To do this, the *in situ* local bottom reverberation is used to calibrate the channel. Probability theory is explored to provide a theoretical basis for the development of a single-hypothesis decision metric to best exploit this information. The methods are tested on a combination of broadband sonar data and detailed ocean simulations. The particular metric proposed for this problem seems to be important for achieving good performance, and may be of some interest in its own right for other types of single-hypothesis decision problems.

9:15

3aSP2. Source location in an urban setting using time reversal with small sensor arrays. Mark E. Todaro (U.S. Army Armament Res., Development, & Eng. Ctr., Benet Labs, Watervliet, NY 12189-4050)

The finite-difference, time-domain (FDTD) method was used with time reversal to explore acoustic pulse source location in an urban setting with the source outside the array of sensors. The FDTD program simulates two-dimensional sound propagation from an omnidirectional point source,

including reflection and diffraction from buildings. Pressure versus time is recorded at various hypothetical sensor locations. During time-reversal playback, each sensor becomes an omnidirectional point source generating a time-reversed version of the pressure it received. A number of sensor arrays were modeled in various urban settings to identify the limitations of using time reversal to locate a non-line-of-sight source outside the array. In general, as the number of sensors and spatial extent of the array increases, the ability to locate the source increases as well. A 19-sensor hexagonal array 8 m across, for example, is capable of locating a source on the opposite side of a building. Two small four-sensor arrays 10 m from each other are adequate for around-the-corner source location. A single small array of three or four sensors spaced 1 to 2 m apart was generally insufficient to locate the source.

9:30

3aSP3. Source localization using the distribution of the interference invariant. Peter M. Daly (1710 SAIC Dr. MS 1-11-15, McLean, VA 22102, peter.m.daly@saic.com)

S.D. Chuprov's "beta" parameter, or interference invariant, has been used to characterize low-frequency broadband acoustic propagation. Baggeroer showed the interference invariant was dependent on frequency, mode grouping, and environment. Rousseff and Spindel proposed representing the invariant as a distribution rather than a single parameter, as well as demonstrating via simulation how the invariant changed with receiver depth. This talk extends Rousseff and Spindel's work to a shallow-water source localization problem. A method of incoherently combining broadband energy for passive source tracking in range and depth utilizing the distribution of the interference invariant as an observable from a single hydrophone will be illustrated, along with relevant performance metrics. [Work sponsored by SAIC.]

9:45

3aSP4. Ray-based acoustic localization applied in a highly reverberant environment. Natasha A. Chang and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 1231 Beal Rd., 2019 Lay Autolab, Ann Arbor, MI 48109; drd@engin.umich.edu)

Acoustic localization techniques for finding and identifying broadband sound sources is of interest for studying low event rate cavitation, and other hydroacoustic noise sources, in water tunnels and other reverberant environments. The measured center frequency and bandwidth of the acoustic pulse from the collapse of a small isolated cavitation bubble are both approximately 100 kHz, and the measured pulse duration is $\sim 15-20$ μs. Here, a sound projector is used to mimic such a pulse, and 16 receiving hydrophones are used to estimate the source location in the test section of a laboratory water tunnel when neither the timing of the event nor the exact signal waveform is presumed known. Fortunately, the direct-path signal arrival is generally distinct enough at each receiver for acoustic localization. The signal coda is not used in the present effort. The received direct-path signals are cross correlated between hydrophones to obtain arrival time differences. These arrival time differences are used in conjunction with a simple, straight-ray acoustic model to estimate the source location in three dimensions. Extension of this effort to use of first reflections and frequency domain modeling of the water tunnel for matched-field processing will be discussed as well. [Sponsored by ONR codes 333, 321.]

10:00

3aSP5. Rate distortion bounds for passive sonar performance. Tianzhu Meng (Dept. of Elec. and Computer Eng., Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747) and John R. Buck (Univ. of Massachusetts Dartmouth, North Dartmouth, MA 02747)

Information theory provides a novel perspective on passive sonar. This approach begins by partitioning the search space and then considers the problem of assigning an unknown source to the correct partition based on pressure observations from a hydrophone array. The goal is to assign the source to the correct partition with the minimum possible probability of error (P_e). Prior work [Buck, Proc. IEEE SAM Workshop, 2002] described necessary conditions for achieving arbitrarily small P_e as a tradeoff between SNR and the range extent, or resolution, of the partitions. This paper presents a method to extend these results using rate distortion theory to find necessary conditions for any P_e, not just arbitrarily small ones. The Gaussian channel bound sets an upper limit on the rate of information received at the array, which implies a lower bound on P_e for a given partition. For a given range resolution, this method provides the minimum achievable P_e for a given SNR, or the minimum SNR to achieve a given P_e. Examples of these bounds will be given for typical shallow water environments. [Work supported by ONR Code 321US.]

10:15–10:30 Break

10:30

3aSP6. Bistatic clutter statistics in underwater acoustics. Brian La Cour (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

For the case of a spatially separated (i.e., bistatic) source-receiver pair, the joint distribution of the time delay, bearing, and Doppler shift of acoustic clutter is derived. The clutter is assumed to be spatially uniform with a fluctuating velocity that is Gaussian distributed with zero mean and components that are uncorrelated and of equal variance. For the Doppler distribution, the source and receiver are assumed, in general, to be moving relative to one another. Unlike the monostatic case, the time delay and bearing are found to be statistically dependent. Conditioned on the time delay, the distribution in bearing exhibits two peaks appearing to either side of the source direction for times less than three direct-blast times. Later times show a single peak in the direction of the source which diminishes in relative amplitude with time. The Doppler shift distribution, con-

ditioned on time delay and bearing, is derived under the assumption that the clutter velocity dispersion is much smaller than the speed of sound. The statistical distribution is then related to bistatic reverberation, and a qualitative comparison is made between measured bistatic data and the derived statistical distributions.

10:45

3aSP7. Analysis of variable length data records for generating spectrally decomposable target-specific features. M.H. Supriya and Pillai P.R. Saseendran (Dept. of Electron., Cochin Univ. of Sci. & Technol., Kochi, India 682 022, supriya@cusat.ac.in)

The instantaneous noise data waveforms generated by surface crafts vary widely, and it is hardly possible to interpret them into a pattern that indicates the type or class of the vessel. Even for the same surface craft, the noise spectra vary widely depending on the length of the noise data waveform. The vessel-generated noises of intermittent nature, coupled with the uneven vibrations of distorted shafts as well as the wear and tear of the moving mechanical structures due to aging effects, may lead to substantial variations in the spectral profiles computed using short-term and long-term noise data waveforms. The spectrally derivable features of a class of vessel gathered from long-term data may remain the same within the class, while the time-series analysis using short-term data may yield more source specific features. Thus, the spectra computed using a time-series autoregressive estimator, capable of supplying variable length data segments, is seen to have a strong dependence on the length of the noise data segment. This paper presents an approach for generating and updating a systematic source data record comprising all the spectrally decomposable source specific features, using variable length noise data segments.

11:00

3aSP8. Wave field synthesis, adaptive wave field synthesis and ambisonics using decentralized transformed control: Potential applications to sound field reproduction and active noise control. Philippe-Aubert Gauthier, Alain Berry (Groupe d'Acoustique de l'Université de Sherbrooke, Dept. of Mech. Eng., Université de Sherbrooke, 2500 boul. Université, Sherbrooke, QC, J1K 2R1, Canada), and Wieslaw Woszczyk (Ctr. for Interdisciplinary Res. in Music, Media, and Technol., Music Faculty, McGill Univ., Montréal, QC, H3A 1E3, Canada)

Sound field reproduction finds applications in listening to prerecorded music or in synthesizing virtual acoustics. The objective is to recreate a sound field in a listening environment. Wave field synthesis (WFS) is a known open-loop technology which assumes that the reproduction environment is anechoic. Classical WFS, therefore, does not perform well in a real reproduction space such as room. Previous work has suggested that it is physically possible to reproduce a progressive wave field in-room situation using active control approaches. In this paper, a formulation of adaptive wave field synthesis (AWFS) introduces practical possibilities for an adaptive sound field reproduction combining WFS and active control (with WFS departure penalization) with a limited number of error sensors. AWFS includes WFS and closed-loop ''Ambisonics'' as limiting cases. This leads to the modification of the multichannel filtered-reference least-mean-square (FXLMS) and the filtered-error LMS (FELMS) adaptive algorithms for AWFS. Decentralization of AWFS for sound field reproduction is introduced on the basis of sources' and sensors' radiation modes. Such decoupling may lead to decentralized control of source strength distributions and may reduce computational burden of the FXLMS and the FELMS algorithms used for AWFS. [Work funded by NSERC, NATEQ, Université de Sherbrooke and VRQ.]

11:15

3aSP9. Wavefront control by wave number domain focusing. Joung-Woo Choi, Ji-Ho Jang, and Yang-Hann Kim (Ctr. for Noise and Vib. Control, Dept. of Mech. Eng., KAIST, Sci. Town, Daejon, 305-701, Korea, jwoo@kaist.edu)

This study introduces a novel method that can manipulate propagating direction of wavefronts within a selected region using source array. The idea of the proposed method stems from a method of acoustic contrast maximization, which has been used to focus sound energy within a zone of interest. This paper attempts to focus sound energy in the wave number domain, so that the sound energy of a selected region is concentrated on a desired wave number area. This makes it possible to generate a plane wave that propagates to a desired direction. The simple pure-tone case is considered to express the idea in the wave number domain; then, the method is extended to more general case, where an excitation signal has a broadband spectrum. Numerical and experimental results obtained in various conditions certainly validate that the direction of the wavefront can be manipulated for some finite region in space. [Work supported by the BK21 project initiated by Ministry of Education and Human Resources Development of Korea.]

11:30

3aSP10. A unified approach for the spatial enhancement of sound. Joung-Woo Choi, Ji-Ho Jang, and Yang-Hann Kim (Ctr. for Noise and Vib. Control, Dept. of Mech. Eng., KAIST, Sci. Town, Daejon, 305-701, Korea, jwoo@kaist.edu)

This paper aims to control the sound field spatially, so that the desired or target acoustic variable is enhanced within a zone where a listener is located. This is somewhat analogous to having manipulators that can draw sounds in any place. This also means that one can somehow see the controlled shape of sound in frequency or in real time. The former assures its practical applicability, for example, listening zone control for music. The latter provides a mean of analyzing sound field. With all these regards, a unified approach is proposed that can enhance selected acoustic variables using multiple sources. Three kinds of acoustic variables that have to do with magnitude and direction of sound field are formulated and enhanced. The first one, which has to do with the spatial control of acoustic potential energy, enables one to make a zone of loud sound over an area. Otherwise, one can control directional characteristic of sound field by controlling directional energy density, or one can enhance the magnitude and direction of sound at the same time by controlling acoustic intensity. Throughout various examples, it is shown that these acoustic variables can be controlled successfully by the proposed approach.

WEDNESDAY MORNING, 19 OCTOBER 2005

MARQUETTE ROOM, 8:45 A.M. TO 12:00 NOON

Session 3aUW

Underwater Acoustics and Acoustical Oceanography: Head Waves and Interface Waves

Peter H. Dahl, Chair

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

Invited Papers

8:45

3aUW1. Acoustic head waves in oceanic waveguides. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Environ. Technol. Lab., 325 Broadway, Boulder, CO 80305)

Wave and ray theories of head (lateral) waves are reviewed, including effects due to absorption, stratification, source directivity, and roughness of interfaces. While the wave theory is instrumental in determining head-wave amplitude and region of existence, the ray theory of head waves is particularly helpful in extending results to multilayered structures and horizontally inhomogeneous environments. Physical significance of the head wave is discussed. As far as inverse problems are concerned, head waves offer a rather direct way of measuring certain physical and geometrical characteristics of the ocean bottom. Head waves typically have smaller amplitudes than other components of the acoustic field in a waveguide, but still can be reliably identified and measured, especially where they are spatially or temporally separated from the more energetic field components. Shallow-water environments are shown to complicate the arrival structure of head waves and their modeling, while providing additional opportunities for retrieving useful geoacoustic information from head-wave data. Application of the head-wave theory to determination of geoacoustic parameters of a stratified ocean bottom, including compressional wave attenuation, is illustrated using data obtained during the Pacific Shelf-93 experiment. [Work supported in part by ONR.]

9:10

3aUW2. Head waves, diving waves, and interface waves at the seafloor. Ralph A. Stephen (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, rstephen@whoi.edu)

Brekhovskikh (1960) summarizes the system of waves that arises from reflection and refraction of spherical waves at the interface between homogeneous solid half-spaces. By eliminating the shear wave potential in one half-space, the system for fluid–solid half-spaces like the seafloor is obtained. There are two cases: one where the shear speed in the bottom is less than the compressional speed in the fluid (soft sediments), and the other where the shear speed in the bottom is greater than the compressional speed in the fluid (hard volcanic basement). This model is the basis for defining interface phenomena such as evanescent waves, head waves, pseudo-Rayleigh waves, and Stoneley/Scholte waves. If a positive gradient is introduced into the compressional and shear sound speeds in the bottom, one obtains diving waves and interference head waves (Cerveny and Ravindra, 1971). There are two types of

interface waves: pseudo-Rayleigh waves that are evanescent in the bottom but propagate in the water, and Stoneley/Scholte waves that are evanescent in both media. In multi-interface models there are of course normal modes. In actual seafloors, low speed layers, sound and shear speed gradients, and interface and volume lateral heterogeneities affect the characteristics of propagation and scattering. [Work supported by ONR.]

9:30

3aUW3. On spectral and amplitude properties of first- and zeroth- order head waves, and implications for geoacoustic inversion. Jee Woong Choi and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, choijw@apl.washington.edu)

Head waves and ground waves are strongly linked with properties of the seabed, and thus provide a useful measure for geoacoustic inversion. For a Pekeris waveguide, the ground wave can be interpreted as sequence of first-order (or classical) head waves, called a head-wave sequence [Choi and Dahl, J. Acoust. Soc. Am. **116**, 2529 (2004)]. (The term first-order originates from a ray series classification.) The presence of a sound-speed gradient in the sediment gives rise to very different effects. Here, the first-order head wave is replaced by either a lower-amplitude interference head wave whose properties are akin to a first-order head wave, or a higher-amplitude interference head wave or related diving wave whose properties are zeroth-order. There is also a shift in the dominant frequency of the first-arriving head wave in the transition regime between these arrivals. In this talk, nomenclature for head waves based on the ray series classification is suggested, parametric dependencies for distinguishing these arrivals are discussed, the criterion to estimate the frequency shift is presented, and some implications for geoacoustic parameter estimation are discussed. Synthetic data and results are obtained using the RAM parabolic equation algorithm. [Research supported by ONR Ocean Acoustics.]

9:50

3aUW4. Critical angle effects and their treatment using ray theory and mode theory. Evan K. Westwood and Clark S. Penrod (Appl. Res. Labs., Univ. of Texas, Austin, TX 78713)

The interaction of the acoustic field from a point source with a fluid half-space is examined in terms of ray theory and mode theory. For ray theory, a complex ray approach [E. K. Westwood, J. Acoust. Soc. Am. **85**, 1872–1884 (1989)] is used to find the reflected and transmitted fields as the sum of one or two eigenrays. The approach uses the method of steepest descent to solve the plane-wave integral for the fields, where the reflection and transmission coefficients are allowed to influence the locations of the saddle points and their steepest descent paths. As a consequence, saddle points are complex, and complicated processes such as the reflected lateral wave, beam displacement, and the transmitted evanescent field are included. For mode theory, the ORCA normal mode model [Westwood *et al.*, J. Acoust. Soc. Am. **100**, 3631–3645 (1996)] is used to illustrate the effects of the critical angle on the mode structure in a Pekeris waveguide. The Pekeris branch cut is shown to correspond to the lateral wave, and a method for replacing its branch line integral with a series of modes is described.

10:10–10:30   Break

Contributed Papers

10:30

3aUW5. Broadband seismo-acoustic modeling in range-dependent layered media using the parabolic equation method. David C. Calvo and Michael D. Collins (Acoust. Div. Naval Res. Lab., Washington, DC 20375-5320)

We present an efficient broadband modeling capability based on synthesizing solutions of the frequency-domain wave equation computed using the wide-angle parabolic equation (PE) method. The PE [a version of NRL's Range-dependent Acoustic Model (RAM)] takes into account recent advances for handling range-varying layer thicknesses, compression and shear waves, and surface/interfacial waves (Rayleigh, Stoneley). The technique allows for multiple sources on land, underground, or offshore. Long-range propagation examples demonstrating the computation of multiply reflected head waves are presented. Examples of using head waves and their related phenomena to perform geoacoustic inversion in multiple layers are shown.

10:45

3aUW6. Lateral waves at the ocean-bottom interface: A feasibility study for geoacoustic inversion using horizontal arrays. Allan D. Pierce, William M. Carey (Boston Univ., Boston, MA 02215, adp@bu.edu), James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and George V. Frisk (Florida Atlantic Univ., Dania Beach, FL 33004)

Lateral waves are the cw counterpart of head waves, the latter being sometimes used for geoacoustic inversion [Godin *et al.*, 1999; Berger and Buckingham, 2000; Choi and Dahl, 2004]. They should be detectable at horizontal ranges r that satisfy the criterion $H^2/(5\lambda) > r > 4h_1 h_2/\lambda$, where the various symbols refer to the water depth, height above bottom of source and receiver, and wavelength in water column. The exent of the usable ranges can be increased appreciably if an end fire (parametic) array is used to direct sound so that the beam hits the bottom interface at the critical angle. Because lateral waves are the diffraction spill-off of energy carried primarily in the bottom, they have the potential for yielding geoacoustic parameters such as bottom attenuation and sound-speed profiles. Textbook models predict waves that radiate up from the bottom at the critical angle; more realistic models suggest a distribution in angles, with the distribution varying with range and depending on the bottom sound-speed profile. Horizontal arrays yield information on this angle distribution. Developed theory uses Biot low-frequency model for the bottom and matched asymptotic expansions.

11:00

3aUW7. Observations of zeroth-order head waves in the Yellow Sea. Peter H. Dahl and Jee Woong Choi (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

As part of the 1996 joint U.S.-China Experiment in the Yellow Sea, head wave arrivals were measured in waters 75 m deep and at range ∼700 m, on a vertical line array (VLA) that spanned the water column. Head waves are strongly linked with properties of the seabed, and thus provide a useful measure for geoacoustic inversion carried out at relatively short range. Based on a time-domain and beam-forming analysis of the VLA data, the sediment sound speed and its gradient are estimated, and the conclusion is made that the observed head waves are kinematically consistent with zeroth-order in their ray-series classification. This conclusion is verified by simulation using the RAM parabolic equation algo-

rithm, and it is found that the dynamic properties of the data also point to a zeroth-order, rather than a first-order, classification. The classification in turn establishes a baseline model for the head wave energy spectrum that is compared with the measured data. It is found that the estimated energy spectrum agrees well with a zero-order model, i.e., proportional to the source spectrum, $S(f)$, modified by sediment attenuation, rather than a first-order model, $S(f)/f$. [Research supported by ONR Ocean Acoustics.]

11:15

3aUW8. Range-dependent seismo-acoustics benchmark problems. Michael D. Collins (Naval Res. Lab., Washington, DC 20375), Finn B. Jensen, Peter L. Nielsen, Mario Zampolli (NATO Undersea Res. Ctr., La Spezia, Italy), Henrik Schmidt (MIT, Cambridge, MA 02139), and William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180)

Developing the capability to accurately solve range-dependent problems in ocean acoustics has been a topic of great interest for many years. At the time of the ASA benchmark efforts [J. Acoust. Soc. Am. **87**, 1497–1545 (1990)], the elastic parabolic equation was in its early stages of development, and it was too early to consider problems involving shear waves. In recent years, there has been much progress in the development of parabolic equation and spectral techniques for range-dependent seismo-acoustics problems. The time is now ripe for extending the ASA benchmarks to cases involving shear waves. Several types of problems will be considered, including problems involving upslopes, downslopes, multiple solid layers, mode cutoff, and interface waves. [Work supported by the ONR.]

11:30

3aUW9. Extension of the variable rotated parabolic equation to problems involving variable topography. Jon M. Collis, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180, collisj@rpi.edu), and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Developing improvements in the capability of handling sloping interfaces and boundaries with the parabolic equation method has been an active area of research. The mapping solution [M. D. Collins and D. K. Dacol, J. Acoust. Soc. Am. **107**, 1937–1942 (2000)] and the variable rotated parabolic equation [Outing et al., J. Acoust. Soc. Am. **115**, 2579 (2004)] are among the recent advances in this area. The variable rotated parabolic equation, which handles variable bathymetry, is extended to solve problems involving variable topography. With this approach, the surface boundary is approximated in terms of a series of regions of constant slope. In each region, the parabolic equation is solved in a coordinate system that is aligned with the boundary. The junctions between regions are treated by overshooting the point at which the slope changes and interpolating and extrapolating to obtain a starting field in the next region. The mapping solution has been applied to some problems involving variable topography, but the variable rotated parabolic equation solution should provide greater accuracy, especially when there are relatively large slopes. [Work supported by ONR.]

11:45

3aUW10. Wave field and evanescent waves produced by a sound beam incident on a simulated sediment. Curtis F. Osterhoudt, Philip L. Marston (Dept. of Phys. and Astron., Washington State Univ., Pullman, WA 99164-2814, gardyloo@mail.wsu.edu), and Scot F. Morse (Dept. of Computer Sci., Western Oregon Univ., Monmouth, OR 97361)

When a sound beam in water is incident on a sediment at a sufficiently small grazing angle, the resulting wave field in the sediment is complicated, even for the case of flat, fluidlike sediments. The wave field in the sediment for a sound beam from a simple, unshaded, finite transducer has an evanescent component and diffractive components. These components can interfere to produce a series of nulls outside the spatial region dominated by the evanescent wave field. This situation has been experimentally simulated by using a combination of previously described immiscible liquids [Osterhoudt et al., J. Acoust. Soc. Am. **117**, 2483 (2005)]. The spacing between the observed nulls is similar to that seen in a wave-number-integration-based synthesis (using OASES) for a related problem. An analysis of a dephasing distance for evanescent and algebraically decaying components [T .J. Matula and P. L. Marston, J. Acoust. Soc. Am. **97**, 1389–1398 (1995)] explains the spacing of the nulls. [Work supported by ONR.]

WEDNESDAY AFTERNOON, 19 OCTOBER 2005

LA SALLE ROOM, 1:00 TO 2:50 P.M.

Session 3pAA

Architectural Acoustics: Selected Topics in Architectural Acoustics

Bruce C. Olson, Chair
Olson Sound Design, 8717 Humboldt Ave., North, Brooklyn Park, MN 55444-1320

Chair's Introduction—1:00

Contributed Papers

1:05

3pAA1. Time-domain edge-diffraction calculations near zone boundaries. Paul T. Calamia (Dept. of Computer Sci., Princeton Univ., 35 Olden St., Princeton, NJ 08544) and U. Peter Svensson (Norwegian Univ. of Sci. and Technol., NO-7491 Trondheim, Norway)

Time-domain edge-diffraction calculations are often used in studies of acoustic scattering from objects with rigid, simple-shaped surfaces, e.g., in computer simulations of room acoustics, noise-barrier performance, and radiation from loudspeakers. Many methods for such calculations are based on the Biot-Tolstoy solution, an explicit, continuous-time expression

for diffraction by an infinite wedge. However, this expression contains two onset singularities which make numerical computations difficult: one which is present for all source-receiver combinations, and a second which occurs only when a receiver crosses a specular-zone or shadow-zone boundary, i.e., a boundary where a geometrical-acoustics component has a discontinuity. The former singularity was eliminated by Svensson et al. using a formulation in which the diffraction impulse response is expressed as a line integral along the diffracting edge [J. Acoust. Soc. Am. **106**, 2331 (1999)]. In this paper, the latter singularity is addressed with analytical approximations of the formulation developed by Svensson et al. These approximations allow for accurate numerical computations for receivers at

or near zone boundaries, and maintain a continuous total sound field when combined with geometrical-acoustics components. The approximations will be presented, along with a demonstration of modeling software into which they have been integrated.

1:20

3pAA2. A blueprint for computational analysis of acoustical scattering from orchestral panel arrays. Thomas Burns (Starkey Labs., 6600 Washington Ave. S., Eden Prairie, MN 55344)

Orchestral panel arrays have been a topic of interest to acousticians, and it is reasonable to expect optimal design criteria to result from a combination of musician surveys, on-stage empirical data, and computational modeling of various configurations. Preparing a musicians survey to identify specific mechanisms of perception and sound quality is best suited for a clinically experienced hearing scientist. Measuring acoustical scattering from a panel array and discerning the effects from various boundaries is best suited for the experienced researcher in engineering acoustics. Analyzing a numerical model of the panel arrays is best suited for the tools typically used in computational engineering analysis. Toward this end, a streamlined process will be described using PROENGINEER to define a panel array geometry in 3-D, a commercial mesher to numerically discretize this geometry, SYSNOISE to solve the associated boundary element integral equations, and MATLAB to visualize the results. The model was run (background priority) on an SGI Altix (Linux) server with 12 CPUs, 24 Gbytes of RAM, and 1 Tbyte of disk space. These computational resources are available to research teams interested in this topic and willing to write and pursue grants.

1:35

3pAA3. Analysis of temporal decay of diffuse broadband sound fields in enclosures by decomposition in powers of an absorption parameter. Donald Bliss, Linda Franzoni (Mech. Eng., Duke Univ., Durham, NC 27708, dbb@duke.edu), Jerry Rouse (Sandia Natl. Lab., Albuquerque, NM 87185), and Ben Manning (Duke Univ., Durham, NC 27708)

An analysis method for time-dependent broadband diffuse sound fields in enclosures is described. Beginning with a formulation utilizing time-dependent broadband intensity boundary sources, the strength of these wall sources is expanded in a series in powers of an absorption parameter, thereby giving a separate boundary integral problem for each power. The temporal behavior is characterized by a Taylor expansion in the delay time for a source to influence an evaluation point. The lowest-order problem has a uniform interior field proportional to the reciprocal of the absorption parameter, as expected, and exhibits relatively slow exponential decay. The next-order problem gives a mean-square pressure distribution that is independent of the absorption parameter and is primarily responsible for the spatial variation of the reverberant field. This problem, which is driven by input sources and the lowest-order reverberant field, depends on source location and the spatial distribution of absorption. Additional problems proceed at integer powers of the absorption parameter, but are essentially higher-order corrections to the spatial variation. Temporal behavior is expressed in terms of an eigenvalue problem, with boundary source strength distributions expressed as eigenmodes. Solutions exhibit rapid short-time spatial redistribution followed by long-time decay of a predominant spatial mode.

1:50

3pAA4. Determining the effects of absorption and aperture size on sound decay in a complex coupled volume system. David T. Bradley and Lily M. Wang (Architectural Eng., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, dbradley@mail.unomaha.edu)

Coupled volume systems may consist of a large, dry main volume connected to a smaller, reverberant coupled volume through an acoustically transparent opening, known as a coupling aperture. Given the right combination of architectural parameters, a particular sound decay may be obtained in the main volume which exhibits the double-slope effect (DSE). DSE is characterized by a double-slope decay profile with a steep early decay and a shallow reverberant tail. This project focuses on two architectural parameters—absorption ratio between the main and coupled volumes, and aperture size—and how these parameters affect sound decay in a complex coupled volume system. A medium detail level computer model of a generalized imaginary concert hall has been constructed, and the computer modeling program ODEON has been used to predict the room impulse responses for several configurations of the architectural parameters in the model. Binaural auralizations have also been created in ODEON and used in subjective testing to determine the effects of parameter changes on subjective cues. Objective results, based on several DSE quantifiers, will be presented. Corresponding subjective results will also be discussed.

2:05

3pAA5. Field testing of sound absorption coefficients in a classroom. Steve Pettyjohn (The Acoust. & Vib. Group, 5700 Broadway, Sacramento, CA 95820)

Formal procedures for determining the sound absorption coefficients of materials installed in the field do not exist. However, the U.S. Air Force requested such tests to prove that the sound-absorbing material used in classrooms at Beale AFB in Marysville, CA, met the specified NRC of 0.80. They permitted the use of two layers of 0.5-in. fiberboard or 1-in.-thick fiberglass panels to meet the specified NRC rating. Post-construction tests showed reverberation times longer than expected. Unrealistic sound-absorption coefficients for room finish materials had to be used with the Sabine equation to achieve agreement between the measured and predicted reverberation time. By employing the Fitzroy equation and generally published absorption coefficients for ceiling tile, carpet, and fiberboard, the model provided excellent agreement with the measured reverberation times. The NRC of the fiberboard was computed to be 0.35, agreeing with published data. Since this did not meet project specifications, the Fitzroy model was used to learn the type and quantity of material needed to meet design goals. Follow-up tests showed good agreement between the predicted and measured reverberation times with material added, and project specifications were met. Results are also compared with the requirements of ANSI 12.60.

2:20

3pAA6. An effective method to predict sound transmission loss for multilayered/patterned partitions. Siddharth Sinha, Mark Lowry, David Tarnowski, and Daryoush Allaei (QRDC, Inc. 125 Columbia Court, Ste. 6, Chaska, MN 55318, dallaei@qrdc.com)

Existing methods to calculate the sound transmission loss (STL) of multilayered/patterned partitions require either fabrication of the partition and-time consuming experimental testing or computationally expensive numerical methodologies. This paper presents two fast and accurate methodologies developed using FEM and wave propagation theory to predict the STL curve for any geometrical/material configuration partition. In the first methodology, a FEM program calculates the frequency-dependant bending and shear-wave velocities of any configuration partition by performing a modal analysis on a two-dimensional cross section of the partition. The critical frequency of that partition is then obtained by calculating the frequency at which the wave-speed curve intersects the curve for speed of sound in air. In the second methodology, the critical frequency of the partition is used to obtain an equivalent homogeneous isotropic panel. The separation impedance (mass + bending + shear-wave impedances) of this equivalent panel is calculated to plot the STL curve, which can be considered as that of the original partition as they have similar wave dynamics and critical frequency. The model is experimentally and theoretically validated by comparing a sample of the results with several validated theoretical and experimental trends previously published in open literature.

3p WED. PM

2:35

3pAA7. Exploratory research on acoustic design of multimedia classrooms. F. D. Zong, Z. L. Zhang, J. W. Fang, Y. J. Yu (Inst. of Acoust., Zhejiang Normal Univ., China, C-321004, fdzong@zjnu.cn), and Q. Chen (Zhejiang Normal Univ., China, C-321004)

All kinds of multimedia classrooms can be found everywhere, especially in universities and schools. It is a pity that there are few multimedia classrooms with good acoustic quality, partly because these multimedia classrooms have been changed from common classrooms using some multimedia equipment. It is realized that an eligible multimedia classroom must meet many acoustic demands. These indexes are as follows: (1) how to assure the best reverberation time and frequency; (2) the integrated consideration of shape, direct sound, diffusion, early reflected sound, and ocular effect; (3) the audience's sound absorption; (4) the room's capacity. How to design an eligible multimedia classroom according to acoustic demands is discussed. This paper gives an overview of research into multimedia classroom acoustics, highlighting the importance of a good acoustic environment in universities and schools to enhance teaching and learning. The paper focuses on exploring research on reverberation time, early reflected sound, direct sound, and directivity of multimedia classrooms. There are some experience data and some advice based on examination and measurement. [Work supported by Zhejiang Nature Science Research Fund China.]

WEDNESDAY AFTERNOON, 19 OCTOBER 2005

CARVER AND HENNEPIN ROOMS,
2:00 TO 3:05 P.M.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Lily M. Wang, Chair

Univ. of Nebraska, Architectural Engineering, 245 Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182

Chair's Introduction—2:00

Invited Papers

2:05

3pID1. Hot topics: Signal processing in acoustics. Charles F. Gaumond (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375)

Signal processing in acoustics is a multidisciplinary group of people that work in many areas of acoustics. We have chosen two areas that have shown exciting new applications of signal processing to acoustics or have shown exciting and important results from the use of signal processing. In this session, two hot topics are shown: the use of noiselike acoustic fields to determine sound propagation structure and the use of localization to determine animal behaviors. The first topic shows the application of correlation on geo-acoustic fields to determine the Greens function for propagation through the Earth. These results can then be further used to solve geo-acoustic inverse problems. The first topic also shows the application of correlation using oceanic noise fields to determine the Greens function through the ocean. These results also have utility for oceanic inverse problems. The second topic shows exciting results from the detection, localization, and tracking of marine mammals by two different groups. Results from detection and localization of bullfrogs are shown, too. Each of these studies contributed to the knowledge of animal behavior. [Work supported by ONR.]

2:25

3pID2. Hot topics in biomedical ultrasound: ultrasound therapy and its integration with ultrasonic imaging. E. Carr Everbach (Eng. Dept., Swarthmore College, 500 College Ave., Swarthmore, PA 19081)

Since the development of biomedical ultrasound imaging from sonar after WWII, there has been a clear divide between ultrasonic imaging and ultrasound therapy. While imaging techniques are designed to cause as little change as possible in the tissues through which ultrasound propagates, ultrasound therapy typically relies upon heating or acoustic cavitation to produce a desirable therapeutic effect. Concerns over the increasingly high acoustic outputs of diagnostic ultrasound scanners prompted the adoption of the Mechanical Index (MI) and Thermal Index (TI) in the early 1990s. Therapeutic applications of ultrasound, meanwhile, have evolved from deep tissue heating in sports medicine to include targeted drug delivery, tumor and plaque ablation, cauterization via high intensity focused ultrasound (HIFU), and accelerated dissolution of blood clots. The integration of ultrasonic imaging and therapy in one device is just beginning, but the promise of improved patient outcomes is balanced by regulatory and practical impediments.

2:45

3pID3. Nanotech materials and how they can revolutionize underwater sensor coatings. Thomas S. Ramatowski (NAVSEA Div. Newport, 1176 Howell St., Newport, RI 02841)

Most of the materials presently used to coat or encapsulate underwater acoustic sensors have been around for decades. The performance of these materials from a variety of perspectives can be characterized as barely adequate. Thus, these limitations have forced design engineers to accept many performance versus cost and/or service lifetime compromises. These compromises now threaten the viability of some of the U. S. Navy's most cherished future hardware concepts, such as miniaturized, distributed sensors,

large-area sensors/smart skins, and high-powered acoustic sources. Recent developments in nanomaterials and polymer nanocomposite technology suggest that a major advance in marine coatings and encapsulation technology would enable many of these concepts and initiatives to reach fruition. Because of the enormous surface area that a dispersion of such particulates possesses, relatively small loadings (typically a few weight percent) in a suitable polymer matrix may exhibit orders of magnitude-scale improvements in certain physical properties by influencing the structure of the polymer matrix in ways not possible to achieve with conventional technology. This presentation will review ongoing nanotech composite research at NAVSEA division Newport, and discuss the potential for nanotechnology to trigger a revolution in naval coatings and encapsulation technology in the next 3–5 years.

WEDNESDAY AFTERNOON, 19 OCTOBER 2005 CONRAD D, 1:00 TO 3:00 P.M.

Session 3pMU

Musical Acoustics: Nonlinear Vibrations of Strings

Antoine Chaigne, Chair
ENSTA, UME, Chemin de la Huniere, 91761 Palaiseau Cedex, France

Chair's Introduction—1:00

Invited Papers

1:05

3pMU1. On the nonlinear models of the vibrating string. Alexandre Watzky (B2OA, Sci. and Technologie, Univ. Paris 12 - 94010 Creteil, France)

Vibrations of strings (threads, wires, cables . . .) are of great interest because of their various domains of application. In musical acoustics, phenomena which could have been neglected elsewhere take a particular importance since perception, which is very sensitive to nonlinear effects, is involved. Some phenomena can also be emphasized when a string is coupled to a sound-radiating structure. Reliable physical models are thus necessary to account for these phenomena, and to understand the true behavior of a vibrating string. Despite the fact that the first nonlinear models were published more than one century ago, and that accurate equations of motion can be naturally achieved within a finite displacement continuum mechanics framework, general models never received the attention they deserved, most authors focusing on particular phenomena and often settling on approximate models. This can be explained by the awkward multiplicity of the involved phenomena. The aim of this presentation is to discuss the consequences of some common assumptions and the true nature of some observed couplings. Particular attention will be paid to the preponderance of the spatial shape of the modes, which are usually underestimated with respect to their temporal form.

1:30

3pMU2. Unusual motions due to nonlinear effects in a driven vibrating string. Roger J. Hanson (Dept. of Phys., Univ. of Northern Iowa, Cedar Falls, IA 50614, roger.hanson@cfu.net)

Usual nonlinear effects observed in a sinusoidally driven vibrating string include generation of motion perpendicular to the driving plane, sudden jumps of amplitude and associated hysteresis, and generation of higher harmonics. In addition, under some conditions, there can be a rich variety of unusual, very complex motions of a point on the string, the pattern of which, together with associated harmonic (and sometimes subharmonic) content, can change dramatically with a slight change in driving frequency or sometimes with constant driving frequency and force. Intrinsic string asymmetries can also have a profound effect on the behavior. In a brass harpsichord string (wire) such asymmetries can cause a small splitting of each natural frequency of free vibration into two closely spaced frequencies (relative separation $\sim 0.2\%$ to 2%, strongly dependent on tension.) The two frequency components are associated, respectively, with the transverse motion along two orthogonal characteristic wire axes. Emphasis will be on display of optically detected unusual motion patterns of a point on the string, including an example of a pattern period of 10 s when driving at 50 Hz. See R. J. Hanson *et al.*, J. Acoust. Soc. Am. **117**, 400–412 (2005) for a more complete treatment.

1:55

3pMU3. Simulation of the nonlinear vibration of a string using the cellular automata. Shozo Kawamura (Dept. of Mech. Eng., Toyohashi Univ. of Technol., 1-1 Tempaku-cho, Toyohashi 441-8580, Japan)

In this study, the nonlinear dynamic responses of a string are simulated using the cellular automata (CA) method. The CA method is a discrete computation method, in which the analytical domain is divided into finite state variables called ecellsf and the state of each cell is updated according to local rules at every discrete time step. In this study, two types of local rule are used: the direct update rule and the reflection rule. In the case of nonlinear systems, the velocity of wave propagation is not constant and depends on the amplitude. A new treatment of the dynamic time step is proposed for the CA method considering the effect of the propagation velocity. As numerical examples, first the dynamic responses of a string with linear characteristic are simulated using the CA method. A typical resonance curve can be obtained. Second, the dynamic responses of a string with nonlinear characteristic are simulated using the proposed method. Some characteristic types of vibration can be obtained. It is concluded that the linear and nonlinear dynamic responses of a string may be obtained by simulation using the CA method.

3pMU4. Modeling of tension-modulated strings using finite difference and digital waveguide techniques. Jyri Pakarinen (Lab. of Acoust. and Audio Signal Processing, Helsinki Univ. of Technol., P.O. Box 3000, FI-02015 HUT, Finland, jyri.pakarinen@acoustics.hut.fi)

Tension modulation is a nonlinear phenomenon where large-amplitude string vibrations cause the tension of the string to vary. This results in an initial pitch glide and energy coupling between modes, causing for example the generation of missing harmonics. The presentation discusses two methods for numerical simulation of the tension modulation nonlinearity from the sound synthesis point of view. The tension modulation is assumed to propagate instantaneously along the string. In the digital waveguide approach, spatially distributed fractional delay filters are used in modulating the string length during run time. Energy-preserving techniques can be used in implementing the fractional delays. In the finite difference approach, time-domain interpolation is used to artificially modulate the wave propagation velocity. The generation of missing harmonics is implemented in the finite difference model by creating an additional excitation point at the string termination. In the waveguide model, the same effect can be obtained by using suitable approximations in the string elongation calculation. Synthesis results for both techniques are presented. Also, a brief comparison of the models with a discussion on stability issues is provided. [This research has been funded by the Academy of Finland (Project No. 104934), S3TK graduate school, and Tekniikan edistamissaatio.]

Contributed Paper

2:45

3pMU5. The non-linear motion of harp strings. Chris Waltham, Gary Chan, and Andrzej Kotlicki (Dept. of Phys. & Astron., Univ. of British Columbia, Vancouver BC, V6T 1Z1, Canada, cew@phas.ubc.ca)

The harp is an instrument with a set of plucked strings that excites the sound board directly, without the medium of a bridge. The strings are positioned at an acute angle to the plane of the sound board. The quality of the sound produced depends on the motion of the string and its interaction with the resonances of the sound board. The string and sound board motions of small and large harps have been studied using small, inexpensive infrared position sensors. The results are compared to those of a simple, nonlinear model based on the measured elastic properties of the string materials, and those of the sound board. The construction and calibration of the position sensors will also be described.

WEDNESDAY AFTERNOON, 19 OCTOBER 2005　　　　　　　　SALON E, 1:00 TO 3:40 P.M.

Session 3pNCa

NOISE-CON and Noise: Predicting and Assessing Community Responses to Noise II

Lawrence S. Finegold, Cochair
1167 Bournemouth Ct., Centerville, OH 45459-2647

Harvey S. Knauer, Cochair
Environmental Acoustics Inc., 5801 Grayson Rd., Unit 1, Harrisburg, PA 17111

Soren Pedersen, Cochair
Catseye Services, 26 Warrender Ave, Etobicoke, ON, Canada M9B 5Z2

1:00

3pNCa1. Community response to blast noise. Edward T. Nykaza, Larry L. Pater (ERDC-CERL, 2909 Farber Dr., Champaign, IL 61822), Sanford Fidell (Fidell Assoc., Woodland Hills, CA 91367), and Paul Schomer (Schomer and Assoc., Inc., Champaign, IL 61821)

Although community response to impulsive noise from military operations is usually discussed for NEPA-related purposes in terms of the prevalence of annoyance, it is managed on a local, daily basis in terms of numbers of recent complaints. Reconciling blast noise complaint rates with the annoyance predicted by dosage-effect analysis would be of considerable benefit to the Army, since it would provide insight into the dynamics of community reaction to this distinctive form of noise exposure, and put its assessment and management on a common footing. This paper describes a systematic approach to the challenges of quantifying community reaction to blast noise. [Work supported by ERDC-CERL.]

1:20

3pNCa2. Evaluation of small arms noise in a natural soundscape—Bear Butte, SD. David Braslau (David Braslau Assoc., Inc., 1313 5th St. SE, Ste. 322, Minneapolis, MN 55414)

Most studies on soundscape intrusion have been limited to moving sources. Of less concern is noise from small arms. Potential impact was predicted from a proposed large small arms facility with 10 000 rounds or more per day on the natural soundscape at Bear Butte, one of the most sacred sites of the Northern Cheyenne and other tribes. The primary impacted activity is meditation and oneness with the natural environment that can continue for several days through day and night. Non-natural sources included limited vehicles on a nearby highway and farm equipment, but few aircraft. Second-by-second ambient octave band readings were taken at 20 sites starting before sunrise. The minimum ambient level observed was 19.6 dBA but limits were encountered with a 1/2 in. microphone. Sound level data on small arms were projected from the proposed range four miles north of Bear Butte to elevated points on the Butte assuming a zero wind environment. Impact was evaluated using audibility, intrusiveness and impulse-weighted DNL. Projected levels were well above ambient. While the DNL was projected to increase by 15 dBA, this metric has

little meaning for this type of activity. Assumptions related to outdoor sound propagation, audibility and impulsive noise perception are discussed.

1:40

3pNCa3. Assessment of community reaction to proposed club racetrack. Christopher W. Menge (Harris Miller Miller & Hanson Inc., 15 New England Executive Park, Burlington, MA 01803)

The paper presents the approach and results of a study to assess the potential community reaction to a controversial proposed "club" racing and driver training facility in a rural community in the foothills of New Hampshire's White Mountains. Commissioned by a municipal foundation, the study was to be neutral and balance the widely varying interests and concerns of all town residents. The study was also to assist the development of noise elements of a town "Race Track Ordinance," in the absence of other noise regulations. Both existing community noise levels and noise emissions of comparable racing "car club" vehicles were measured, and are summarized in the paper. The local terrain, track layout and speed profiles were coded into SoundPLAN using ISO 9613-2 to compute summer-long Leq, leading to normalized DNL throughout the surrounding community as the primary metric for community reaction. Regularity of the audibility of racing on busy days was also estimated as an additi! onal metric. Based on these metrics, reaction was expected to be negative and widespread without such measures, therefore specific noise abatement measures and ordinance provisions were recommended and presented to the community at large and town officials.

2:00

3pNCa4. Queuing for quiet—the natural soundscape microstructure from a visitor perspective—II. Richard D. Horonjeff (Consultant in Acoust. and Noise Control, 81 Liberty Square Rd. #20-B, Boxborough, MA 01719) and Grant S. Anderson (Harris Miller Miller & Hanson Inc., Burlington, MA 01803)

The passage of PL100-91 "The National Parks Overflights Act of 1987" and PL106-181 "The National Parks Air Tour Management Act" chartered the National Park Service (NPS) and Federal Aviation Administration (FAA) to restore "natural quiet" to park settings, especially those with lengthy occurrences of nonindigenous sounds. Since 1991, several field investigations have produced over 400 hours of source identification logs maintained by trained, audiometrically screened observers in a number of parks. The data acquisition protocol used a forced-choice, hierarchical menu structure with "audible source state" changes timed to the nearest second. This protocol provided the basis for examining the amount of time a visitor would have to wait (T) to experience a contiguous block of indigenous-only sounds of duration (D). This paper expands on an earlier presentation at the 148th Meeting of the Acoustical Society of America, 2004, by (1) presenting data from additional parks, (2) identifying the similarities and differences in findings within and across parks, (3) demonstrating the effects of terrain shielding, atmospheric propagation, and ambient masker sound levels, and (4) providing recommendations for employing the technique to achieve various park management objectives and its introduction into predictive models.

2:20

3pNCa5. Another "new" metric for outdoor amphitheater criteria. Robert S. Berens (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Since the late 1960s, when amplified musical performances began being held there, Atlanta's open-air Chastain Park Amphitheater has been the source of enormous friction between the City, the venue's owner, and the wealthy, politically-connected residential community abutting the Park. To identify the characteristics of concert event sound to which neighbors are particularly sensitive, sound levels were monitored during 17 concerts, ranging from quiet jazz and classical performances to rock-and-roll and hip-hop. Community sound levels were monitored at 25 locations, including nine where measurements were made simultaneously inside and outside homes. The study team confirmed that low-frequency sound was the

one feature of concert-related sound that community residents identified as most problematic, but that only a relatively small proportion of the Chastain concerts resulted in any significant community annoyance. After assessing the spectral characteristics of the most troublesome concerts, a new compliance metric has been proposed to address low-frequency annoyance issues: a two-tiered exceedence threshold, based on 1-minute LEQ levels in the 63 Hz octave band measured at the rear of the amphitheater, with a concert-event "exceedence" defined to be either a 1-minute LEQ(63 Hz) level greater than 95 dB or more than ten 1-minute LEQ(63 Hz) levels greater than 90 dB.

2:40

3pNCa6. Recommended noise criteria for siting industrial facilities near residential communities with extremely low ambient sound levels. David M. Hessler (Hessler Assoc., Inc., Consultants in Eng. Acoust., 3862 Clifton Manor Pl., Ste. B., Haymarket, VA 20169)

Noise limits or criteria in many residential communities are based on an allowable increase in level above the prevailing ambient sound level, usually on the order of 3 to 5 dBA. The theory is that such a small increase results in masking of the new or existing intruding facility noise by the environmental ambient sound. However, many facilities are planned for installation or already exist in remote and very quiet ambient environments with residual levels as low as 25 dBA during the day or nighttime. This paper summarizes a study conducted to determine appropriate criteria in such remote and quiet areas where little or no noise masking will occur from ambient sources. The paper discusses both audible noise measured by the A-weighted level, low frequency noise characterized by the C-weighted level and tonal noise.

3:00

3pNCa7. Mitigation of structureborne noise nuisance. Wing P. Ko (Thorburn Assoc., Inc., 20880 Baker St., P.O. Box 20399, Castro Valley, CA 94546)

This paper presents a noise complaint case which was solved by me a few years ago in Hong Kong. A newlywed couple in the residential unit complained to the Government that the noise emitted from the pump room directly beneath their unit was very annoying, especially in the night-time period. The owner of the building was then required by the Government to mitigate the noise to the night-time statutory noise requirement within 30 days, otherwise he would be prosecuted. Ideally, the structureborne noise from the pump room could be effectively mitigated by installation of floating slab and vibration isolators under the pumps. Also, the water tanks and water pipes were required to be isolated from the walls and floor. However, this work was impossible to be completed within 30 days to stop the prosecution. Water supply to the above residents would be seriously interrupted during the construction period. As the only noise parameter of the statutory requirement was 30 minute A-weighted L_{eq}, the most effective and practical way in this exigent situation was to reduce the pump operation time within any 30 minute period to decrease the L_{eq} values. In addition, the water pipes and pumps were also required to be isolated from the walls and floor with resilient materials to break the vibration channels. These noise mitigation measures were successfully applied to the pump room before the end of the 30 days. Finally, the noise levels inside the complainant's unit were found to meet the statutory requirement. The noise complaint case was then closed by the Government.

3:20

3pNCa8. A study on construction equipment noise. Ramachandraiah Alur (Dept. of Civil Eng., IIT-Madras, Chennai-36, India, ram_alur@yahoo.com)

In India, the construction industry is one of the major industries comprising huge manpower and equipment. Currently the government is focusing on infrastructure development for projects such as expanding highways, construction of buildings, and various other activities. This paper is focused toward the noise and vibration generated by such construction equipment, and stresses the requirement for a standard code for noise

control. The sites covered are construction sites such as highway expanses, construction of various buildings, and also operations involved in ground-leveling techniques. It is noticed that the noise levels from various sources exceed the acceptable limits generally suggested. Some of the pile-driving operations generate noise levels of the order of 108 dBA and pneumatic tools generate 85 to 90 dBA in the site. A questionnaire survey is also carried out by many of these companies to evaluate the opinion of the user from the noise and vibration viewpoint. The majority (59%) complains that noise and vibration have to be mitigated. However, as Indian industries have to comply with many international standards, it becomes imperative to arrive at acceptable standards. In this context a discussion has been carried out.

WEDNESDAY AFTERNOON, 19 OCTOBER 2005   SALON F, 12:45 TO 2:25 P.M.

Session 3pNCb

NOISE-CON and Noise: Products for Noise Control

Stephen I. Roth, Chair
Roth Acoustical Assoc., 2352 Norton Rd., Pittsburgh, PA 15241

12:45

3pNCb1. Acoustical prediction methods for heating, ventilating, and air-conditioning (HVAC) systems. S. R. Ryherd and L. M. Wang (Architectural Eng. Program, Univ. of Nebraska–Lincoln, 1110 S. 67th St., Omaha, NE 68182)

The goal of this project is to compare and contrast various aspects of acoustical prediction methods for heating, ventilating, and air-conditioning (HVAC) systems. The three methods include two commonly used software programs and a custom spread sheet developed by the authors based on the American's Society of Heating, Refrigeration, and Air-conditioning Engineers (ASHRAE) Applications Handbook. Preliminary results indicate relatively good agreement between the three methods analyzed. The degree of disparity is predominantly effected by the assumptions required by the end user. Research methods and results will be presented. This project provides a greater understanding of these acoustical prediction methods and their limitations.

1:05

3pNCb2. Two case studies: QuietRock QR-530 drywall panels in new and remediated multifamily construction. Brian D. Tinianov (Quiet Solution, 522 Almanor Ave., Sunnyvale, CA 94085)

Reliable acoustical isolation continues to be a high risk element of contemporary multifamily construction. Traditional construction techniques, offering potentially high acoustical performance, exist but may be compromised during typical construction or occupation. This paper presents two case studies using a new class of construction material—drywall panels employing constrained layer damping. QuietRock QR-530 damped gypsum board panels are used in a new construction project and as part of a remediation treatment. In a first study, QR-530 panels were used as a drywall alternative in a 2×6, semistaggered, framed wall separating luxury condominiums. Field evaluation per ASTM E36 revealed a normalized noise isolation class of 56. In a second case study, a single layer of QuietRock was applied directly to an existing single stud assembly in a resort hotel. Before and after testing yielded a change of 14 points, raising the noise isolation class from 36 to 50. This paper reviews the details of the test cases and the underlying physical principals of the subject materials.

1:25

3pNCb3. Case studies of pre-engineered and manufactured sound isolation rooms for music practice and radio broadcast. Ron N. Probst and Dann Rypka (Wenger Corp., 555 Park Dr., Owatonna, MN 55060)

Pre-engineered and manufactured sound isolation rooms were developed to ensure guaranteed sound isolation while offering the unique ability to be disassembled and relocated without loss of acoustic performance. Case studies of pre-engineered sound isolation rooms used for music prac-

tice and various radio broadcast purposes are highlighted. Three prominent universities wrestle with the challenges of growth and expansion while responding to the specialized acoustic requirements of these spaces. Reduced state funding for universities requires close examination of all options while ensuring sound isolation requirements are achieved. Changing curriculum, renovation, and new construction make pre-engineered and manufactured rooms with guaranteed acoustical performance good investments now and for the future. An added benefit is the optional integration of active acoustics to provide simulations of other spaces or venues along with the benefit of sound isolation.

1:45

3pNCb4. Combination sound and vibration isolation curb for rooftop air-handling systems. Thomas S. Paige (Kinetics Noise Control, Inc., Vibron Products Group, 1720 Meyerside Dr., Mississauga, ON, Canada ON L5T 1A3)

This paper introduces the new Model ESSR Sound and Vibration Isolation Curb manufactured by Kinetics Noise Control, Inc. This product was specially designed to address all of the common transmission paths associated with noise and vibration sources from roof-mounted air-handling equipment. These include: reduction of airborne fan noise in supply and return air ductwork, reduction of duct rumble and breakout noise, reduction of direct airborne sound transmission through the roof deck, and reduction of vibration and structure-borne noise transmission to the building structure. Upgrade options are available for increased seismic restraint and wind-load protection. The advantages of this new system over the conventional approach of installing separate duct silencers in the room ceiling space below the rooftop unit are discussed. Several case studies are presented with the emphasis on completed projects pertaining to classrooms and school auditorium applications. Some success has also been achieved by adding active noise control components to improve low-frequency attenuation. This is an innovative product designed for conformance with the new classroom acoustics standard ANSI S12.60.

2:05

3pNCb5. Insertion loss class—Enhancing the single number rating system for duct silencers. John Duta (Dumont, NJ) and Nathan Clute (Industrial Acoust. Co., 1160 Commerce Ave., Bronx, NY 10462)

Single number rating systems have been used successfully to select architectural acoustics products for many years. Recently a single number rating system was proposed to evaluate sound attenuators in HVAC applications. The proposed system does not include the very low frequencies. This paper discusses the addition of the 31 Hz octave band to the proposed standard contour for calculating insertion loss class (ILC). A number of examples are shown using actual published silencer test data how the system can be applied to common HVAC noise problems.

Session 3pNCc

NOISE-CON and Noise: Product Noise and Vibration Control − Case Studies III

Karl B. Washburn, Chair
John Deere, P.O. Box 538 Dubuque, IA 52004-0538

1:00

3pNCc1. Optimization of isolators for vibration suppression in an integrated motor propulsor. Kiran Konde and Marty Johnson (Vib. and Acoust. Labs, Virginia Tech, 140 Durham Hall, VA 24061-0238)

Integrated motor propulsors are ducted propulsors for underwater autonomous vehicles. The noise and vibration created by such propulsors is of great concern for stealth applications. In the present paper, a general approach to the design and optimization of passive elastomers embedded with dynamic vibration absorbers for reducing the vibration of such ducted propulsors will be presented with some results showing the optimal location and dimensions of the isolators. The strong fluid structure coupling between the propulsor, its surrounding duct, annular fluid and the surrounding water is taken into account while designing the isolators. A genetic algorithm with objective function as mean square velocities is used to get the optimal location and optimal size of these isolators. Finally the results obtained from the experiments for the performance of these designed isolators under the operating conditions of the propulsor will be presented. Also the approach adopted will be compared with the vibration control methods available in literature for such applications.

1:20

3pNCc2. Development of a low profile acoustical door for use on racks and cabinets for the information technology industry. Michael D. O'Connell and William James Anderl (IBM Rochester Acoust. Lab., IBM Corp., Hwy. 52 and 37th St. N., Rochester, MN 55901)

This paper presents the design of 19 inch rack acoustical doors balancing acoustical attenuation, airflow impedance and distribution in a short depth by combining air foil technology with acoustic baffle design. Design optimization was done utilizing fluid flow analytical modeling and verified with a air flow bench and an acoustical rack door test fixture. Higher heat loads in rack mounted computer equipment drive higher cooling requirements. In order to provide air cooling solutions, higher volumetric air flow is required resulting in higher acoustical noise levels. These noise levels can result in noise levels that are unacceptable to the customer. Acoustical doors lower noise levels but are prone to high flow impedance, uneven flow distribution and large physical depth. High impedances require higher air moving device speeds to offset the lost volumetric air flow. This decreases the effective acoustical attenuation. Various rack modules have different inlet and outlet air flow locations making the distribution of the air from the door (front) or into the door(rear) important. Solutions to these problems usually require large depths in order to provide blockage of line of site and gradual air flow lines to keep impedance low and provide even distribution of the air.

1:40

3pNCc3. Case study: An appraisal of wet/dry vacuum noise. A. E. Sandman, Z. Tao, and D. W. Herrin (Univ. of Kentucky, 151 RGAN Bldg., Lexington, KY 40506-0503)

An investigation was conducted to assess the difference in sound between a high end wet/dry vacuum and a consumer version. It was noted that the high end model had a convoluted flow path as well as foam in the exhaust. The sound power for each vacuum was measured. Additionally, the sound power for the high end model was measured with and without the foam in the exhaust. As anticipated, the higher end model was better at higher frequencies due to the absorbing material in the exhaust. Additionally, a crude sound quality assessment was conducted. The jury appeared to be more sensitive to the high frequency sound.

3p WED. PM

Session 3pNCd

NOISE-CON and Noise: Measesurement of Product Noise Emissions II

Jeff G. Schmidt, Chair
ViAcoustics, 2512 Star Grass Cir., Austin, TX 78745

1:00

3pNCd1. The challenges of lawnmower noise reduction. Carol J. Drutowski and Keith Fetzer (The Toro Co., 8111 Lyndale Ave. S., Bloomington, MN 55420)

The European Noise Directive (2000/14/EC) became effective in the year 2000, limiting the allowable noise level for lawnmowers sold in the European Union. Noise level limits are based on cutting width. The Directive requires that manufacturers meet these limits and include statistical uncertainties, then declare their noise levels and have these results certified by an independent third party agency. There are two major challenges with this Directive. First by taking statistical uncertainties into account, this creates a limit lower than the published limit. Second is to lower the sound power level while still maintaining acceptable performance. The physical phenomena that generate the unwanted noise of a lawnmower are also used to perform work. For rotary lawnmowers, a dominant noise source is the cutting deck assembly, which is required to lift and cut grass, then dispose of the clippings. To cut grass with a good quality of cut

appearance, the blades are required to turn at a very high tip speed to generate the necessary airflow. For lawnmowers, noise levels and unit performance are intrinsically linked. This has frustrated our industry for over 20 years.

1:20

3pNCd2. Effect of discharge duct geometry on centrifugal fan performance and noise emission. David A. Nelson (Nelson Acoust., P.O. Box 879, Elgin, TX 78621), William Butrymowicz, and Christopher Thomas (Orbital Technologies Corp., Madison, WI 53717)

Non-ideal inlet and discharge duct geometries can cause significant changes to both the aerodynamic performance ("fan curve") and specific sound power emission of a fan. A proper understanding of actual installed performance, as well as a good estimate of the system backpressure curve, is critical to achieving flow and acoustic goals as well as other criteria such as power consumption, mass and volume. To this end a battery of ISO 10302 tests was performed on a blower assembly which supports the Advanced Animal Habitat, being developed by ORBITEC for deployment on the International Space Station. The blower assembly consists of (4) identical centrifugal fans that, amongst themselves and across two proto-types, incorporated several discharge geometries. The inlet geometries were identical in all cases. Thus by comparing the dimensionless pressure-flow and noise emission characteristics across the cases, significant insight into the nature and potential magnitude of these effects is gained.

1:40

3pNCd3. An automated system for the acoustical and aerodynamic characterization of small air moving devices. Jeff G. Schmitt (ViAcoust., 2512 Star Grass Circle, Austin, TX 78745), David A. Nelson, (Nelson Acoust., Elgin, TX 78621), and John Phillips (Acoust. Systems, Austin, TX 78745)

A plenum fixture for use in the measurement of acoustic emissions of air moving devices used to cool electronic equipment under the actual aerodynamic conditions of operation has been standardized in ISO 10302 and ANSI S12.11. This fixture has proven to be a valuable tool for use in the characterization of these devices. However, as many in industry have discovered, the construction of the plenum to the standardized specifications can quite complex, and the use of the plenum to fully characterize air

moving devices can be quite laborious and tedious. Under contract to the NASA Glenn Research Center, which has a significant interest in the acoustic emissions of the air moving devices it uses to cool racks and payloads that are installed on the International Space Station, the authors have developed a fully automated fan test plenum that operates under software control. This plenum has been developed to facilitate rapid acoustic characterization of fans and other air moving devices, both inde-pendently and when operating into real world inlet conditions, obstruc-tions and aerodynamic loads. The plenum slider has been calibrated to allow full development of fan curve data in parallel with acoustic emission data.

2:00

3pNCd4. Acoustic emissions verification testing of International Space Station experiment racks at the NASA Glenn Research Center Acoustical Testing Laboratory. James C. Akers, Paul J. Passe (Analex Corp., NASA John H. Glenn Res. Ctr. at Lewis Field, 2100 Brookpark Rd., M.S. 333-1, Cleveland, OH 44135), and Beth A. Cooper (Acoust. Testing Lab., NASA John H. Glenn Res. Ctr. at Lewis Field, Cleveland, OH 44135)

The Acoustical Testing Laboratory (ATL) at the NASA John H. Glenn Research Center (GRC) in Cleveland, OH, provides acoustic emission testing and noise control engineering services for a variety of specialized customers, particularly developers of equipment and science experiments manifested for NASA's manned space missions. The ATL's primary cus-tomer has been the Fluids and Combustion Facility (FCF), a multirack microgravity research facility being developed at GRC for the USA Labo-ratory Module of the International Space Station (ISS). Since opening in September 2000, ATL has conducted acoustic emission testing of compo-nents, subassemblies, and partially populated FCF engineering model racks. The culmination of this effort has been the acoustic emission veri-fication tests on the FCF Combustion Integrated Rack (CIR) and Fluids Integrated Rack (FIR), employing a procedure that incorporates ISO 11201 ("Acoustics–Noise emitted by machinery and equipment–Measurement of emission sound pressure levels at a work station and at other specified positions–Engineering method in an essentially free field over a reflecting plane"). This paper will provide an overview of the test methodology, software, and hardware developed to perform the acoustic emission verification tests on the CIR and FIR flight racks and lessons learned from these tests.

WEDNESDAY AFTERNOON, 19 OCTOBER 2005

ROCHESTER ROOM, 1:30 TO 2:30 P.M.

Session 3pNS

Noise and NOISE-CON: Noise Exposure and Impact

Michael R. Stinson, Chair
National Research Council, Inst. for Microstructural Sciences, Ottawa, ON, K1A 0R6, Canada

Contributed Papers

1:30

3pNS1. Characterization of vibration and noise exposure in Canadian Forces armored vehicles. Ann M. Nakashima, Matthew J. Borland, and Sharon M. Abel (Defence Res. and Development Canada–Toronto, 1133 Sheppard Ave. West, Toronto, ON M3M 3B9, Canada)

A study to characterize the vibration and noise exposure in several Canadian Forces (CF) armored vehicles is in progress. Measurements of whole-body vibration and ambient noise levels are being made in the LAV III, Bison, Coyote, and M113 vehicles at three different positions: driver, crew commander, and passenger bench (or navigator seat in the case of the Coyote). The measurements are being made while the vehicles are idling,

driven over rough terrain, and driven at a high speed on paved highways. There are several standards that provide guidance on the measurement and assessment of whole-body vibration, but they are difficult to implement in practice, particularly in adverse environments. The whole-body vibration measurements in this study are particularly difficult to interpret in the case of the crew commander, who often stands on the seat, and the passenger, who is seated but unrestrained by a seatbelt. The preliminary results-suggest, that according to the International Organization for Standardiza-tion guidelines (ISO 2631-1:1997), there may be potential health risks for the driver and passenger after driving on rough terrain for less than 10 min. Noise levels were as high as 100 dBA during high-speed highway driving.

1:45

3pNS2. Noise exposure in marching bands. Joseph Keefe (Ostergaard Acoust. Assoc., 315 Grist Mill Dr., Basking Ridge, NJ 07920)

Previous studies involving orchestras have shown that music ensembles can produce hazardous noise levels. There are no similar data for marching bands and pep bands. In order to evaluate the noise levels produced by marching and pep bands, 1/3-octave-band sound-pressure levels were measured while these groups rehearsed and performed. Data were collected while marching with the bands to ensure a realistic environment. Comparing these data to OSHA and NIOSH criteria, marching and pep band exposures often exceed safe values. For typical exposures, OSHA doses range from 11% to 295%, while NIOSH doses range from 35% to 3055%. Exposures that would be considered hazardous in the workplace are common in marching and pep bands; students and band directors should take steps to recognize the risk posed by various instruments and various locations, and should implement hearing conservation efforts.

2:00

3pNS3. Why noise-induced hearing loss of industrial workers is dramatic while that of similarly assessed musicians has been described as trivial. David Bies (Resonance Technol. Intl., 103 Callaghan St., Mooroobool QLD, Australia)

Criteria for noise exposure considered acceptable for hearing protection are based upon industrial experience, yet these same criteria do not describe the experience of musicians. Investigation of the physics of the human ear reveals a basic design compromise that explains this anomaly. Acoustic stimulation is encoded in the velocity response of the basilar membrane, which makes possible the use of damping control to achieve the dynamic range of the ear. The use of damping control for this purpose without unacceptable distortions is possible if damping is slowly varying.

The ear is free running and guided by previous instruction, making it vulnerable to loud impulsive sounds. To protect the ear the aural reflex is provided, but this protection is limited to frequencies below about 1 to 2 kHz. In the natural environment this design compromise is satisfactory, but in the industrial environment loud impulsive sounds are common and the compromise fails. It is to be noted that impulsive sounds of high frequency and level for which the ear has no defense, and which are not characteristic of music, are averaged to zero using standard assessment procedures.

2:15

3pNS4. Noise around electrical energy substations. Fabiano B. Diniz (Lab. of Acoust., Mech. Eng. Dept, Federal Univ. of Parana 81531-990 Curitiba, PR, Brazil) and Paulo H. Zannin

This survey is intended to characterize the noise impact due to electrical energy substations in the city of Curitiba over the population living in their vicinity. This impact has been studied with the aid of a computational tool capable of mapping the acoustical field of substations and their vicinity. Several factors have been considered in this survey: sound power of the transformers; vehicle flow on the surrounding roads; positioning of the firewalls, of the buildings and of the walls; terrain topography. Seven substations have been analyzed, and acoustical maps have been traced for each of them for the day and night periods. With these maps it was possible to visualize what was the incident noise levels on the building facades. The predicted noise levels have been compared to the environmental legislation of the noise emissions in effect in the city. Finally the construction of noise barrier walls surrounding the transformers has been simulated, via the software, for some cases, in order to evaluate the reduction on the acoustical discomfort caused by the transformers over the neighborhood population.

ASA/NOISE-CON 2005 Plenary Session, Business Meeting and Awards Ceremony

William A. Yost, President
Acoustical Society of America

Gerald C. Lauchle, President
Insitute of Noise Control Engineering

Welcoming Remarks by President of INCE-USA

Presentation of INCE-USA Awards

INCE Student Paper Prizes

Martin Hirschorn IAC Student Project Prize

Introduction by President of Acoustical Society of America

Business Meeting of the Acoustical Society of America

Presentation of Certificates to New Fellows

Wolfgang Ahnert
Kim C. Benajmin
Charles C. Church
Robert L. Clark
James P. Cottingham
C. Craig Formby
Samir N. Y. Gerges

Cheryl Gracewski
Kenneth W. Grant
Paul Johnson
James M. Kates
Qing Huo Liu
Peter Mapp
Peggy B. Nelson
Jixun Zhou

Presentation of Certificates to Science Writing Award Recipients

Science Writing Award in Acoustics for Journalists
Kate Ramsayer
''Infrasonic Symphony'' published in the 10 January 2004 issue of *Science News*

Science Writing Award in Acoustics for Journalists
Declan Butler
''Sound and Vision'' published in the 5 February 2004 issue of *Nature*

Presentation of Acoustical Society Awards

Rossing Prize in Acoustics Education to Katherine S. Harris

Pioneers of Underwater Acoustics Medal to Henrik Schmidt

Silver Medal in Animal Bioacoustics to James A. Simmons

Silver Medal in Speech Communication to Katherine S. Harris

Trent-Crede Medal to Jerry H. Ginsberg

3p WED. PM

ACOUSTICAL SOCIETY OF AMERICA

PIONEERS OF UNDERWATER ACOUSTICS MEDAL

Henrik Schmidt

2005

The Pioneers of Underwater Acoustics Medal is presented to an individual irrespective of nationality, age, or society affiliation, who has made an outstanding contribution to the science of underwater acoustics, as evidenced by publication of research in professional journals or by other accomplishments in the field. The award was named in honor of five pioneers in the field: H. J. W. Fay, R. A. Fessenden, H. C. Hayes, G. W. Pierce, and P. Langevin.

PREVIOUS RECIPIENTS

Harvey C. Hayes	1959	Fred N. Spiess	1985
Albert B. Wood	1961	Robert J. Urick	1988
J. Warren Horton	1963	Ivan Tolstoy	1990
Frederick V. Hunt	1965	Homer P. Bucker	1993
Harold L. Saxton	1970	William A. Kuperman	1995
Carl Eckart	1973	Darrell R. Jackson	2000
Claude W. Horton, Sr.	1980	Frederick D. Tappert	2002
Arthur O. Williams	1982		

CITATION FOR HENRIK SCHMIDT

⋯ for pioneering contributions in numerical modeling and at-sea experiments in underwater acoustics.

MINNEAPOLIS, MINNESOTA • 19 OCTOBER 2005

Henrik Schmidt has told me that the luck of being at the right place at the right time is a family trait. Born August 13, 1950 in Aabenraa, Denmark, a town near the German border, he earned his Ph.D in Structural Mechanics at the Technical University of Denmark working on nondestructive testing and acoustic emission with the intention of becoming a world-renowned, soccer playing, bridge builder. However, adhering to the family trait, he followed the example of his grandfather who, as a young man had been forced into the German Army in World War I. There he distinguished himself as the only soldier from the first to the last day of the war who was never promoted. However, he accidentally stumbled over a secret French telephone line and was awarded the Iron Cross. Henrik continued this accidental military tradition when, during his eight months of required Danish military service, he was ordered to fix up some underwater acoustics ray trace models. A quarter of a century later, he is receiving the ASA's Pioneers of Underwater Acoustics Medal, which is silver rather than iron.

For the four years after his military service, Henrik was a research fellow, first at the Technical University, then at Risoe National Laboratories (Niels Bohr's lab) modeling stress wave propagation and scattering. Next, he took a research position at the SACLANT Undersea Research Centre (renamed NATO Undersea Research Centre) for five years where he had another accidental push toward a certain research direction. I had just left the Centre and upon his arrival, Henrik was asked by Finn Jensen, then the head of the Modeling Group, to continue with a seismic-acoustic code I had been working on at that time. Finn told me that Henrik studied the programs, decided they were junk, tossed them in the garbage and started from scratch. After a short but intense period, he laid the foundation for his subsequent pioneering effort, the OASES/SAFARI set of full solutions to the seismo-acoustic wave equations. In this activity, he received much encouragement from Hank Kutschale of Columbia/Lamont. Aside from quantifying complicated acoustic phenomena, the models he developed at SACLANTCEN and afterwards were used by underwater acousticians as benchmark solutions over the next quarter century.

In 1987 he left SACLANTCEN with an international reputation in Underwater Acoustics for an Associate Professorship in the Massachusetts Institute of Technology (MIT) Department of Ocean Engineering. His rise to tenured Professor, Associate Research Director of the MIT Sea Grant Program and Department Head was as unambiguous and as fast as such transitions can possibly happen at MIT. He has graduated about 20 Ph.D. students from ten countries. He has also coauthored the now standard text, *Computational Ocean Acoustics.*

Henrik's initial seminal contribution of propagation modeling of elastic layered media was followed by a series of papers with colleagues and students that changed the concept of attainable accuracy from modeling seismo-acoustic phenomenon and from modeling and devising new signal processing methods. These papers dealt with such topics as scattering from rough interfaces and other complex structures that are embedded in an elastic waveguide, modeling ambient noise in complex ocean environments, devising (nonlinear) inversion methods, developing robust matched field processing (MFP) methods, studying radiating structures in acoustic channels, and modeling 3-D waveguide reverberation. The above body of work showed his gradual shift to predicting phenomena that could be experimentally confirmed or to developing methods that would lay the foundation for robust measurements and signal processing in complex environments.

Henrik's transition to intense experimental activity started with the Arctic, a fluid waveguide bounded by rough, elastic layered media. He participated in Arctic expeditions and ultimately was Chief Scientist of an Office of Naval Research (ONR) Sea Ice Mechanics Initiative. In this kind of endeavor, initial success is often claimed if everyone returns alive and so a Chief Scientist of an Arctic Expedition is typically someone who has devoted his/her total career to expedition-like experiments. This accomplishment in the context of the totality of Henrik's prolific record is itself quite amazing. Of course, the important outcome was that the detailed theory of propagation and scattering in the presence of rough, fluid loaded elastic plates that he had developed was confirmed with Arctic data. Noise was also an area where his scientific trajectory was from theory/modeling to experiment; some of this work on cracking ice was modeled from his general theoretical study of Doppler in a waveguide.

In the last decade Henrik has been conducting precise experiments studying the properties of the ocean and ocean bottom while pioneering the combined use of emerging Autonomous Underwater Vehicle (AUV) and acoustic communication technology, synthetic aperture processing and the latest data assimilation techniques. Embedded in these efforts is a very exciting application to developing procedures using multiple AUV's for mine hunting.

Much of his latest work has been done with SACLANTCEN in Italy, where he and his wife Satu maintain a second home in a small mountain village. When not in Italy, Satu is a social worker while also working on her Master's degree in Sociology. When in Italy, they graciously entertain family, friends, students, colleagues or local villagers. There, the multi-lingual conversations are as likely to be about the olive harvest, or denying knowledge of locations of the valuable porcini (mushrooms indigenous to the region) as they are to be about acoustics or AUV's. Wherever they are, they always have the pleasure of an active family life stimulated by their three sons: Mads, a robotics engineer, Jakob, a mechanical engineering student and Erik, a high school senior.

Henrik Schmidt's research career in Underwater Acoustics has followed a trajectory of ever increasing accurate modeling of complex ocean acoustic phenomenon, and experimentation, of using his knowledge gained from his modeling studies to devise robust, physics-based signal processing and inversion methods and finally of combining all of the above with emerging ocean technologies and developments in ocean sciences to bring Underwater Acoustics to a level of accuracy not previously thought possible. His work with other colleagues, education of students, activity as a Chief Scientist on a variety of complex Arctic and ocean experiments, service in ASA is unambiguous testimony to his leadership in our community. He is truly a Pioneer in Underwater Acoustics.

WILLIAM A. KUPERMAN

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal

in

Animal Bioacoustics

James A. Simmons

2005

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics. The previous recipient is Whitlow W. L. Au—1998.

CITATION FOR JAMES A. SIMMONS

··· for contributions to understanding bat echolocation

MINNEAPOLIS, MINNESOTA • 19 OCTOBER 2005

James A. Simmons is an eminent scholar and pioneer in the field of Biosonar, and his research includes sophisticated behavioral and neurophysiological studies of sound processing in the echolocating bat. For over three decades, he has been in the forefront of bat echolocation research, spanning the time between his graduate research to the present.

Jim obtained his bachelor's degree from Lafayette College in Easton, Pennsylvania in 1965 with a double major in Psychology and Chemistry. He then earned a master's degree in 1968 and a Ph.D. in Psychology in 1969 from Princeton University. Jim's graduate research involved studies of echolocation in bats, under the mentorship of E. Glen Wever, one of the giants in physiological acoustics. At that time, acceptance of the processes underlying spatial perception by echolocation was not universal, and one of the exciting moments of his graduate training came when a skeptical Nobel Laureate, Georg von Békésy, on one of his periodic visits to Wever's lab, came to see the behaving bats in ''Building B.'' The demonstration that Jim conducted not only convinced Békésy that bats echolocated but that they also use echo delay to estimate target distance. It was not until some years later that Jim found out that this was a set-up engineered by Wever and by Donald R. Griffin, who was then at the Rockefeller University, to convince Békésy about the bat's extraordinary use of echolocation to determine target range. Jim's dissertation was entitled ''Perception of target distance by echolocating bats.'' After receiving his Ph.D., Jim remained at Princeton University as a National Institutes of Health (NIH) Postdoctoral trainee for two years.

Jim continued his research on bat echolocation after he moved to Washington University in St. Louis, Missouri in 1971, as an assistant professor in the Neural Science Program that was housed in the Psychology Department. Between 1980 and 1984, he taught and conducted research as a professor in the Department of Biology and Institute of Neuroscience, University of Oregon in Eugene, Oregon. In 1984, Jim moved to Brown University in Providence, Rhode Island, where his wife Andrea had been hired as an assistant professor in the Psychology Department. Both are now professors in the Psychology and Neuroscience Departments at Brown University, respectively

Jim developed methods for conducting psychophysical studies of sonar processing by bats, and researchers around the world have adopted these methods to address a wide range of research questions. Jim was the first to use electronically delayed playbacks of the bat's echolocation signals to simulate target echoes for the study of perception in bats. He used such ''phantom'' target echoes to measure, for the first time, target range difference discrimination thresholds of echolocating bats. Jim studied sonar ranging performance in bats by determining the minimum difference in echo delay that an echolocating animal can discriminate. With this paradigm, Jim estimated that bats can discriminate a range difference of approximately 1 centimeter, corresponding to an echo delay difference of approximately 60 microseconds.

Jim continued to study sonar ranging performance in echolocating bats, and in the late 1970's, he introduced a new behavioral task, requiring the bat to discriminate a sonar target returning echoes at a fixed delay from one returning echoes that alternated between two delays. In this experiment, Jim found that the echolocating bat can discriminate a jitter in echo delay in the submicrosecond range, corresponding to a change in target distance of less than 0.1 mm. This result, originally published in 1979 in *Science*, ''Perception of echo phase information in bat sonar,'' demonstrated astonishing ranging accuracy by the echolocating bat. Many researchers in the field challenged the report, because they asserted it was not biologically possible for the bat's sonar system to discriminate such small time differences at ultrasonic frequencies. Jim continues to work on this problem to explore biological processes that could support sensitivity to small changes in echo delay.

Through clever behavioral experiments, Jim demonstrated time-varying gain in the sonar receiver of echolocating bats. The hearing sensitivity of the big brown bat decreases before each sonar pulse is emitted and then recovers in a logarithmic fashion to compensate for the two-way transmission loss of sonar returns, thereby maintaining a constant echo sensation level over a distance of about 1.5 meters. This is functionally important to the bat, as it stabilizes the bat's estimate of echo arrival time, its cue for target distance.

In addition to Jim's extraordinary contributions to our understanding of perception by sonar, he has conducted groundbreaking neurophysiological experiments in echolocating bats. One of his

manuscripts published with co-authors Albert Feng and Shelley Kick in *Science* (1971), entitled ''Echo detection and target-ranging neurons in the auditory system of the bat, *Eptesicus fuscus*,'' had a profound impact on the study of the neurophysiology of echolocating bats. This paper describes the response properties of auditory neurons in the bat central nervous system that show facilitated responses to pairs of sounds separated by a limited range of biologically relevant delays. These neurons exhibit the response characteristic known as ''echo delay-tuning'' or ''range-tuning,'' which could provide the neural substrate for target distance coding. The published report on this population of delay-tuned neurons by Simmons and colleagues preceded the first papers by Nobuo Suga and his group, who have since published widely on this topic.

There are so many firsts associated with Jim's research that it is difficult to adequately convey the magnitude of his impact on the field of biosonar. His experiments have been theoretically rooted, innovative, and carefully executed. In the past five years, he has used new methods for making thermal infrared video recordings of bats flying in natural situations. He developed a stereo video viewing system that lets him observe bats in 3D and listen to their sounds while they behave. These studies have led to new discoveries that challenge our understanding of echolocation behavior in bats.

Jim has won many awards and honors in his career such as the National Institutes of Health Research Scientist Development Award (9/1/84–7/31/95), and the James McKeen Cattel Prize, Lafayette College Psychology Department in 1965. He was elected a fellow of the Acoustical Society of America in 1996 and a fellow of the American Association for the Advancement of Science in 2000. Jim presented the tutorial lecture at the 143rd Meeting of the Acoustical Society of America in Pittsburgh, PA, June 2002. He is also one of the associate editors in Animal Bioacoustics for the *Journal of the Acoustical Society of America (JASA)*.

James Simmons is a highly regarded scientist who has published extensively, with over 95 journal articles appearing in prestigious journals including 8 in *Science*, 2 in *Nature*, 20 in *JASA* and 2 in *Acoustics Research Letters Online (ARLO)*. For many of us, to have a single paper published in *Science* or *Nature* would be a tremendous accomplishment. To have 10 manuscripts in these two journals is absolutely astounding. He is also a much sought-after speaker, having being invited to speak at over 85 universities and research institutions throughout the world. An example of the high regard colleagues have of Jim came from a comment by Dr. Prof. Hans Ulrich Schnitzler, who in his own right is an internationally renowned scientist, ''Jim Simmons has provoked me to think more than any other individual in the field.''

Jim and Andrea are proud parents of their two children, Jessica, a junior art major at Brown, and Ryan, a senior in high school. Both children spent much of their formative years acting as field assistants at bat sites all over the world.

We could not have selected a more deserving individual to receive the Silver Medal in Animal Bioacoustics than James A. Simmons.

WHITLOW W. L. AU
CYNTHIA F. MOSS

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal

in

Speech Communication

Katherine S. Harris

2005

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Franklin S. Copper	1975	Arthur S. House	1991
Gunnar Fant	1980	Peter Ladefoged	1994
Kenneth N. Stevens	1983	Patricia K. Kuhl	1997
Dennis H. Klatt	1987		

CITATION FOR KATHERINE SAFFORD HARRIS

⋯ for research and leadership in speech production

MINNEAPOLIS, MINNESOTA • 19 OCTOBER 2005

Katherine Harris was born and spent her early years in Lowell, Massachusetts. She moved to Greenwood, Mississippi in time to attend high school, where one of her classmates was James Flanagan. She received a B.A. in Psychology from Radcliffe College and a Ph.D. in Psychology from Harvard University. She then joined the research staff at Haskins Laboratories, her lifelong research home, to work on questions of speech perception and later on questions of speech production. In speech perception she is most closely associated with studies of the relationship between phone identification and discrimination, and with the cues for fricative consonants. However, it is her pioneering research in electromyographic (EMG) studies of speech that incorporates her most significant contributions to speech science. Today EMG research falls under the area of neuroscience research. With her first EMG publication in 1964 (Harris, Rosov, Cooper and Lysaught, 1964), she is clearly one of the founders of a research tradition that studies human behavior through neuroscience methods.

It is especially notable that Kathy's research focused on speech, one of the most complex of human behaviors. EMG in the 1960's was primarily used to measure steady-state contractions of muscles. Kathy's work led to innovations in hardware and software to track dynamic changes in the muscles of speech production. Other innovative procedures have been incorporated in her speech production research, for example the use of fiber-optic scopes to view laryngeal function.

Kathy's leadership in the area of speech production has been acknowledged from the beginning of her career. At the National Institutes of Health (NIH) she has participated on grant review panels, as well as serving as a consultant on taskforces and advisory committees for the National Institute for Deafness and Other Communication Disorders. For 20 years she was the principal investigator on an NIH Clinical Research Center Grant, directing a large number of investigators and diverse research projects. Notably, Kathy has been selected as a Fellow by four distinguished societies, the Acoustical Society of America, the American Association for the Advancement of Science, the American Speech-Language-Hearing Association, and the New York Academy of Sciences. Kathy has frequently been invited to present her research and reviews on all aspects of the field of speech production at major national and international meetings. The significant role she has played in speech production research was further demonstrated by her leading role in the history lectures on Speech Communication in preparation for celebrating the 75th anniversary of the Acoustical Society of America (ASA) (Harris, Ladefoged, and Stevens, 2000).

As a Distinguished Professor in the doctoral program in Speech and Hearing Sciences at the City University of New York Graduate Center, Kathy has directly influenced generations of speech scientists. She is an outstanding mentor and teacher who instills the love for research by example. Moreover, her vast knowledge of speech communication, coupled with an ability to provide critical insights into research, made her invaluable to her doctoral students at the City University of New York (CUNY), as well as to the many new research scientists who came to Haskins Labs. Of the numerous doctoral students that she mentored, we might mention the two most recent past chairs of the Speech Technical Committee, Emily Tobey (Nelle C. Johnston Chair in Communication Disorders at the University of Texas at Dallas) and Diane Kewley-Port (Associate Professor of Speech Science at Indiana University) who are recognized as eminent scientists in speech production. It should also be noted that Kathy has been a role model for women in acoustics as one of the women to become prominent in ASA. In addition to her role as mentor, Kathy's importance as a teacher of acoustics and speech science has been established through her textbook, *Speech Science Primer*, which was written with her former doctoral students Gloria Borden and Lawrence Raphael. It has been one of the most widely used textbooks in speech science during the past 20 years, and is now in its fourth edition.

Kathy's service to the Acoustical Society is well known, including her recent term as President of the Society. Perhaps what is not as well known is the vital role she has played in being a prominent liaison with the American Speech-Language-Hearing Association (ASHA). Kathy's primary teaching has been in a speech and hearing science department. This has given her a natural association with ASHA where she has served on important committees, such as chairing the Publications Board. Kathy received the Honors of the Association from ASHA, their highest award for service. Thus Kathy has been able to facilitate communication between ASA and

ASHA, as for example to set up formal arrangements between the Societies to cooperate in promoting and selecting young investigators for the Research Grant in Speech Science in memory of Dennis Klatt.

Katherine Harris has truly been a pioneer in speech communication. Her contributions to theories of speech perception, her leadership in developing neuroscience techniques for speech, and her role as a prominent woman in the Acoustical Society have all provided inspiration for generations of young scientists to join in her vision. In these efforts she has been fully supported by her husband, George Harris, and her daughters Maud and Louise. Her selection for the Society's Silver Medal in Speech Communication reflects the importance her lifelong contributions to both speech science and the Society.

DIANE KEWLEY-PORT
LAWRENCE J. RAPHAEL
FREDERICKA BELL-BERTI

ACOUSTICAL SOCIETY OF AMERICA

TRENT–CREDE MEDAL

Jerry H. Ginsberg

2005

The Trent-Crede Medal is presented to an individual, irrespective of nationality, age, or society affiliation, who has made an outstanding contribution to the science of mechanical vibration and shock, as evidenced by publication of research results in professional journals or by other accomplishments in the field.

PREVIOUS RECIPIENTS

Carl Irwin Vigness	1969	Eric E. Ungar	1983
Raymond D. Mindlin	1971	Miguel C. Junger	1987
Elias Klein	1973	Gideon Maidanik	1991
J. P. Den Hartog	1975	Preston W. Smith, Jr.	1996
Stephen H. Crandall	1978	David Feit	1999
John C. Snowdon	1980	Sabih I. Hayek	2003
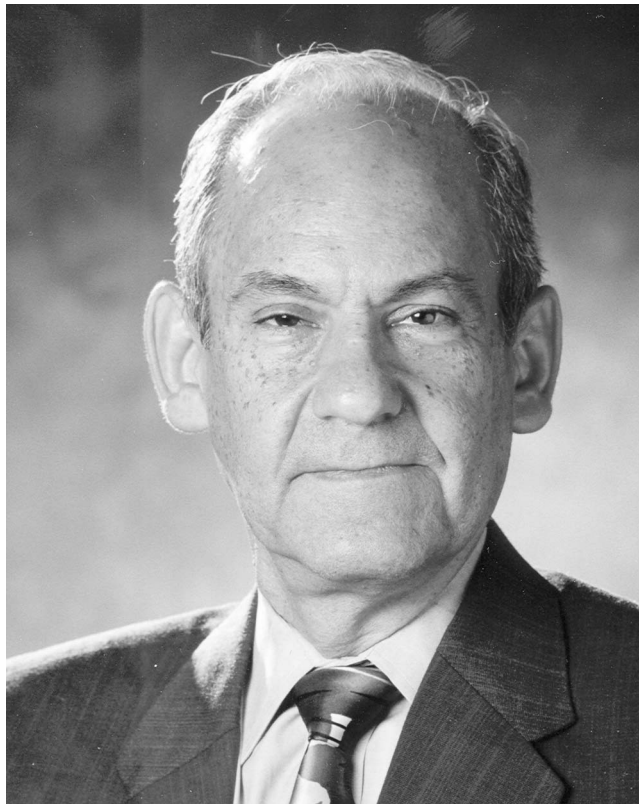
CITATION FOR JERRY H. GINSBERG

··· for contributions to the theory of vibrations of complex systems.

MINNEAPOLIS, MINNESOTA • 19 OCTOBER 2005

Jerry H. Ginsberg was born in New York City on September 18, 1944, the youngest of the three children of David and Raechel Ginsberg. In most respects, he was a normal child, who enjoyed sports and the fellowship of his peers, but he was undeniably very bright, and a youthful interest in science and technology led him to the Bronx High School of Science, from which he graduated in 1961. (Among his fellow classmates was William A. Kuperman, who also went on to a distinguished career in acoustics.)

Jerry received his undergraduate education at the Cooper Union, graduating in 1965 with the Cooper Union Alumni Award for Excellence in Civil Engineering. Admission was and still is extremely competitive, and the quality of education, distinguished by small classes, a challenging curriculum, and elective instructors, was the chief priority. (Another ASA medallist with a degree from Cooper Union is Martin Greenspan, who received the Gold Medal in 1983.)

During his undergraduate years, Jerry discovered that he had an exceptional aptitude for dynamics, and was influenced by the advice of one of his professors, Anthony Armenakis, to pursue graduate work in engineering mechanics. There was several good possibilities, but Columbia was chosen, partly because it was in New York, but primarily because of the exceptional program it had at that time in engineering mechanics, especially the dynamics of continuous structures. [One may note that the two previous recipients of the Trent-Crede Medal, David Feit (1999) and Sabih Hayek (2003), had preceded Jerry at Columbia and had received doctorates from the same program. One of the faculty members, Raymond D. Mindlin, was the second recipient (1971) of the Medal.]

Jerry received the Master of Science degree from Columbia in 1966 and the Doctorate in Engineering Science (E.Sc.D.) in 1970. While a graduate student he held a NASA Traineeship from 1965–1968 and was a graduate research assistant in the final year of his graduate studies. His doctoral work, advised by Hans Bleich and supported by the Office of Naval Research (ONR), was on the nonlinear vibrations of cylindrical shells, and was completed during the summer of 1969.

In 1968, Jerry married Rona Axelrod, and in the Fall of 1969, Jerry and Rona moved to Lafayette, Indiana, where Jerry began a long and distinguished role in engineering education with an appointment as a Assistant Professor in the School of Aeronautics, Astronautics, and Engineering Science at Purdue University. Engineering mechanics (which was at one time a separate department) at Purdue was then in somewhat of a state of flux, and Jerry eventually was transferred to the School of Mechanical Engineering in 1973, where he was promoted to Associate Professor in 1974. His mentor during the early years was Joseph Genin, and they collaborated for a period on work supported by the American Railroad Association to understand the complex dynamical interaction of trains with tracks. A number of papers were also published on vibrations and stability, and Jerry's interests began to be drawn toward fluid-structure interaction: he was substantially influenced by his reading of Sydney Goldstein's *Modern Developments in Fluid Mechanics*. A proposal in 1973 to the National Science Foundation on the nonlinear dynamics of shells surrounded by fluid received strongly positive reviews and commenced a series of grants that led to increasingly greater involvement with nonlinear acoustics.

Jerry was attracted to the Acoustical Society of America by its *Journal*, which was the leading forum for research in nonlinear acoustics, and he was impressed by the intellectual quality of the review that he got back on his first submission (published in 1978). He joined the Society in 1979 and presented a paper at the spring meeting in June 1979. It was the beginning of a long and very active involvement with the Society, leading down through the years to serving on many committees, chairing several (including the Technical Committee on Structural Acoustics and Vibration and the Books+ Committee), serving as Technical Chair for one of the Society's meetings (Atlanta, 2000), and serving as an Associate Editor for the *Journal*. He was made a Fellow of the Society in 1987.

Jerry's avocation as a textbook writer began in the Purdue years. The first was a two volume set, *Statics and Dynamics*, coauthored by Genin and published by Wiley in 1977. The book received a prize from the Council on Graphics Excellence of the Printing Industries of America and was well received, although the market for introductory mechanics texts for engineering students was quite crowded. A second edition, a substantial rewrite to more nearly accommodate

less capable students, appeared in 1984. Jerry turned his attention to more advanced texts in later years, publishing his *Advanced Engineering Dynamics*, first edition in 1988 and second in 1995, and then publishing *Mechanical and Structural Vibrations—Theory and Applications* in 2001. Both represent innovative departures from the previous texts on these subject, and are widely used today as textbooks.

In 1980, Stothe P. Kezios, the Director of Georgia Tech's School of Mechanical Engineering, met Jerry and subsequently recruited him for the Georgia Tech faculty. There Jerry continued his work on nonlinear acoustics and also teamed with Allan Pierce to work on ONR supported projects on structural acoustics. Peter Rogers, Yves Berthelot, Jacek Jarzynski, Ken Cunefare, and other outstanding individuals whose names are well-known in acoustics later joined the group. Jerry continued to excell as a classroom teacher, with near perfect evaluations from the students, and a succession of awards for excellence in teaching, including a Faculty Member of the Year Award in 1983, the Georgia Tech Distinguished Professor Award in 1994, and the Archie Higdon Distinguished Educator in Mechanics Award from the American Society for Engineering Education in 1998. The School received a substantial endowment and changed its name to the G. W.Woodruff School, and in 1989 Jerry became the first holder of the G. W. Woodruff Chair in Mechanical Systems.

There were a long succession of doctoral students; those currently active in the ASA include Sean F. Wu (coadvised with Allan Pierce and now with Wayne State University), J. Gregory McDaniel (now with Boston University), Pei-Tai Chen (now with the National Taiwan Ocean University), Gee-Pinn Too (now with Cheng Kung University in Taiwan), and Kuangcheng Wu (now with Northrop Grumman Newport News). Every meeting of the ASA invariably has papers presented by one or more of Jerry's students, and they frequently win Best Student Paper awards; the most recent one was to Matt Allen in 2004.

In addition to those honors mentioned above, Jerry is a Fellow of the American Society of Mechanical Engineers (ASME, 1989) and received a Fullbright-Hays Fellowship (for advanced study in France) in 1975–1976. He gave the ASME's Rayleigh Lecture in 2001, and was a Keynote Speaker at the ASME Design Engineering Technical Conference in 1997.

Jerry's research accomplishments are varied and extensive, and reported in over a hundred publications, as well as in the theses written by his students. Among the topics that he has addressed and for which he has substantially increased our understanding are the effects of damping on nonlinear systems, dynamics of air-cushioned vehicles, multi-dimensional nonlinear propagation of sound from vibrating surfaces, the doubly-asymptotic approximation for sound radiation from vibrating structures, variational formulations for sound-structure interaction and radiation, nonlinear reflection from elastic media, the nonlinear progressive equation (NPE) model, veering of eigenvalues in complex vibrating systems, radiation modes and complex power for submerged bodies, fuzzy structures, vibration control using parametric excitation, and the dynamics of rotating machinery. Perhaps his greatest legacy is in his work on vibrations of complex systems; and it is likely that his book on vibrations will long endure as a suitable high-level introduction to this field.

No account of Jerry's professional life would be complete without an acknowledgment of the superb help from his wife Rona Axelrod Ginsberg (who herself has had a extensive publishing career) in the editing of Jerry's writings. Jerry and Rona presently live in Dunwoody, a suburb of Atlanta. They have two sons, Mitchell and Daniel, of whom they are very proud.

Mitchell and his wife Tracie Sears Ginsberg have three daughters, Leah, Elizabeth, and Abigail, who are a continuous delight to their grandparents.

The Acoustical Society of America is proud to have the list of distinguished recipients of the Trent-Crede Medal further enhanced with the addition of Jerry H. Ginsberg to the list. We wish him continued success and honor in all his endeavors.

ALLAN D. PIERCE

Session 4aAA

Architectural Acoustics: Early Reflections on Reflections

Peter D'Antonio, Cochair

RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Trevor Cox, Cochair

RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Chair's Introduction—9:00

Contributed Papers

9:05

4aAA1. Creative reflections—the strategic use of reflections in multitrack music production. Alexander Case (Dept. of Music, Sound Recording Technol., Univ. of Massachusetts Lowell, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

There is a long tradition of deliberately capturing and even synthesizing early reflections to enhance the music intended for loudspeaker playback. The desire to improve or at least alter the quality, audibility, intelligibility, stereo width, and/or uniqueness of the audio signal guides the recording engineer's use of the recording space, influences their microphone selection and placement, and inspires countless signal-processing approaches. This paper reviews contemporary multitrack production techniques that specifically take advantage of reflected sound energy for musical benefit.

9:20

4aAA2. Simulating the acoustics of classical sound recording techniques in rooms. Jonas Braasch, William L. Martens, and Atsushi Marui (CIRMMT, Faculty of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, QC H3A 1E3, Canada)

Recently, a method to position sound sources in 3-D space using virtual microphone control has been proposed [J. Acoust. Soc. Am. **117**, 2391]. In this computer-generated environment, gains and delays between a virtual sound source and virtual microphones are calculated according to their distances and the axis orientations of the microphone directivity patterns. In the follow-up study reported here, it was investigated how to best simulate the influence of a room on the virtual microphone recording. For this purpose, a virtual rectangular room was created using the mirror-image technique (up to second-order reflections). The room dimensions were copied from an existing concert space at McGill University. Late reverb was created using a multiple feedback delay network with a time-variant architecture to enable modulation. To evaluate the system, measured impulse responses between a sound source and a five-channel microphone setup were compared to the virtual impulse responses for the same room/microphone-placement configuration. Among the tested parameters was the signal ratio between direct sound source and reflections as a function of the microphone placement and choice of microphone directivity patterns. The coherence between the simulated microphone signals was adjusted as well from the measured data. [Work supported by NSERC and VRQ.]

9:35

4aAA3. Sound-pressure level distribution in a long, narrow hallway: Measurements versus results from a computer model with scattering from surface roughness and diffraction. Lily M. Wang, Jonathan Rathsam (Architectural Eng. Program., Univ. of Nebraska-Lincoln, Peter Kiewit Inst., 1110 S. 67th St., Omaha, NE 68182-0681, lwang4@unl.edu), Claus L. Christensen, and Jens H. Rindel (Tech. Univ. of Denmark, DK-2800 Kgs. Lyngby, Denmark)

The sound-pressure level distribution down a long, narrow hallway due to a sound source at one end does not decrease linearly along the length of the hall. This characteristic may be due to the changing behavior of scattering that occurs down the length of the hallway, which is distance- and angle-dependent. A new scattering method that incorporates these effects has been implemented in the room acoustic computer modeling program, ODEON [C. L. Christensen and J. H. Rindel, Forum Acusticum, Budapest (2005)]. A comparison of the results from an ODEON model with real-world measurements along a long, narrow hallway on the campus of the Technical University of Denmark is provided in this paper. [Work supported by the NSF.]

9:50

4aAA4. Surface characterization from the point of view of an acoustical consultant. Scott D. Pfeiffer, Jacob Ament, Zackary Belanger, Molly Norris (Kirkegaard Assoc., 801 West Adams St., 8th Fl., Chicago, IL 60607), and Tim Gulsrud (Kirkegaard Assoc., Boulder, CO 80302)

The practical characterization of the acoustical properties of absorption, diffusion, and transparency of surfaces is an area where the application of physical acoustic techniques may be more applicable than typical architectural acoustical parameters. Understanding the diffusive behavior of surfaces is crucial to the design of spaces, but cumbersome to quantify, requiring different techniques depending on the question being asked. Past studies conducted by Kirkegaard Associates on products for partial reflection, diffusion, nominal transparency, and the results of a recent implementation of the current proposed Audio Engineering Society standard AES-4id-2001 are discussed with respect to the need to make design decisions and implement products for fulfillment of an acoustic design process. The viability of using a single-valued diffusion coefficient is discussed with respect to the needs of the acoustical consulting community, while still exploring the possibility of alternative characterization methods.

10:05

4aAA5. A narrow-band analysis of reflected magnitude and phase from six reflector panel arrays. Jonathan Rathsam, Lily M. Wang (Architectural Eng. Program, Univ. of Nebraska-Lincoln, 247 PKI, 1110 S. 67th St., Omaha, NE 68182-0681, jrathsam@mail.unomaha.edu), and Rendell R. Torres (Rensselaer Polytechnic Inst., Troy, NY 12180-3590)

This investigation analyzes measurements of the reflected sound field from various reflector panel arrays. A previous study by the third author [R. R. Torres and M. Vorlaender, "Scale-model MLS-measurements of scattering from overhead panel arrays," Acta Acustica (in press)] measured impulse responses using the maximum length sequence method at various receiver positions from six scale-model panel arrays of different sizes and densities. This study included an octave band analysis of reflected magnitude as a function of receiver position. The current authors have extended the research by conducting a narrow-band analysis of magnitude and phase of the reflected sound fields for three of the panel arrays [J. Rathsam, L. M. Wang, and R. R. Torres, J. Acoust. Soc. Am. **117**, 2499(A) (2005)]. This study revealed a close relationship between phase and magnitude. The current work continues the analysis of reflected phase and magnitude for the three additional reflector panel arrays. The newer arrays are more complex and may reveal the effects of multiple diffraction more clearly than the previous three arrays.

10:20–10:35   Break

10:35

4aAA6. A stable transient boundary element method (BEM) for diffuser scattering. Jonathan A. Hargreaves and Trevor J. Cox (Univ. of Salford, M5 4WT, UK)

Boundary element methods (BEM) may be used to model scattering from hard rigid surfaces such as diffusers. They have the advantage over volumetric methods that only the surface need be meshed and the surface velocity potential found. Unlike the more widely used single frequency methodology, transient BEM discretizes integral equations to produce an iterative system that is marched on in time from known initial conditions to calculate how the velocity potential varies over time. This iterative process can be unstable, and this is one reason why transient BEM is not more widely used. Previous works on transient BEMs have focused on idealized surfaces, such as spheres and plates. However, little is published on the performance of these methods for more complex surfaces of interest, such as Schröder diffusers. Consequently, this paper presents an implicit scheme suitable for a surface comprising thin and solid sections. Such an implicit scheme has the benefits of not constraining time-step duration to the smallest surface detail, and brings stability benefits. Numerical integration is carried out efficiently and accurately by conversion to contour integrals. Accuracy and stability is investigated by comparison to a verified single frequency BEM.

10:50

4aAA7. Wavelet transform use for reflection analysis in architectural acoustics. Zackery Belanger (Kirkegaard Assoc., 801 West Adams St. 8th Fl., Chicago, IL 60607)

The use of wavelet transforms in the realm of architectural acoustics is discussed as an alternative to more traditional frequency-analysis techniques. Wavelet transforms allow the time and frequency information of a signal to be viewed simultaneously, providing a valuable way to study the frequency content of individual reflections and the frequency structure of impulse responses. Discussion will focus on the application of wavelet transforms to project work at Kirkegaard Associates.

11:05

4aAA8. Acoustics of a planetarium. Micah Shepherd, Timothy W. Leishman, and Sentagi Utami (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., C110 ESC BYU, Provo, UT 84602)

Brigham Young University has recently constructed a planetarium with a 38-ft.-diameter dome. The facility also serves as a classroom. Since planetariums typically have poor acoustics due to their domed ceiling structures, acoustical recommendations were requested before its construction. The recommendations were made in an attempt to create an acceptable listening environment for lectures and other listening events. They were based in part on computer models and auralizations intended to predict the effectiveness of several acoustical treatments on the outer walls and on the dome itself. The recommendations were accepted and the planetarium was completed accordingly. A series of acoustical measurements was subsequently made in the room and the resulting acoustical parameters were mapped over the floor plan. This paper discusses these results and compares them with the predictions of the computer models.

11:20

4aAA9. A novel approach to achieving significant reverberation control in performance halls. David A. Conant and William Chu (McKay Conant Brook, Inc., 5655 Lindero Canyon Rd., Ste. 325, Westlake Village, CA 91362, DConant@MCBinc.com)

Conventional methods for achieving broadband, variable sound absorption in large halls normally include heavy application of sound-absorptive drapery and/or thick fibrous panels, applied near available surfaces below, at, and in volumes above the catwalk plane. Occasionally, direct adjustments to room air volume are also provided to effect double-sloped decays. The novel method described here combines carefully located, broad scattering and absorption in singular architectural elements and was applied to a new, 1200-seat concert hall. A change of 0.70 s RT60 in midfrequency is achieved in a visually dramatic manner while neither materially changing room volume nor introducing often-maligned drapery. The aggregate of reverberation control methodologies employed reduces the unoccupied RT60 at midfrequencies from about 3.2 to 1.7 s in this space programed principally for music, including pipe organ. Results of MLS measurements including binaural measurements and binaural recordings of anechoic material and CATT-acoustic modeling and auralizations are discussed.

11:35

4aAA10. Development of a new fabric canopy material. Zackery Belanger (Kirkegaard Assoc., 801 West Adams St. 8th Fl., Chicago, IL 60607)

The unique requirements of two concert halls, one existing and one new, have driven the development of a new fabric canopy material with beneficial reflection and transmission spectra. The development process is highlighted, including motivation, material choice, testing, and challenges such as absorption issues and architectural requirements. The potential of the fabric for use in other acoustic environments is also discussed.

11:50

4aAA11. Evaluation of diffuse reflections in a 1:10 scale model hall. Jin Yong Jeon, Jong Kwan Ryu, and Yong Hee Kim (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea, jyjeon@hanyang.ac.kr)

The acoustical properties of a multipurpose hall (opera mode) were measured in a 1:10 scale model to investigate the effect of diffuse reflections. Measurements were made in stall areas with and without diffusers on the walls adjacent to the pit and on the side walls. The diffusers consisted of 15- to 25-cm-high hemispheres and 25-cm high polygons. A mix of these diffusers was used to cover about 40% of the area of the side walls. It was found that 1-IACCE3 increased at positions close to the walls

4a THU. AM

1999   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **1999**

adjacent to the pit and the first reflection with maximum amplitude increased in some seats. The possibility of diffuse reflections either hitting the proscenium wall or soffit close to the side wall first, as well as reflections coming from both sides of the hall explains the reason for the bunch of early reflections in the impulse responses. The effect of diffusers at each receiver position was investigated via subjective evaluation. The results from the subjective evaluation indicated that diffuse reflections were preferred to specular reflections.

THURSDAY MORNING, 20 OCTOBER 2005

HENNEPIN ROOM, 9:00 TO 11:00 A.M.

Session 4aAB

Animal Bioacoustics: Topics in Bioacoustics

Michael J. Ferragamo, Chair

Gustavus Adolphus College, Dept. of Biology, St. Peter, MN 56082

Contributed Papers

9:00

4aAB1. Whistle repertoire of bottlenose dolphins (*Tursiops truncatus*) in the Mississippi Sound. Erica N. Hernandez, Stan Kuczaj (Univ. of Southern Mississippi, 118 College Dr., Box 5025, Hattiesburg, MS 39402), and Moby Solangi (Inst. for Marine Mammal Studies, Gulfport, MS 39502)

The whistle repertoire of wild bottlenose dolphins (*Tursiops truncatus*) in the Mississippi Sound, part of the northern Gulf of Mexico, was investigated. There is a large population of dolphins in this area, and many dolphins that are now housed in zoos and aquariums were captured in the Mississippi Sound. This paper reports the types of whistles that are predominant in this area, and how these whistles are used in the context of concurrent surface behavior. Over the course of 1 year (April 2004–March 2005), dolphin whistles were recorded as part of an ongoing study of the effects of human activity on wild bottlenose dolphins. The surface behavior of the focal group was categorized at 1-min intervals as follows: mill, travel, mill/travel, feed, social, with boat, or with shrimp boat. Whistles were then categorized as one of the following: upsweep, downsweep, convex, concave, sine, or constant frequency. Preliminary analysis of the data suggests that both the rate of whistling and the types of whistles produced vary as a function of dolphin behavior. Further analysis of the data will reveal if different types of whistles are associated with specific surface behavior categories. [Research supported by Department of Commerce.]

9:15

4aAB2. Sound-conducting mechanisms for echolocation hearing of a dolphin. Vyacheslav A. Ryabov (Karadag Natural Reserve Natl. Acad. of Sci. of Ukraine, Kurortnoe, Feodosiya 98188, Crimea, Ukraine, ryaboff@ukr.net)

The morphological study of the lower jaw of a dolphin (*Tursiops truncatus p.*), and the modeling and calculation of its structures from the acoustic point of view have been conducted. It was determined that the cross-sectional area of the mandibular canal (MC) increases exponentially. The MC represents the acoustical horn. The mental foramens (MFs) is positioned in the horn throat, representing the nonequidistant array of waveguide delay lines (NAWDL). The acoustical horn ensures the traveling wave conditions inside the MC and intensifies sonar echoes up to 1514 times. This ''ideal'' traveling wave antenna is created by nature, representing the combination of the NAWDL and the acoustical horn. The dimensions and sequence of morphological structures of the lower jaw are optimal both for reception and forming the beam pattern, and for the amplification and transmission of sonar echoes up to the bulla tympani. Morphological structures of the lower jaw are considered as components of the peripheral section of the dolphin echolocation hearing.

9:30

4aAB3. Automatic recognition of harmonic bird sounds. Jason R. Heller (Dept. of Appl. Math, Stony Brook Univ., Stony Brook, NY 11794, hellerjr@ams.sunysb.edu) and John D. Pinezich (Adv. Acoust. Concepts, Hauppauge, NY 11788)

The method of sound recognition relies on a transformation of a sound into a spectrogram followed by extraction of the harmonics as curves. The extracted curves are called *frequency tracks*. A procedure called *find_feasible_sets* is used to extract sets of tracks that may correspond to harmonic sounds. If a set of tracks overlap each other sufficiently in time, then the set is designated a *feasible set*. Following the extraction of the feasible sets, the procedure *find_maximal_subsets* is applied to each feasible set. This procedure uses a function called *harmonic_relate* that determines if two tracks are harmonically related. All tracks that are not harmonically related to any other tracks in the feasible set are discarded. Furthermore, the feasible set is divided into *maximal subsets*. A maximal subset is a subset of the feasible set in which every track is harmonically related to one fixed track in the set called the *reference track* but no other tracks in the feasible set are related to the reference track. Each frequency track in a track set is transformed into a *feature vector* whose components describe the frequency, slope, and shape of the track. The species of birds analyzed are bluejay and herring gull.

9:45

4aAB4. An optimized toolchain for predicting directivity patterns from digital representations of biological shapes. Rolf Müller (School of Phys. and Microelectronics, Shandong Univ., Hongjia Lou 5, Jinan, China, mueller@sdu.edu.cn)

Animals have evolved intricate shapes which diffract emitted or received sound and thereby generate a specific directivity pattern. Computer-tomographic methods can generate high-resolution digital representations of these morphological structures in the form of three-dimensional voxel arrays. However, predicting acoustic directivity patterns from these representations with numerical methods can incur high computational cost, e.g., for large structures with fine detail and/or high wave numbers (as in bats and dolphins). Here, the design of a toolchain is described which can handle all steps of deriving a directivity prediction from a voxel representation: generation of a finite-element mesh, assembly of the system matrix, computation of an approximate solution, forward projection into the far field. All individual operations are performed by self-contained tools, which communicate through files. This gives access to intermediate results and limits re-execution upon parameter changes to downstream steps. At each stage, optimizations can be made based on the specifics of the problem such as the regular structure of the voxel array and the distance independence of the directivity. Use of these optimizations has resulted in a highly efficient performance, which is documented by measures for execution speed, memory usage, and accuracy.

2000 J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005 **2000**

10:15

4aAB5. Equine acoustics: Anatomy of a whinny. David G. Browning (Dept. of Phys., Univ. of Rhode Island, 2 Lippitt Rd., Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Dept. of Animal Sci., Univ. of Connecticut, Storrs, CT 06269)

Of the roughly nine different vocalizations of a horse, the whinny appears to be the most interesting. A whinny is a horse's primary means of long range vocal communication; the bandwidth and variability offer the possibility of expression, at least at a primitive level. Acoustic analysis of a whinny shows three distinct domains: the initial frequency ramp-up running from 1 to 2 kilohertz in roughly 1 second, matched by a similar response in the second and third harmonics; secondly, this is followed by a nasal tremolo of a longer duration with generally a slight downslope in frequency (this is perhaps the stage most associated with the human perception of a whinny); and, lastly, a guttural tremolo, essentially the same sound as a nicker. In the samples analyzed, each domain seems to vary independently both in strength and duration. Attempts to link an aspect of a whinny with a particular behavior is still in a formative stage, complicated by the fact that a horse's behavior is usually based primarily on visual rather than acoustic inputs.

10:30

4aAB6. Identifying individual sperm whales acoustically using self-organizing maps. Juliette W. Ioup and George E. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148)

The Littoral Acoustic Demonstration Center (LADC) is a consortium at Stennis Space Center comprising the University of New Orleans, the University of Southern Mississippi, the Naval Research Laboratory, and the University of Louisiana at Lafayette. LADC deployed three Environmental Acoustic Recording System (EARS) buoys in the northern Gulf of Mexico during the summer of 2001 to study ambient noise and marine mammals. Each LADC EARS was an autonomous, self-recording buoy capable of 36 days of continuous recording of a single channel at an 11.7-kHz sampling rate (bandwidth to 5859 Hz). The hydrophone selected for this analysis was approximately 50 m from the bottom in a water depth of 800 m on the continental slope off the Mississippi River delta. This paper contains recent analysis results for sperm whale codas recorded during a 3-min period. Results are presented for the identification of individual sperm whales from their codas, using the acoustic properties of the clicks within each coda. The recorded time series, the Fourier transform magnitude, and the wavelet transform coefficients are each used separately with a self-organizing map procedure for 43 codas. All show the codas as coming from four or five individual whales. [Research supported by ONR.]

10:45

4aAB7. Spectral identification of sperm whales from Littoral Acoustic Demonstration Center passive acoustic recordings. Natalia A. Sidorovskaia, Blake Richard (Phys. Dept., UL Lafayette, Lafayette, LA 70504, nsidorovskaia@louisiana.edu), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The Littoral Acoustic Demonstration Center (LADC) made a series of passive broadband acoustic recordings in the Gulf of Mexico and Ligurian Sea to study noise and marine mammal phonations. The collected data contain a large amount of various types of sperm whale phonations, such as isolated clicks and communication codas. It was previously reported that the spectrograms of the extracted clicks and codas contain well-defined null patterns that seem to be unique for individuals. The null pattern is formed due to individual features of the sound production organs of an animal. These observations motivated the present studies of adapting human speech identification techniques for deep-diving marine mammal phonations. A three-state trained hidden Markov model (HMM) was used with the phonation spectra of sperm whales. The HHM-algorithm gave 75% accuracy in identifying individuals when it had been initially tested for the acoustic data set correlated with visual observations of sperm whales. A comparison of the identification accuracy based on null-pattern similarity analysis and the HMM-algorithm is presented. The results can establish the foundation for developing an acoustic identification database for sperm whales and possibly other deep-diving marine mammals that would be difficult to observe visually. [Research supported by ONR.]

4a THU. AM

Session 4aAO

Acoustical Oceanography and Underwater Acoustics: Environmental Variability and Sonar Systems

James F. Lynch, Cochair
Woods Hole Oceanographic Inst., 203 Bigelow Bldg., Woods Hole, MA 02543

William M. Carey, Cochair
Boston Univ., College of Engineering, Boston, MA 02215

Chair's Introduction—8:25

Contributed Papers

8:30

4aAO1. Time and frequency analysis of seismic airgun calibration data from an environmental acoustic recording system (EARS) buoy.
George E. Ioup, Juliette W. Ioup, Sean R. Chapin, Arslan M. Tashmukhambetov (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), Joal J. Newcomb, William M. Sanders (Naval Res. Lab., Stennis Space Center, MS), Christopher D. Walker, Benjamin Brack, Grayson H. Rayborn, James M. Stephens (Univ. of Southern Mississippi, Hattiesburg, MS 39406), and Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA 70504)

In the summer of 2003 two Environmental Acoustic Recording System (EARS) buoys were deployed in the northern Gulf of Mexico by the Littoral Acoustic Demonstration Center. The buoys were collocated and recorded ambient noise and seismic airgun array shots up to approximately 25 kHz. The gains and hydrophone sensitivities were set such that one EARS buoy could record the seismic shots without clipping and the other could record ambient noise. The M/V Kondor towed an airgun array on parallel linear tracks with horizontal closest points of approach to the buoy of 0, 500, 1000, 2000, and 5000 m, giving experimental data for a wide range of horizontal distances (up to 7 km) and arrival angles. The raw data were calibrated using the EARS system parameters to produce calibrated pressure time series for each shot. These data are analyzed in both the time and frequency domains. Maximum pressures for each shot as well as sound exposure levels (pressure squared integrated over time for 200 ms in this case) are presented. Also presented is a spectrogram analysis. The maximum time-domain peak pressure recorded is 200 dB *re*: 1 μPa. The maximum sound exposure level is 177 dB *re*: 1 μPa^2 s. [Research supported by Industry Research Funding Coalition.]

8:45

4aAO2. Littoral Acoustic Demonstration Center 2003 seismic calibration experiment: Modeling acoustic energy distribution from seismic exploration array. Natalia A. Sidorovskaia (Phys. Dept., UL Lafayette, Lafayette, LA, 70504, nsidorovskaia@louisiana.edu), Arslan Tashmukhambetov, George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

In the summer of 2003 the Littoral Acoustic Demonstration Center (LADC) conducted a calibration experiment to measure the acoustic energy output of an industrial seismic exploration array in a frequency range up to 25 kHz. A standard acoustic propagation model, RAM (Range-dependent Acoustic Model by M. Collins), was adapted to model the broadband (up to 1000 Hz) waveguide transfer function for a 21-element moving seismic exploration source array generating acoustic data. Notional source signatures for each airgun in an array were generated by the calibrated airgun modeling package GUNDALF, which is based on original theoretical work by Ziolkowski, Hatton, Laws, and others. The package allows modeling of the very close bubble interactions from each airgun in an array. Experimental and simulated data demonstrate good agreement in

the frequency range up to 300 Hz. Factors responsible for the discrepancies between measured and modeled data in the higher frequency region are discussed. Environmental implications of variations in the acoustic energy distribution with seasonal and geographic changes in the propagation environment are addressed. [Research sponsored by the Industry Research Funding Coalition through the International Association of Geophysical Contractors.]

9:00

4aAO3. Arrival time analysis for seismic airgun data classification. James Stephens (Dept. of Phys. & Astron., Univ. of Southern Mississippi, 6971 Lincoln Rd., Hattiesburg, MS 39402)

Survey data were collected in order to characterize the output of a marine seismic airgun array for various propagation geometries. Near uniform sound speed in the survey area allowed for classification of first-break and secondary arrivals as direct arrival and either reflected or refracted path arrival, respectively, using simple propagation geometries. Plots of initial arrival versus horizontal offset between the airgun and receiver allow classification of the arrival as via direct, reflected, or refracted path in a manner similar to that employed in the analysis of seismic reflection data in the shot domain. Beyond a certain horizontal range (approximately 4.5 km), the critically refracted arrival overtakes the reflected arrival, affording an estimate of the propagation speed in the seafloor layer. Analysis of this sort allows separation of the direct arrival for analysis purposes.

9:15

4aAO4. Performance variability of matched-field processors on a shelf-break region in the Mid-Atlantic Bight. Young-Nam Na (Agency for Defense Development, Chinhae, Kyungnam, ROK 645-600), Peter C. Mignerey, and Bruce H. Pasewark (Naval Res. Lab., Washington, DC 20375-5350)

An acoustics experiment (RAGS) was conducted by NRL in December 2003 in the Mid-Atlantic Bight. The deployed equipment included two sources broadcasting 300- and 500-Hz cw signals and three vertical arrays distributed at 10-km intervals over a 30-km range. The water depth varied between 63-125 m. Temperature profiles were obtained from sensors attached to the vertical arrays. Fourier analysis of temperature and vertical array data show that the region is dominated by the M2 tide, the effects being strongest up slope. Several storms significantly influenced the received acoustic fields. Simulations, using several matched-field processors, show significant mismatches between replica and data fields are introduced by changes in the water depth and the sound speed profiles. Just 1 m of tidal depth changes cause significant phase differences between replica and data fields that induce nearly 20-dB variation in matched-field processor output. The study will apply several existing matched-field processors to the measured data with modeled replicas and discuss source

2002   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **2002**

localization results. The discussion will focus on the inevitable degradation of processors that utilize coherence between replica and measured data in shallow waters. [Work supported by the U.S. ONR and the R.O.K. Agency for Defense Development.]

9:30

4aAO5. Littoral tidally induced environmental influences on horizontal array performance. Roger Oba, Steven Finette (Acoust. Div. Naval Res. Lab., Washington, DC 20375), Colin Shen, and Thomas Evans (Naval Res. Lab., Washington, DC 20375)

A modeled continental shelf-break region is used to assess the combined and relative impact of bathymetry and internal wave/bore influences on horizontal array performance in a littoral environment. Tidal forcing over a smooth break varying from a maximum depth of 100-m flat bathymetry to 50 m is used in conjunction with a submesoscale hydrodynamic model to produce internal wave activity in a fully stratified water column. A continuous wave source (100 Hz, for example) is sited at different aspects relative to the shelf break, and the acoustic field is propagated to broadside arrays that are placed at various bearings. A three-dimensional parabolic equation code propagated the field to the arrays. Time-variable three-dimensional acoustic effects such as beam wander due to horizontal refraction, resulting from either slope or internal waves, are both illustrated. The time-dependent bathymetric effects can be either enhanced or reduced by the internal tidal activity, depending upon the relative geometry of propagation, bathymetry, and internal tide. The influence of bathymetry versus the full evolving environment due to hydrodynamic action can be compared using array correlation calculations, signal gain degradation, and beam wander. [This research is sponsored by the ONR.]

9:45

4aAO6. Environmentally associated signal-to-interference variability in low-frequency active sonar in Korea Strait. P. G. Cable (BBN Technologies, 11 Main St., Mystic, CT 06355) and W. M. Carey (Boston Univ., Boston, MA 02215)

An analysis has been conducted of spatial and temporal variability of target echo-to-interference measurements made during Area Characterization Test III (ACT III) in 1995 in the Korea Strait. The measurements were made over a 5-day period for five fixed bistatic geometries (spatial scale of order 200 square km) using explosive sources, bottom-mounted horizontal receiving hydrophone arrays, and passive reflector targets in 100-m-depth water under downward-refracting acoustic conditions. The bottom at the ACT III site was nearly flat, of sand–silt composition. Meteorological conditions were relatively calm over the conduction of the measurements, but the main proximate oceanographic feature, the South Korean Coastal Front, did result in a relatively varied and complex sound-speed environment. The signal-to-interference fluctuation statistics associated with a single ping was determined to give a standard deviation of order 1 dB, while the observed variability of signal-to-interference over the spatial and temporal scales of the measurements gave a standard deviation of about 2 dB. The observed variability of received signal-to-interference will be discussed in terms of physical causes and measurement error. [Work supported by ONR Code OA321.]

10:00–10:20   Break

10:20

4aAO7. A review of spatial coherence and array signal gain results. William M. Carey (College of Eng., Boston Univ., Boston, MA 02215)

A question discussed over the last three decades is: "how does scattering in the deep and shallow-water-sound channels limit coherent array processing?" In particular, for the horizontal array, what is the limitation of resolution in terms of the mutual coherence function and its coherence length? For frequencies (400 Hz) the measurement of the magnitude-squared coherence in a multipath environment with a partly coherent-dynamic noise field is difficult. C. Carter has determined that the measurement of phase coherence in the presence of noise is a poor statistical

estimator since it is a function of the signal-to-noise ratio with large confidence bounds [IEEE Trans. Audio Electroacoust. **AU21**, 388–389 (1973)]. However, array signal gain and a theoretical coherence functional form can also provide a measure of the horizontal coherence length. This paper reviews single path coherence results and those derived from array measurements and shows that representative deep-water lengths for a frequency of 400 Hz are 100 wavelengths at 400 km with a frequency scaling of the 5/2 power while for shallow water they are 30 wavelengths at 40 km. These numbers will be discussed in terms of internal wave scattering theories.

10:35

4aAO8. Calculation of fine-scale sound-field structure based on towed CTD (conductivity, temperature, and depth) chain measurements. Joy E. Lyons, David L. Bradley, and R. Lee Culver (Grad. Program in Acoust., Penn. State Univ., State College, PA 16804, jel246@psu.edu)

In July 2002, an experiment was conducted off the coast of southern California in which the two-dimensional temperature and salinity field was measured directly using a towed CTD chain. The sensor spacing was smaller (2 to 8 m) near the surface (above 30 m) in order to better resolve the mixed layer and thermocline; the depth of the lowest sensor was about 150 m. The path of each tow was back and forth across the same section of ocean such that ocean dynamic features observed in the CTD data were recorded in expanded and compressed states depending on the ship's heading. The range of each run was a few kilometers, with data taken in 1-s increments. This yielded a high-resolution field suitable for analyzing the propagation behavior of high frequencies. Using the fine-scale measurements of the sound-speed field and an acoustic propagation model based on an FFT-realized parabolic equation, the sound field was calculated reflecting the observed ocean dynamic features. Modeling of scintillation in acoustic intensity will be discussed in the context of spice theory. [Work supported by ONR Code 321US.]

10:50

4aAO9. A hydro-acoustic source model in calculating noise field generated by breaking waves. Xuemei Chen, Steven L. Means, William G. Szymczak (Acoust. Div. Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, xchen@wave.nrl.navy.mil), and Joel C. W. Rogers (Polytechnic Univ., Brooklyn, NY 11201)

To develop a complete model for the breaking wave noise, it is necessary to relate the source quantities to the physical parameters of the wave-breaking and noise-generation processes. In this paper, the source structure of an individual breaking wave is simulated using a coupled hydro-acoustic model, which incorporates the physical processes underlying the mechanisms of the generation of the noise. The physical processes of wave formation and breaking are modeled using a generalized hydrodynamics formulation, providing the hydrodynamic parameters, such as pressure variations and air cavity shapes, etc., for the acoustic calculation. In the acoustic simulation, an algorithm has been developed in handling wave propagation in irregular regions, such as the bubbly liquid generated by wave-breaking. For the noise field modeling, the locations and occurrence times for the individual breaking waves are specified as stochastic quantities using Poisson simulations and the total noise field is calculated as the superposition of the noise contributed by those breaking waves. [Work supported by ONR.]

11:05

4aAO10. Sound field temporal fluctuations due to mode coupling in shallow water with internal soliton. Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), Valery Grigorev, Boris Katsnelson (Voronezh Univ., Voronezh, 394006, Russia), and James Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Sound field temporal fluctuations due to mode coupling, resulting from traveling internal solitons (IS) approximately along the acoustic track is addressed. The spectrum of received signal amplitude fluctuations in the range of very low frequency ($\sim F = 0.001$–0.01 Hz) is obtained. Maxi-

4a THU. AM

mums in this spectrum, correspond to the times which IS passes the sound field ray cycles. A relationship between F-spectrum and acoustic frequency is illustrated by showing Frequency-frequency $(F-f)$ diagram. It is shown that depending on the sound field frequency (f), different mode pairs give the most pronounced contribution to the sound intensity fluctuations. At the same time invariant frequency, F, of maximal fluctuations remains approximately the same for all sound frequency (f) bands. For example in the case of the SWARM95 experiment, F is about 0.0015 Hz for sound frequency, f in the range of 30–150 Hz. Generally, F depends on waveguide parameters and velocity of IS. In this paper, the physical meaning of the invariant frequency and its relationship to the sound frequency is discussed. In addition, the experimental setup for observation of F in future shallow-water acoustic experiments is proposed. [Work supported by CRDF, PFBR and ONR-321OA.]

11:20

4aAO11. Environmental influence on acoustic impulse responses across the New Jersey Shelf in the winter. Peter C. Mignerey (Acoust. Div. 7120, Naval Res. Lab., Washington, DC 20375-5350)

In December 2003 as part of the RAGS03 experiment, the Naval Research Laboratory moored three vertical arrays at ranges of 10, 20, and 30 km distant from fixed 300- and 500-Hz LFM acoustic sources. Data were recorded continuously for approximately 20 days, which covered periods of severe storms and various phases of internal tides. Associated temperature profile data show the internal tide arriving as either an internal jump or series of elevation waves superimposed on a bottom layer associated with the foot of a shelf-break front. Impulse responses derived from acoustic data received by the vertical arrays provide insight into the acoustic modal structure within the ocean for various sea states and phases of the internal tide. Results to date show significant loss of high modes during those time periods when either the internal tide is present or the sea state is high in comparison with periods of low sea state and absent internal tide. The extent of mode stripping increases with down-slope range. Apparently high acoustic modes are being scattered and stripped by near-boundary interactions. Evidence for these interactions will be presented along with a discussion of stripping processes. [Work supported by ONR.]

11:35

4aAO12. Acoustic sampling volume of a bistatic sonar. Kenneth G. Foote (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

The expression for the acoustic sampling volume of a monostatic sonar [K. G. Foote, J. Acoust. Soc. Am. **90**, 959 (1991)] is generalized to the case of a bistatic sonar, with separate, non-collocated transmitting and receiving transducers. The bistatic sampling volume is the integral of a counting function over the physically accessible space. The argument of the counting function, a Heaviside step function, is the difference between the scattering amplitude, or other measure of scattering strength, and a noise-dependent threshold. When this difference is positive, the integrand is unity; when zero, one-half; when negative, zero. Both the sampling volume and a related quantity, the equivalent beam angle, are evaluated numerically for several bistatic geometries, transducer shapes, and narrow-band frequencies for the cases of point scatterers and directional scatterers. The application of the bistatic acoustic sampling volume to the quantification of suspended particulate matter and aquatic organisms is discussed.

THURSDAY MORNING, 20 OCTOBER 2005   CARVER ROOM, 8:15 A.M. TO 12:00 NOON

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Acoustic Radiation Force Methods for Medical Imaging and Tissue Evaluation

Mostafa Fatemi, Chair

Mayo Clinic, Physiology Biophysics, 200 First St., SW, Rochester, MN 55905

Chair's Introduction—8:15

Invited Papers

8:20

4aBB1. Principles of radiation force. Oleg Rudenko (Blekinge Inst. of Technol., Karlskrona, Sweden and Moscow State Univ., Russia, rudenko@acs366.phys.msu.ru)

A brief review of general problems of radiation force (RF) is given. RF was discovered by Lord Rayleigh in 1902 and further fundamental contributions to the RF phenomenon were made by Brillouin and Langevin. In spite of century-old history, RF has not been studied exhaustively and there are many grey areas. The difference between the Rayleigh and Langevin RF appearing in fluids, the difference between averaging procedure in Lagrange and Euler representations, the connections between mean characteristics of the RF-induced streaming (velocity, pressure, mass transfer) and acoustic parameters are still under discussion. Everyone knows that RF is produced by a change in the density of ultrasonic wave energy due to the absorption or due to the reflection from obstacles and inhomogeneous inclusions. However, can RF appear in lossless homogeneous medium or not? The pumping effect of sound beam caused by hydrodynamic nonlinearity was first discussed more than 50 years ago, and similar effects in biomedical applications RF produced by high-intensity focused ultrasound are being discussed now. The talk reviews recent progress in investigation of the problems concerning RF in fluids, biological tissues, and solids with account for specific properties of wave field like nonlinearity, diffraction, absorption, focusing, and scattering.

4aBB2. Generation of ultrasound radiation force with the use of time reversal acoustics principles. Armen Sarvazyan and Alexander Sutin (Artann Labs., Lambertville, NJ 08530, armen@artannlabs.com)

There are numerous medical applications of ultrasound radiation force (RF) which could be made more effective using the time reversal acoustics (TRA) principles. This paper gives an overview of research into physical and technical bases of RF generation in heterogeneous biological media using TRA focusing systems. A custom-designed compact multichannel TRA system for receiving, digitizing, storing, time reversing, and transmitting acoustic signals in a wide frequency range from 0.01 to 10 MHz has been developed and extensively tested in model systems and *ex vivo* tissues and bones. Shear strain and shear waves remotely induced in soft tissues and bones by radiation force were detected using various acoustical and optical means. Experimental studies fully confirmed the feasibility of TRA generation of RF and demonstrated several advantages over conventional means of remotely inducing shear stress in biological media. These advantages include a possibility to create highly localized (close to diffraction limit) shear stress in heterogeneous media stir focused ultrasound beam in 3-D volume using very simple hardware. [Work supported by NIH grant.]

4aBB3. *In vivo* noninvasive method for measuring local wave velocity in femoral arteries of pig. Xiaoming Zhang, Randall Kinnick, Cristina Pislaru, Mostafa Fatemi, and James Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905)

We have proposed generating a bending wave in the arterial wall using ultrasound radiation force and measuring the wave velocity along the arterial wall [Zhang *et al.*, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 642–652 (2005)]. Here, we report the results of *in vivo* studies on pigs. The pig was anesthetized, and a micromanometer tip catheter was inserted into the femoral artery to measure luminal pressure. A water bath was created on the animal's groin to allow unimpeded access of the ultrasound beams to the femoral artery. The femoral artery was first located using a 13-MHz linear-array transducer. Then, a vibro-acoustography image was obtained to ensure precise positioning of the excitation force relative to the artery. The artery was excited by the force transducer and the resulting vibration of the arterial wall was measured by a sensing Doppler transceiver. Measured wave velocity was 3.1 m/s at 300 Hz. With this new method wave velocity over a distance of 5 mm, and therefore stiffness of arteries, can be measured locally and non-invasively. Measurement time is short in a few tens of milliseconds, which allows pressure dependence and pharmacological effect on the wall properties to be measured at different cardiac times.

4aBB4. Vibro-acoustography: Recent advances and future promises. Mostafa Fatemi, Azra Alizad, and James F. Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, fatemi@mayo.edu)

Vibro-acoustography is an imaging modality that uses ultrasound as the source of energy to explore objects at two ends of the spectrum. Employing two ultrasound beams at slightly different frequencies, this technique produces a highly localized radiation stress field that vibrates the object at the difference frequency. In response, the object produces a secondary acoustic field. This sound, which contains information about the dynamics of the object at the low difference frequency as well as objects characteristics at the ultrasound range, is used to produce an image of the object. This combination makes vibro-acoustography a unique tool to study biological tissues at low and high frequencies simultaneously. Utility of this method has already been demonstrated for a wide range of applications *in vitro*. *In vivo* applications of vibro-acoustography have been demonstrated in a recent study on human breast. It has been shown that this technique can produce high quality *in vivo* images at ultrasound intensities within the recommended FDA limits. Vibro-acoustography may also be used for imaging solids, such as metallic implants. This paper describes recent achievements in vibro-acoustography as well as possible future role of this technology in medicine. [This abstract includes materials patented by MF and JFG.]

10:00–10:30 Break

Contributed Papers

4aBB5. A bioreactor for the dynamic mechanical stimulation of vocal-fold fibroblasts based on vibro-acoustography. Roger W. Chan (Otolaryngol. Head & Neck Surgery, Grad. Program in Biomed. Eng., Univ. of Texas Southwestern Medical Ctr., Dallas, TX 75390-9035, roger.chan@utsouthwestern.edu) and Maritza Rodriguez (UT Southwestern Medical Ctr., Dallas, TX 75390)

During voice production, the vocal folds undergo airflow-induced self-sustained oscillation at a fundamental frequency of around 100–1000 Hz, with an amplitude of around 1–3 mm. The vocal-fold extracellular matrix (ECM), with appropriate tissue viscoelastic properties, is optimally tuned for such vibration. Vocal-fold fibroblasts regulate the gene expressions for key ECM proteins (e.g., collagen, fibronectin, fibromodulin, and hyaluronic acid), and these expressions are affected by the stress fields experienced by the fibroblasts. This study attempts to develop a bioreactor for cultivating cells under a micromechanical environment similar to that *in vivo*, based on the principle of vibro-acoustography. Vocal-fold fibroblasts from primary culture were grown in 3D, biodegradable scaffolds, and were excited dynamically by the radiation force generated by amplitude modulation of two confocal ultrasound beams of slightly different frequencies. Low-frequency acoustic radiation force was applied to the scaffold surface, and its vibratory response was imaged by videostroboscopy. A phantom tissue (standard viscoelastic material) with known elastic modulus was also excited and its vibratory frequency and amplitude were measured by videostroboscopy. Results showed that the bioreactor was capable of delivering mechanical stimuli to the tissue constructs in a physiological frequency range (100–1000 Hz), supporting its potential for vocal-fold tissue engineering applications. [Work supported by NIH Grant R01 DC006101.]

4aBB6. Modulated acoustic radiation pressure and stress-coupling projections. Philip L. Marston and David B. Thiessen (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

Low-frequency deformation can be induced at a single frequency using radiation stress oscillations of double-sideband suppressed-carrier ultrasound [P. L. Marston and R. E. Apfel, J. Acoust. Soc. Am. **67**, 27 (1980)]. The transducer voltage is proportional to a product of low- and high-frequency sine waves. To anticipate the shape and magnitude of induced deformations, it is helpful to expand the distribution of the radiation stress on the object to be deformed as a series of projections [P. L. Marston, J. Acoust. Soc. Am. **67**, 15 (1980)]. Stress projections are also useful for unmodulated waves: the radiation force is an example. In addition to spherical and nearly spherical objects, recent experiments and calculations have concerned cylindrical objects [S. F. Morse, D. B. Thiessen, and P. L. Marston, Phys. Fluids **8**, 3 (1996); W. Wei, D. B. Thiessen, and P. L. Marston, J. Acoust. Soc. Am. **116**, 202 (2004)]. In standing waves the following projections are nonvanishing in the low acoustic frequency limit for appropriately situated dense objects: radial projection [M. J. Marr-Lyon, D. B. Thiessen, and P. L. Marston, Phys. Rev. Lett. **86**, 2293 (2001)] and quadrupole projection [P. L. Marston *et al.*, J. Acoust. Soc. Am. **69**, 1499 (1981)].

4aBB7. Non-dissipative mechanisms of radiation force and shear wave generation. Lev Ostrovsky (Zel Technologies/NOAA ETL, Boulder, CO 80305), Yurii Ilinskii (Univ. of Texas, Austin, TX 78713), Armen Sarvazyan, and Alexander Sutin (Artann Labs., Lambertville, NJ 08530)

We describe new mechanisms of shear stress generation in tissue by radiation force (RF) that are not related to attenuation or reflection of ultrasound waves. The suggested theoretical model is based on the five-constant theory for elastic solids that is extended to waterlike biological media. It is shown that in the absence of dissipation, a potential and linear sound beam in a homogeneous medium creates a potential radiation pressure rather than the shear stress. However, in inhomogeneous non-dissipative media, that is in a lossless medium with spatially varied linear and/or nonlinear parameters, the shear stress generation by this RF is possible. Additional non-dissipative mechanism is associated with the an-harmonicity of the primary beam. Although the corresponding effect is proportional to the third power of the beam amplitude, it can be significant in the focal area. Estimates demonstrate that the contribution of the above mechanisms to the value of the generated RF can be comparable with and even exceed the classical dissipative radiation force in realistic situations. Experiments with inhomogeneous tissue mimicking phantoms are presented and experimental results are interpreted in the framework of the above theory. [Work partly supported by NIH grant.]

4aBB8. Radiation force of ultrasound as shear wave source in microscopic magnetic resonance elastography. Shadi F. Othman, M. Bulent Ozer, Huihui Xu, Thomas J. Royston (Univ. of Illinois at Chicago, 842 West Taylor St., MC 251, Chicago, IL 60607, troyston@uic.edu), and Richard L. Magin (Univ. of Illinois at Chicago, Chicago, IL 60607)

Microscopic magnetic resonance elastography (micro-MRE) is a high-resolution imaging technique for measuring the viscoelastic properties of small synthetic and biological samples. Taking MRE to the microscopic scale requires stronger static fields, stronger magnetic field gradients, higher performance RF coils, and more compact, higher frequency shear wave actuators. Prior work by our group has been conducted at 11.74 T. A needle attached to a vibrating cantilever beam was placed in contact with the surface of the sample to generate shear waves up to 800 Hz. At higher frequencies, the excited shear waves attenuate within an extremely short distance such that only a very small region in the vicinity of the actuator can be studied due to inherent dynamic range limitations. In principle, modulated focused radiation force of US should be able to create a localized shear wave source within the test sample at a distance from the US transducer, thereby enabling micro-MRE probing of the sample at very high frequencies (up to 5 kHz). A confocal US transducer was fabricated to create such a source within the working constraints of the micro-MRE system. Initial feasibility studies are reviewed in this presentation. [Research supported by NIH Grant No. EB004885-01.]

4aBB9. Finite-difference time-domain approach to acoustic radiation force problems. Glauber T. Silva (Departamento de Tecnologia da Informacao, Universidade Federal de Alagoas, Maceio, AL, Brazil 57072-970)

Acoustic radiation force plays a major role in elastography methods such as vibro-acoustography, acoustic radiation force, shear wave elasticity, and supersonic shear wave imaging. The radiation force (dynamic or static) exerted on an object by an incident wave can be obtained by solving the acoustic scattering problem for the object. However, only in rather simple cases the scattering of waves can be described by exact analytical expressions. In this work, we developed an algorithm based on the finite-difference time-domain (FDTD) method to compute the radiation force exerted on arbitrary shaped objects. The algorithm simulates the wave propagation in a finite extended medium with an embedded object. The radiation force is obtained by numerically calculating a surface integral of the momentum flux, which depends on the incident and scattered fields. Absorbing boundary conditions are used to truncate the medium. We compute the radiation force exerted on a rigid and soft cylinder by a plane wave. Results are in agreement with the theoretical predictions. Discrepancies due to numerical dispersion in the algorithm are under investigation. The presented method might be used to calculate the radiation force on complex objects present in elastography techniques. [Work supported by FAPEAL/CNPq, Brazil.]

4aBB10. Stress field formation for multifrequency vibro-acoustography: A simulation study. Matthew Urban, Mostafa Fatemi, and James Greenleaf (Mayo Clinic and Foundation, 200 First St. SW, Rochester, MN 55905)

Vibro-acoustography is a method that uses the dynamic radiation force (stress) of ultrasound to locally excite an object at low frequency [Fatemi and Greenleaf, Science **280**, 82]. A multifrequency vibro-acoustography method is proposed that uses a multifrequency radiation stress produced by an array transducer driven with N ultrasound frequencies. A multifrequency image at the different $N(N-1)/2$ frequencies is produced with one scan of the region of interest, increasing the information yield by a factor of $N(N-1)/2$. Image formation theory is presented for multifrequency vibro-acoustography using an annular array transducer. Simulations of the stress field were performed for a 20-element, 3.5-MHz annular array transducer. Four ultrasound frequencies produce six unique low-frequency components in the point spread function. Computer phantoms with small spheres are used to demonstrate the usefulness of multifrequency vibro-acoustography for microcalcification detection in breast imaging. The lateral and axial width of the main lobe at -6 dB is 0.68 and 6.47 mm, respectively. The lateral and axial sidelobe levels are -27.8 and -16.8 dB, respectively. The proposed method holds the potential for large gains of information with good spatial resolution and no increase in scanning time. [Work supported by Grant EB00535-03 from NIH.]

Session 4aID

Interdisciplinary: Electronic Tools and Services for ASA Authors and Readers

Allan D. Pierce, Chair

Boston Univ., Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Chair's Introduction—9:00

Many powerful tools and services are available to authors and readers who publish in and use *The Journal of the Acoustical Society of America* and *Acoustics Research Letters Online*. This seminar will feature presenters from the American Institute of Physics who host the online versions of Acoustical Society journals.

9:05−10:00

Tools and Services for Authors

A brief overview of the online manuscript submittal process for the ASA journals using the all-electronic paper management system, Peer X-press.

10:00−11:00

Tools and Services for Readers

A demonstration of research tools available through AIP's online platform, "Scitation," including searchable databases, reference linking featuers, article collection folders, citation exports, email alerts, and RSS data feeds to assist readers who access ASA and AIP journals.

THURSDAY MORNING, 20 OCTOBER 2005　　　　　　　　　　　　　CONRAD D, 10:00 TO 11:25 A.M.

Session 4aMU

Musical Acoustics and Architectural Acoustics: Acoustics of Choir Singing I

Sten O. Ternstrom, Cochair

Kungliga Tekniska Hogskolan, Speech Music and Hearing, Lindstedtsvagen 24, S-10044 Stockholm, Sweden

Thomas D. Rossing, Cochair

Northern Illinois Univ., Physics Dept., DeKalb, IL 60115

4a THU. AM

Chair's Introduction—10:00

Invited Papers

10:05

4aMU1. Multitrack analysis of amateur and professional choirs. Harald Jers (Franzstrasse 33, D-50931 Cologne, Germany, harald.jers@gmx.de)

Many choir singers in the world know the fascinating phenomenon of choir sound, which is the result of multiple voices singing in an ensemble. This so-called chorus effect, where the normal mechanisms of auditory localization of the single voices are disrupted, may be caused by complex interactions between the choir singers, but has not been researched in detail. Each singer of an amateur and a professional vocal ensemble of 16 singers was recorded on separate tracks while singing in the choir. The evaluation of different choir pieces and exercises provided information and predictions about $F0$, SPL, timing/synchronization, vibrato behavior, and the produced choir sound. The results reveal differences between amateur and professional choirs for homophonic and polyphonic choir pieces, and suggest new considerations for choir rehearsals and concert performances.

10:30

4aMU2. The effects of choir spacing and choir formation on the tuning accuracy and intonation tendencies of a mixed choir. James F. Daugherty (Div. of Music Education and Music Therapy, The Univ. of Kansas, 1530 Naismith Dr., Ste. 448, Lawrence, KS 66045, jdaugher@ku.edu)

The tuning accuracy and intonation tendencies of a high school mixed choir ($N=46$) were measured from digital recordings obtained as the ensemble performed an *a cappella* motet under concert conditions in $N=3$ singer spacing configurations (close, lateral, circumambient) and $N=2$ choir formations (sectional and mixed). Methods of analysis were modeled on Howard's (2004) pitch-based measurements of the tuning accuracy of crowds of football fans. Results were discussed in terms of (a) previous studies on choir spacing (Daugherty, 1999, 2003) and self-to-other singer ratios (Ternstrm, 1995, 1999); (b) contributions of choir spacing to vocal/choral pedagogy; and (c) potential ramifications for the design and use of auditoria and portable standing risers for choral performances.

Contributed Papers

10:55

4aMU3. Directivity of singers. Harald Jers (Franzstrasse 33, D-50931 Cologne, Germany, harald.jers@gmx.de)

Studies of acoustical balance between singers within a choir by means of room acoustical measurements have shown that the directional sound propagation of the source is important. For this reason the directivity of female and male singers for different vowels has been measured in this investigation. Measurements of a pilot study and some first measurements in 1998 have been supplemented with new measurements and an enhanced setup. A special measurement setup with reference and recording microphones was used to collect the directivity data. A resolution of 10 deg for azimuth and elevation angle was obtained. The results will be shown in 3D spherical plots with frequency adjustments in semitones from 80 to 8000 Hz. The measurements are compared to an artificial singer's directivity,

and the influence of a sheet music binder in front of a singer will be shown. The results give information on the directivity of singers and are relevant for the prediction of self-to-other-ratios that result from placement and formation aspects within a choir.

11:10

4aMU4. Auditorium design for choral performance. Timothy Foulkes and Christopher Storch (Cavanaugh Tocci Assoc. Inc., 327F Boston Post Rd., Sudbury, MA 01776)

Design for a 500 seat recital hall to support an award winning high school choral program is discussed. Acoustic design strategy, important acoustic parameters (calculated and measured), and photos of the completed project are reviewed.

Note: Session 4pMU, Acoustics of Choir Singing II, will conclude at Central Lutheran Church with a panel discussion and a short Concert by the St. Olaf Cantorei

---

THURSDAY MORNING, 20 OCTOBER 2005　　　　　　　　　　　　　　CONRAD A, 9:00 TO 11:00 A.M.

Session 4aPA

Physical Acoustics: Topics in Atmospheric Acoustics

Roger M. Waxler, Chair

Univ. of Mississippi, National Center for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

9:00

4aPA1. Investigation of two-dimensional numerical shock propagation using a weighted essentially non-oscillatory scheme. Mark S. Wochner and Anthony A. Atchley (Grad. Prog. in Acoust., Penn State Univ., University Park, PA 16802)

This work is focused on the computational simulation of realistic jet noise propagation in air. Consequently, in addition to accurate absorption, dispersion, wave steepening, etc., propagation of multidimensional shocks is necessary. Time-domain propagation of shocks requires some special technique to ensure stability and this is usually accomplished using artificial viscosity. In this investigation, instead of using artificial viscosity, a weighted essentially non-oscillatory (WENO) scheme is used. The WENO scheme is unique in that it uses a combination of multiple stencils to calculate derivatives. Shock propagation is achieved in this method through its use of smoothness indicators, which work to leave out any candidate stencils that contain a discontinuity in the calculation of derivatives. Its ability to handle a wide variety of acoustical conditions and the future potential of such a model is shown. [Work supported by ONR and the Pennsylvania State University Applied Research Laboratory Exploratory and Foundational Program.]

9:15

4aPA2. Modeling volcanic infra-sound propagation. Joseph F. Lingevitch, Geoffrey F. Edelmann, Douglas P. Drob, and Michael D. Collins (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375, jfl@nrl.navy.mil)

Active volcanoes are significant sources of infra-sound and can be exploited for studying atmospheric sound. The propagation is influenced by the vertical and horizontal wind structure where wind-induced refraction can trap or disperse sound energy, leading to large variations of propagation with wind conditions. Stratospheric winds influence long-range observations and exhibit variations on seasonal, diurnal, and semi-diurnal time scales. These effects impact the observability and interpretation of acoustic signals at remote listening stations. The propagation of infra-sound over propagation paths of several hundred kilometers from volcanoes in the Archipelago of Vanuatu in Oceania is studied over a frequency range of 0.1–4 Hz. Atmospheric wind and sound-speed profiles constructed every 6 h for a 2-year period are used as inputs to a parabolic equation propagation model. The results are compared with experimental observations and other models to assess the validity of our numerical simulations. [Work supported by ONR.]

9:30

4aPA3. Atmospheric sound propagation at higher altitudes. Xiao Di, Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677), and Richard Clark (Millersville Univ., Millersville, PA 17551)

During the daytime, upward refraction dominates atmospheric sound propagation. An elevated sensor can potentially give much higher sound-pressure levels and higher signal coherence. To understand daytime upward atmospheric sound propagation at higher altitudes, a comprehensive sound propagation experiment for receivers at various heights (up to 220 m) was performed in November 1998. In the experiment, both meteorological data and acoustic data were taken. Three single tone sources were used in the experiment (210, 380, and 600 Hz). The horizontal propagation distances were 300, 600, and 900 m. In this presentation, we will present the experimental results and calculations of sound-pressure level and coherence using the measured meteorological data as input. Numerical calculations are compared with the experimental data. Both experimental data and theoretical predictions show much higher sound-pressure levels and higher coherence at higher attitudes.

9:45

4aPA4. The pressure distribution around foam spheres outdoors. Jeremy Webster, Richard Raspet, and Jiao Yu (Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

A previous investigation of the pressure fluctuations around foam windscreens outdoors indicated both lower pressure levels and lower correlations between microphones than expected from wind noise theory [J. Acoust. Soc. Am. **116**, 2517]. The measurement set was limited and suffered from uncertainty in the incoming wind direction. In this paper we show results from several new measurements which use a three-axis sonic anemometer to establish the wind direction and speed simultaneous to pressure correlation measurements. [Prepared in part through collaborative participation in the Collaborative Technology Alliance for Advanced Sensors sponsored by the U.S. Army Research Laboratory.]

10:00

4aPA5. Performance comparison of compact cylindrical and spherical windscreens. Qamar A. Shams, Allan J. Zuckerwar, and Bradley S. Sealey (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23681)

An improved and compact windscreen was conceived for a microphone of a type used outdoors to detect atmospheric infrasound from a variety of natural and artificial sources. This cylindrically shaped compact windscreen, made of closed-cell polyurethane foam, had dimensions of 0.0762 m i.d. ×0.2286 m height ×0.0127 m wall (3×9×0.5 in.). The low acoustic impedance of the foam showed a transmission coefficient near unity. The same closed-cell polyurethane foam was used to fabricate a spherical windscreen of 0.254 m (10 in.) diameter with 0.0127-m (0.5 in.) wall thickness. In this paper the effectiveness of cylindrical and spherical windscreens is evaluated using a low-speed wind tunnel facility that enables controlled and repeatable experiments. Wind noise levels are characterized for both windscreens as a function of frequency and mean wind velocity.

10:15

4aPA6. Soaker hose versus compact nonporous windscreen: A comparison of performance at infrasonic frequencies. Allan J. Zuckerwar, Qamar A. Shams (NASA Langley Res. Ctr., M.S. 238, Hampton, VA 23681, Qamar.A.Shams@nasa.gov), Krish K. Ahuja, and Robert Funk (Georgia Inst. of Technol., Atlanta, GA 30332)

A compact nonporous windscreen described previously [J. Acoust. Soc. Am. **114**, No. 4, Pt. 2, 2323 (2003)] was tested in the field against a soaker hose array to compare performance at infrasonic frequencies. The cylindrically shaped compact windscreen, made of closed-cell polyure-thane foam, had dimensions 0.0762 m i.d. ×0.2286 m height ×0.0127 m wall (3×9×0.5 in.). The low acoustic impedance of the foam permits the propagation of infrasound through the walls of the windscreen with a transmission coefficient near unity. The soaker hoses were 15.24-m (50 ft.) long and coupled to a Chaparral model 5 low-frequency microphone. The hose plenum was removed and replaced with the compact windscreen for testing. A sonic boom simulator, located at a distance of 400 m (1/4 mile) from the microphone, generated tones at 3, 4, 5, and 6 Hz. Analysis of the signals received by the interior microphone revealed that the tones transmitted through the windscreen, as well as the reduction of background noise due to naturally occurring wind-generated turbulence, were nearly identical for the two types of windscreen.

10:30

4aPA7. Low-frequency ground impedance from the surface mode. Roger Waxler, Carrick L. Talmadge, and Kenneth E. Gilbert (The Natl. Ctr. for Physical Acoust., 1 Coliseum Dr., University, MS 38677)

At the last meeting of the Acoustical Society of America we presented data and theoretical analysis showing that pulse arrivals in downward refracting atmospheres have a universal low-frequency tail. This tail was identified with the surface mode. In this presentation we discuss the possibility of obtaining low-frequency ground impedance from the surface mode arrival. Signals received on a 10-m vertical array are analyzed. The vertical waveform of the surface mode is obtained as a function of frequency and used to infer a ground impedance. The inferred ground impedance is compared to impedances predicted by ground impedance models.

10:45

4aPA8. Outdoor sound propagation effects on aircraft detection through passive phased-array acoustic antennas: 3D numerical simulations. Ivan Roselli, Pierluigi Testa, Gaetano Caronna (Dept. of Tech. Phys., Faculty of Eng., Univ. of Rome "La Sapienza," Via Eudossiana 18 - 00184, Rome, Italy), Andrea Barbagelata, and Alessandro Ferrando (D'Appolonia S.p.A., 16145, Genoa, Italy)

The present paper describes some of the main acoustic issues connected with the SAFE-AIRPORT European Project for the development of an innovative acoustic system for the improvement of air traffic management. The system sensors are two rotating passive phased-array antennas with 512 microphones each. In particular, this study focused on the propagation of sound waves in the atmosphere and its influence on the system detection efficiency. The effects of air temperature and wind gradients on aircraft tracking were analyzed. Algorithms were implemented to correct output data errors on aircraft location due to acoustic ray deviation in 3D environment. Numerical simulations were performed using several temperature and wind profiles according to common and critical meteorological conditions. Aircraft location was predicted through 3D acoustic ray triangulation methods, taking into account variation in speed of sound waves along rays path toward each antenna. The system range was also assessed considering aircraft noise spectral emission. Since the speed of common airplanes is not negligible with respect to sound speed during typical airport operations such as takeoff and approach, the influence of the Doppler effect on range calculation was also considered and most critical scenarios were simulated.

4a THU. AM

Session 4aSA

Structural Acoustics and Vibration: Vibration, Radiation and Scattering

Courtney B. Burroughs, Chair

Noise Control Engineering, 1241 Smithfield St., State College, PA 16801

Contributed Papers

9:00

4aSA1. Analysis of broadband acoustic radiation from periodic structures. Pavel Danilov (Continuum Dynam., Inc., 34 Lexington Ave., Ewing, NJ 08618) and Donald Bliss (Duke Univ., Durham, NC 27708)

Radiation from a 2D finite-length plate with periodic discontinuities modeled by line impedances, and with arbitrary boundary conditions, is considered. Under broadband excitation such structures often exhibit a fairly simple directivity pattern that is nearly symmetric and relatively smooth. The differences between single-frequency and broadband response are analyzed. Due to discontinuities the structure exhibits passband and stop-band behavior, with structural waves decaying in stop bands. This behavior leads to a different energy distribution between the structure and the acoustic field for frequencies in different bands. Acoustic radiation from periodic structures is characterized in terms of Bloch waves. Closed-form formulas for harmonic and broadband response from a single structural Bloch wave are obtained. Total response is recovered by superposition of Bloch waves. A single Bloch wave contains a primary harmonic, corresponding to a long-wavelength, smooth global solution, and higher spatial harmonics induced by discontinuities which describe the short-wavelength local solution. Radiation characteristics are discussed in terms of global and local solutions.

9:15

4aSA2. Numerical simulation of an aircraft style panel excited by a random pressure field. Noah H. Schiller, Chris R. Fuller (Dept. of Mech. Eng., Virginia Polytechnic Inst. and State Univ., Blacksburg, VA 24061, nschille@vt.edu), and Randolph H. Cabell (NASA Langley Res. Ctr., Hampton, VA 23681)

This study describes the development of a model that predicts the sound radiation from an aircraft style panel excited by a dynamic pressure field with arbitrary spatial correlation. Eventually, this model will be used to develop distributed feedback control strategies for systems with spatially correlated inputs. For this study, the input pressure field is simulated using an array of point forces on the panel. The excitation at each point is defined in the time domain using the spectral representation method [M. Shinozuka and C.-M. Jan, J. Sound Vibr. **25**(1), 111–128 (1972)]. This method generates sample functions that match the spatial and temporal correlation characteristics of the desired pressure field. Three different types of excitations are considered: a turbulent boundary layer excitation, a diffuse field, and a spatially uncorrelated disturbance. This approach could be extended to a variety of other types of spatially correlated stochastic inputs as well. A convergence study is also presented to show how many point forces are required to approximate the spatial characteristics of each type of pressure field. [Work supported by NASA.]

9:30

4aSA3. A boundary-element method using broadband vibrating-wall sources to predict high-frequency interior sound fields produced by wall vibration. Linda P. Franzoni and Tracy A. DuVall (Mech. Eng., Duke Univ., Box 90300, Durham, NC 27708-0300)

In the high-frequency limit, vibrating panels subject to spatially random, temporally broadband forcing are shown to have broadband power and directivity properties than can be characterized by a limited set of parameters, based on numerical simulations. The radiated pressure field is parametrized in terms of direction, wave speed ratio, panel damping, and dimensionless frequency. A source directivity equation dependent on these variables is presented. The radiation properties of this equation are incorporated to simulate vibrating wall panels in an energy/intensity–based boundary-element method (BEM) developed for the prediction of steady-state, broadband, reverberant sound fields in enclosures having either diffusely or specularly reflecting boundaries. The BEM method uses uncorrelated broadband directional intensity sources to construct the source and reflection sound fields and predict mean-square pressure distributions in enclosures. Because uncorrelated broadband directional intensity sources are used, the system does not require a frequency-by-frequency-based solution, thereby reducing computational expense. Simulations are compared to exact solutions obtained by computationally expensive frequency-by-frequency modal methods. When fully developed, the directed application of this method is aircraft interior noise caused by exterior boundary layer excitation on fuselage panels.

9:45

4aSA4. Sound transmission modeling of advanced multilayered composite structures using enhanced T-matrices for two-phase materials. Roland L. Woodcock and Rebecca S. Bryant (United Technologies Res. Ctr., 411 Silver Ln., M.S. 129-17, East Hartford, CT 06108)

Advanced composite structures have been used for many years in the aerospace industry. When designing multilayered structures special attention must be paid to the bonding techniques since the interface conditions have a direct effect on the mechanical coupling between the individual layers. Previous studies have shown the overall acoustical performance such as transmission loss and surface absorption to be sensitive to this structural path mainly in the lower frequency range. The state of the art shows that a comprehensive model is still lacking in the framework of the transfer matrix Method. The present paper proposes a new analytical modeling approach to tackle systems with stiffeners in the low-frequency range. This technique is based on the so-called Series-Parallel network of the transmission line theory in the framework of the classical electro-acoustical analogies. The simulations show that in the typical case of a double plate with stiffeners, with regards to transmission loss the design change due to the mechanical path is captured and the increase of the overall stiffness of the system shifts the resonance to the higher frequencies. The other important acoustical properties of multilayered systems will be highlighted during the presentation with regards to optimizing the overall acoustical performance.

10:00

4aSA5. Axisymmetric acoustic scattering from submerged prolate spheroidal shells. Jeffrey E. Boisvert (NAVSEA Newport, Newport, RI 02841) and Sabih I. Hayek (Penn State Univ., University Park, PA 16802)

The equations of motion for nonaxisymmetric vibration of prolate spheroidal shells of constant thickness were derived using Hamilton's principle [S. I. Hayek and J. E. Boisvert, J. Acoust. Soc. Am. **114**, 2799–2811 (2003)]. The shell theory used in this derivation includes shear deformations and rotatory inertias. The shell displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular

coordinate and circular functions in the azimuthal angle coordinate. The shell is insonified by a plane wave incident along the major axis. The external (heavy) fluid loading impedance was computed using an eigenfunction expansion of prolate spheroidal wavefunctions. Far-field scattered acoustic pressure spectra are presented for several shell thickness-to-half-length ratios ranging from 0.005 to 0.1, and for various shape parameters, a, ranging from an elongated spheroidal shell ($a = 1.01$) to a spherical shell ($a \sim 100$). The far-field directivity of acoustic scattering is presented at selected frequencies. [Work supported by the ONR/ASEE Summer Faculty Research Program.]

10:15−10:30   Break

10:30

4aSA6. Local parametric identification for *in situ* health monitoring of aircraft using distributed sensors. Philippe Micheau, Jérôme Pinsonnault, and Patrice Masson (GAUS, Mech. Eng. Dept., Université de Sherbrooke, Sherbrooke, QC J1K 2R1, Canada, Philippe.Micheau@USherbrooke.ca)

The objective of this study is to provide a strategy for *in situ* structural health monitoring with MEMS sensors in order to reduce the high costs associated with periodic prescribed inspections of aircraft. The presented strategy focuses on the flexural waves in the medium-frequency range in order to obtain a good trade-off between damage localization and distant propagation, and to be efficient for composite materials. The method consists of performing local parametric identification of equivalent MISO systems in order to detect a parametric discontinuity when a defect is present. Simulations were performed with a hierarchical trigonometric functions set on an aluminum plate (100 cm × 75 cm × 1 mm) with a crack (5 cm × 0.007 in.). For a chirp excitation in the medium-frequency range at 1 kHz, wave propagation simulation shows the scattering around the crack in the plate. A distribution of MEMS sensors over the plate is efficient to detect a significant change in the identified local parameters, and consequently to localize the crack. However, this approach is limited to configuration with high signal-to-noise ratio. Experimental validation of the model is conducted both in frequency and time domains for healthy and cracked beams and plates. [Work supported by the Consortium for Research and Innovation in Aerospace in Quebec.]

10:45

4aSA7. Investigation of the resonance frequency shift in parts with cracks. Krasimir Zahariev, Yulian Kin (Purdue Univ. Calumet, 2200 169th St., Hammond, IN 46323, kin@calumet.purdue.edu), and Alexander Sutin (Stevens Inst. of Technol., Hoboken, NJ 07030)

It is known that development of crack in various parts leads to resonance frequency variation and that phenomena can be used for crack detection and remaining lifetime prediction. We have investigated this effect on a steel specimen (25×150×6 mm). The crack was initiated at the root of preliminary machined notch and propagated under cycling loading on fatigue machine. The finite-element analysis was applied for calculation of frequency shift for three flexural modes of vibration and it was observed that the frequency shift increases with the increase of crack size. The maximum detected frequency shift was 3.8% for the crack size 23% of a sample width. The experimental measurements were conducted by measurements of sound produced by free vibration of the sample after impact excitation (impact resonance acoustic spectroscopy). The relative difference between acoustical measurements and FEA results did not exceed 1%. The conducted research provides a good basis for development of the remaining life prediction methods, for example, by Paris formulation. Parameters of the formulation for our case were determined experimentally. [Research supported by 21 Century Fund of Indiana.]

11:00

4aSA8. Effects of material anisotropy on the amplitudes of waves in cylindrical structures. J. Gregory McDaniel (Dept. of Aerosp. and Mech Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, jgm@bu.edu)

Composite materials may be designed and configured to substantially alter the generation and propagation of elastic waves with a relatively small weight penalty. The anisotropy introduced by composite structures leads one to intuitively expect that wave amplitudes will depend strongly on the orientations of the applied force as well as the composite material. The present work investigates this effect for cylindrical structures with point forces. Using a mixture of analytic and numerical models, an understanding of the relationship between force directivity, material orientation, and wave amplitudes is developed. Results of this study lead directly to design strategies for controlling power flow to the structure. [Work supported by ONR.]

11:15

4aSA9. Tri-axial hybrid (passive-active) vibration isolator using single crystal piezoelectric actuators. Peter Herdic, Robert Corsaro, Brian Houston, and Robert Baden (Naval Res. Lab., Code 7130, Washington, DC 20375)

Results are presented for a laboratory study of a tri-axis hybrid (active-passive) vibration isolator. A previous report [J. Acoust. Soc. Am. **112** (2002)] described control studies performed using single-axis control. It identified the optimum physical control law by examining the placement order within the device consisting of a passive compliant spring, sensing layer, and piezoelectric actuator. Additionally, actuator materials were evaluated by examining the performance levels of three piezoelectric materials having different nonlinearity values: PZT-4, PZT-5A, and PMN-PT single crystal. The study described here extends this previous work to the required, more complicated case of three-axis control. Described are the development, design, and fabrication of a hybrid device that demonstrates stable, robust, and reproducible "local" vibration reductions of $\sim 30 - -45$ dB simultaneously in all three excitation axes. These results are found for both single and multiple tone tri-axis disturbances and with negligible out-of-band enhancement, where the total harmonic distortion of the system is less than 0.25%. The control approach, using base acceleration minimization, demonstrates ~ 35 dB of "downstream" performance, which is ~ 25 dB higher than that of the passive element alone. The device is sufficiently small and simple to fabrication that it offers potential for being used in practical applications.

11:30

4aSA10. Minimization of resonance excitation of moving actuator and joint structures. Svetlana Kovinskaya (Mechmath LLC, 14530 Bluebird Trail, Prior Lake, MN 55372, mechmath@mechmath.com) and Yuvenaly Khozikov (IEE Russian Acad. of Sci., St. Petersburg 195265, Russia)

Numerous devices have actuators moving from one immobile position to another under action of applied forces. Repositioning of actuators requires its acceleration and deceleration for faster moving. It is provided by subsequent application of constant forces of opposite directions. Duration of their action is much longer than the resonance periods of actuators or joint structures, and such constant forces excite no resonance. However, a sudden change of the force value results in the resonance excitation. Minimization of resonance excitation can be performed by control of the applied force. Considering the problem in the time domain, one must minimize energy at excited resonance by shaping the force profile while keeping the fast actuator reposition. An effective algorithm for shaping of initial profile based on an iterative procedure is suggested. At the first step of the iteration the profile is deformed (to minimize resonance excitation) by shaping functions which are equal unit outside a short instance. At the second step the deformed initial profile is shaped by the polynomial to meet basic requirements on moving (given distance and zero velocity by the end of reposition). Employment of the developed algorithm significantly reduces the residual vibration and noise radiated by a device.

Session 4aSC

Speech Communication: Phonetic Linguistics: Honoring the Contributions of Peter Ladefoged

Patricia A. Keating, Chair

Univ. of California, Los Angeles, Dept. of Linguistics, Los Angeles, CA 90095-1543

Chair's Introduction—8:30

Invited Papers

8:35

4aSC1. 50 years of phonetics with Peter Ladefoged.　Jenny Ladefoged　(Phonet. Lab., Univ. of California, Los Angeles, CA 90095-1543)

Ladefoged's many different takes on phonetics, in many countries, will be described.

9:00

4aSC2. Peter Ladefoged and phonetics in the field.　Ian Maddieson　(Dept. of Linguist., Univ. of California, Berkeley, Berkeley, CA 94720)

Among many distinctive contributions to phonetics by Peter Ladefoged is an insistence on the immense diversity of phonetic phenomena in the languages of the world, particularly at the segmental level. Because of this Peter has maintained a flexible approach to any scheme of classification or description, adapting to both new approaches and new data. Perhaps more than any other phonetician he has expected to find surprises, and has gone to far corners of the world in search of them. His ground-breaking *Phonetic Study of West African Languages* from 1964 laid out a template for synthesizing a large mass of data which is echoed in later works such as *Preliminaries to Linguistic Phonetics* and *Sounds of the World's Languages*. His widely used Course in Phonetics and other textbooks have shown generations of students the richness of spoken sound. But not only novel research results have marked Peter's contributions to this aspect of phonetics; equally significant are his methodological innovations in taking experimental techniques to the field. Some of the most salient steps along both these paths will be reviewed from personal and professional perspectives.

9:25

4aSC3. The influence of Peter Ladefoged.　Kenneth N. Stevens　(Res. Lab of Electron, Dept. Elec. Engin. and Comp. Sci., and Div. Health Sci. and Tech., Massachusetts Inst. of Technol., Cambridge, MA 02139)

The research and publications of Peter Ladefoged and his students have been a strong influence in elevating the field of phonetics to a science, and in bringing together phonetics and phonology. I will cite some examples of how Peter's work and that of his colleagues has influenced the research of many of us at MIT and around the world, where attempts to develop unifying theories often need to undergo continuous revision based on the meticulous articulatory and acoustic data that he reports in his books and other publications. These include his descriptions of the involvement of the tongue root in vowel systems, the fricatives of Mandarin Chinese, the inventory of places of articulation he catalogs in the languages of the world, and the many uses of laryngeal states in providing phonological contrasts. [Supported in part by grants from NIDCD.]

9:50−10:05　　Break

10:05

4aSC4. Taking the measure of phonetic structure.　Louis Goldstein　(Haskins Labs. and Dept. of Linguist., Yale Univ., 370 Temple St., New Haven, CT 06511, louis.goldstein@yale.edu)

From his earliest work, Peter Ladefoged has insisted that phonetic dimensions and categories (e.g., vowel charts, IPA symbols, features) must be measurable if they are to provide an adequate basis for a universal phonetic representation. This view has led him to two questions, the pursuit of which has been enormously informative and has set the agenda for much contemporary phonetic research. (1) What are the appropriate reference frames in which to make these measurements? The debate triggered by Ladefoged as to the correct choice of reference frame is a key component of current theoretical discussion in phonetics. While the empirical data yield tantalizing hints, their interpretation is far from unambiguous, as can be seen even in some of Ladefoged's earliest work. (2) Are there universal fixed phonetic categories within these reference frames from which individual languages choose (and if so, what are they)? Here, Ladefoged's research has been at the leading edge of a growing consensus that there are, in fact, no fixed categories and that the division of continua into categories in particular languages is largely a random process, constrained by the nature of speech production and perception. [Work supported by NIH.]

4aSC5. Betting on bits: Contextual influences on the perception of "phonetic categories." Sarah Hawkins (Dept. of Linguist., Univ. of Cambridge, Sidgwick Ave., Cambridge CB3 9DA, UK)

Peter Ladefoged's outstanding qualities include meticulous description of differences between "the same" phonetic categories in different languages, and a tendency to explore new research areas, not least by investigating things that others judge better left alone. This paper pays tribute to these qualities by discussing the phonetic category as a theoretical construct in speech perception. By demonstrating that $F1$ of a preceding phrase influences whether a word is heard as "bet" or "bit," Ladefoged and Broadbent (1957) stressed that phonetic categorization is plastic: sense depends on context. Many other contextual influences on sound categorizaton have been identified, yet are poorly integrated into theory, especially when they extend over long durations, or indicate something other than lexical contrast. I argue from (human) perceptual and (animal) neurophysiological and behavioral data that category identification depends on context and perceived relevance—hence on potential meaning. Human data confirm the centrality of context and meaning to categorization. Animal data show that a sound which is behaviorally significant produces long-lasting changes to response characteristics of single neurons in the primary auditory cortex, and probably earlier. Such data encourage radical re-evaluation of how sensory information is used in understanding speech. [Funded by the Leverhulme Trust.]

4aSC6. Pronunciation models for conversational speech. Keith Johnson (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

Using a pronunciation dictionary of clear speech citation forms a segment deletion rate of nearly 12% is found in a corpus of conversational speech. The number of apparent segment deletions can be reduced by constructing a pronunciation dictionary that records one or more of the actual pronunciations found in conversational speech; however, the resulting empirical pronunciation dictionary often fails to include the citation pronunciation form. Issues involved in selecting pronunciations for a dictionary for linguistic, psycholinguistic, and ASR research will be discussed. One conclusion is that Ladefoged may have been the wiser for avoiding the business of producing pronunciation dictionaries. [Supported by NIDCD Grant No. R01 DC04330-03.]

Contributed Papers

4aSC7. Phonological acquisition of a Korean child: An acoustic study. Sun-Ah Jun (Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543, jun@humnet.ucla.edu)

Studies on child phonology suggest that there exist phonological universals in the timing of phonological events and the ordering of phonological categories, but the acquisition of speech sounds is influenced by the language-specific aspects of the ambient language such as phonetics, phonology, and the frequency of the sound in child-directed speech. This study investigates a Korean child's phonological acquisition based on tape recordings of longitudinal data (from 2 months to 2 years, recorded in 1- to 2-week intervals). Special attention is given to the change in prosody and the acquisition of the Korean three-way manner contrast (fortis, aspirated, lenis). It is known that Korean fortis and aspirated obstruents trigger high pitch at vowel onset while lenis obstruents trigger low pitch [Jun (1993), (1998)]. Preliminary results suggest that fortis obstruents are acquired first, followed by aspirated, and then lenis. The segmental properties (e.g., voice onset time, breathy phonation) appropriate for the lenis category were acquired later than the pitch. In addition, unlike the universal tendencies, velar and labial consonants were acquired earlier than alveolar consonants. Factors affecting the order of acquisition, including frequency effect and perceptual salience, will be discussed.

4aSC8. Comparing the acoustics of voiced and voiceless fricatives in Deg Xinag. Richard Wright, Sharon Hargus, and Julia Miller (Dept. of Linguist., Univ. of Washington, Box 354340, Seattle, WA 98195-4340)

Few studies have looked at the acoustic properties of fricative voicing and place in Native American languages despite their relatively rich fricative inventories of rarely studied fricative places. Deg Xinag, an endangered Athabaskan language spoken in Alaska, provides us with a rare opportunity to investigate fricative place and voicing within a single language: it has eight places of articulation for voiceless fricatives, six of which have voiced counterparts, including some rarely studied place contrasts (e.g., palato-alveolar versus retroflex, uvular versus glottal, lateral versus alveolar). In this study, pre- and post-vocalic fricatives were digitally recorded in the field from eight speakers (two males, six females) using a head-mounted mic to control for distance from the source. The segmental context was also controlled for, the neighboring vowel being [a] in all cases. Each speaker produced four repetitions of each word. Each fricative was analyzed qualitatively using impressionistic transcription and spectrographic investigation, and quantitatively using a set of widely employed measures: (a) widely employed spectral measures (center of gravity, skew, kurtosis, standard deviation, lowest spectral peak), peak and rms intensity of frication, overall duration and duration of voicing. [Work supported by NSF.]

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, Animal Bioacoustics, Noise, Acoustical Oceanography and Engineering Acoustics: Non-Blind Deconvolution in Acoustics

George B. Smith, Cochair

Naval Research Laboratory, Code 7183, Stennis Space Ctr., MS 39529-5004

Juliette Ioup, Cochair

Univ. of New Orleans, Dept, of Physics, New Orleans, LA 70148

George E. Ioup, Cochair

Univ. of New Orleans, Dept, of Physics, New Orleans, LA 70148

Invited Papers

9:00

4aSP1. Introduction to deconvolution. Peter A. Jansson (College of Optical Sci., Univ. of Arizona, 1630 E. University Blvd., Tucson, AZ 85721)

Deconvolution tasks will always lie at the frontier of human knowledge in many fields, almost by definition. Rising in the latter 20th century from near disreputability, first to usefulness, then necessity in some disciplines, the varied techniques of deconvolution have assumed an important role in the scientists tool kit. This talk will trace deconvolutions development with examples, including many "firsts," drawn from spectroscopy, radio astronomy, photography, cell biology, color science and diverse other fields. Following a tutorial introduction, detail will be provided on modern super-resolving methods and lesser known topics such as selected-ordinate image (SORI) processing.

9:45

4aSP2. Nonlinear deconvolution approaches. B. Roy Frieden (College of Optics, Univ. of Arizona, Tucson, AZ 85721)

A review of past methods of digital signal restoration is given. Emphasis is upon nonlinear methods, owing to their extra precision in allowing inequalities and other forms of prior knowledge to be entered as constraints. Some specific methods of this type are maximum entropy, maximum bounded entropy, maximum shannon information, recursive median window filtering, Jansson recursion, and minimum Fisher Information. Algorithms and demonstrations will be given.

10:30

4aSP3. Identification of rapidly time-varying wideband acoustic communication channels. Weichang Li (MIT and Woods Hole Oceanograph. Inst., M.S. 16, Woods Hole, MA 02543, lwc@mit.edu) and James Preisig (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Wideband acoustic communications in very shallow-water are typically confronted with channels that are both rapidly time varying and sparsely structured [J. C. Preisig and G. Deane, J. Acoust. Soc. Am. **116(4)**, 2067–2080 (2004)]. Techniques for coherent demodulation require either implicit or explicit identification of channel impulse response. Model based channel identification is able to capture the channel dynamics only when certain persistent excitation condition involving both the transmitted symbol sequence and the sequence of channel estimates holds. This is generally not the case for channels with highly sparse impulse response. On the other hand, dimension reduction techniques such as subspace decomposition and sparse processing, inherently assuming that the channel is stationary over certain time period, often fail to track the rapid channel fluctuations. Therefore both the channel dynamics and redundant dimension have to be addressed simultaneously in order to achieve accurate and robust channel identification. This paper presents methods of model based channel tracking with sparse preprocessing. The limitations of model based tracking and sparse processing when applied separately are derived. The performance gain obtained from the combining approach is then demonstrated through experimental results. [Work supported by ONR Ocean Acoustics.]

11:00–11:30
Panel Discussion

Session 4pAAa

Architectural Acoustics: Distinguished Lecture: From Philharmonic Hall to Number Theory: The Way to More Diffusion

Peter D'Antonio, Cochair

RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Trevor Cox, Cochair

RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Chair's Introduction—2:00

Invited Paper

2:05

4pAAa1. From Philharmonic Hall to number theory: The way to more diffusion. Manfred R. Schroeder (Univ. of Göttingen, Rieswartenweg 8, 37077 Göttingen, Germany, mrs17@aol.com)

In September 1962, in the presence of Mrs. Jacqueline Kennedy, Philharmonic Hall in New York was inaugurated—the first building of the new Lincoln Center for the Performing Arts. To address the soon-apparent acoustic problems, Lincoln Center turned to Bell Laboratories for help, and I was asked to join a "committee of experts," chaired by Vern O. Knudsen of UCLA. My work on Philharmonic Hall, assisted by B.S. Atal, G.M. Sessler, and J.E. West, and later, after my move to Göttingen, by my students D. Gottlob, F.K. Siebrasse, and U. Eysholdt, indicated a need for energetic early *lateral* sound. It was clear that better lateral diffusion could improve the acoustic quality and the feeling of "envelopment" by the sound. Knowing some Galois field mathematics, I lucked upon the design of diffusors which scattered incident waves into broad lateral patterns—but only for a single musical octave. Then, in 1977, during a celebration of the 200th anniversary of Gauss's birth, I heard a talk by André Weil on Gauss sums and quadratic residues and, in a flash, it became clear to me that diffusors based on *quadratic residues* were the answer to broadly scattering waves comprising many musical octaves.

2:50–3:00
Question and Answer Period

Session 4pAAb

Architectural Acoustics: Reflections on Reflections

Peter D'Antonio, Cochair
RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Trevor Cox, Cochair
RPG Diffusor Systems Inc., 651C Commerce Dr., Upper Marlboro, MD 20774

Invited Papers

3:05

4pAAb1. Thirty years since diffuse sound reflection by maximum length. Trevor J. Cox (Acoust. Res. Ctr., Univ. of Salford, Salford M5 4WT, UK) and Peter D'Antonio (RPG Diffusor Systems, Upper Marlboro, MD 20774)

This year celebrates the 30th anniversary of Schroeder's seminal paper on sound scattering from maximum length sequences. This paper, along with Schroeder's subsequent publication on quadratic residue diffusers, broke new ground, because they contained simple recipes for designing diffusers with known acoustic performance. So, what has happened in the intervening years? As with most areas of engineering, the room acoustic diffuser has been greatly influenced by the rise of digital computing technologies. Numerical methods have become much more powerful, and this has enabled predictions of surface scattering to greater accuracy and for larger scale surfaces than previously possible. Architecture has also gone through a revolution where the forms of buildings have become more extreme and sculptural. Acoustic diffuser designs have had to keep pace with this to produce shapes and forms that are desirable to architects. To achieve this, design methodologies have moved away from Schroeder's simple equations to brute force optimization algorithms. This paper will look back at the past development of the modern diffuser, explaining how the principles of diffuser design have been devised and revised over the decades. The paper will also look at the present state-of-the art, and dreams for the future.

3:35

4pAAb2. Measuring, predicting, and characterizing reflections. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774) and Trevor J. Cox (Univ. of Salford, Salford M5 4WT, UK)

Over the past 100 years, since the founding of architectural acoustics by Sabine, there has been considerable effort devoted to studying surface absorption. Over this period, a considerable library of absorption coefficients has been tabulated, based on accepted standards of measurement, and the role of absorptive surfaces is now well understood. In contrast, scientific knowledge about how scattering surfaces should be measured and characterized and how, where, when, and why these surfaces should be used is still evolving. Over the past three decades, how to design, optimize, predict, measure, and characterize the performance of scattering surfaces has been learned. This presentation will describe the procedures developed to measure, predict, and quantify scattering surfaces. Standardized procedures for determining scattering coefficient and correlation scattering coefficient, which indicate the amount of sound scattered away from the specular direction, and the diffusion coefficient, which indicates the uniformity of scattering, will be described and examples given. Three-dimensional balloons, which display the polar distribution of scattered sound, will also be displayed. The goal is that, as these standards become more widely understood, a growing library of coefficients will develop and soon absorption, scattering, and diffusion coefficients will all find their way into architectural acoustic specifications.

4:05

4pAAb3. Reflection of sound from finite-size plane and curved surfaces. Jens H. Rindel (Oersted-DTU, Tech. Univ. of Denmark, Bldg. 352, DK-2800 Kgs. Lyngby, Denmark)

The author's research on reflectors over nearly 25 years is summarized. The influence of curvature was analyzed by a geometrical model in order to quantify the attenuation by a simple expression. Reflection from a finite-size plate was studied using the Kirchhoff–Fresnel approximation, and the design frequency for a single reflector was derived. Above the design frequency the attenuation due to the finite size can be neglected and the reflection is efficient in the specular direction. The method was extended to the case of a reflector array, and it was demonstrated that the performance of a reflector array can improve if the size of the panels is decreased. The same design frequency applies to a single reflector and a reflector array, but with different meaning; in the latter case the design frequency is the upper limit for useful reflections. This design rule was first used in the refurbishment of the concert hall of the Danish Radio in Copenhagen 1989, and later in many other halls. In order to describe the scattering due to edge diffraction, the directional characteristic of reflections from a finite-size plate has been studied and a simple approximation valid for octave bands has been derived.

4:25

4pAAb4. Virtual reflections in electronic acoustic architecture. Bjorn van Munster (P.O. Box 720, 5400 AS UDEN, The Netherlands b.v.munster@siap.nl)

In the era of the ancient Greeks and Byzantines, the first attempts for increasing reverberation time are noted. In the 1950s, the Ambiophonic system accomplished this by means of an electronic device, for the first time. The early systems only increased the reverberation time by delaying the picked-up reverberation. With the introduction of multichannel feedback-based systems, the reverberation level also could be increased. Later, it was understood that it was important to also fill in the missing reflections, address reflection density, frequency dependence, etc. This resulted in the development of the SIAP concept. Current DSP technology led to the development of a processor whereby density, length, level, and the frequency content can be controlled for different areas in the same room or different rooms, leading to the concept of the acoustic server. electronic acoustic architecture has become the current state-of-the-art approach for solving acoustic deficiencies in, among others, rehearsal rooms, theaters, churches, and multipurpose venues. Incorporation of complementary passive acoustic solutions provides an optimum solution for all room problems. This paper discusses the utilization of virtual reflections in the new approach of electronic acoustic architecture for different environments. Measurements performed in the Sejong Performing Arts Centre, Seoul, South Korea, show the power of this approach.

4:45

4pAAb5. Reflections in computer modeling of rooms: Current approaches and possible extensions. U. Peter Svensson (Acoust. Group, Dept. of Electron. and Telecommunications, Norwegian Univ. of Sci. and Technol., Trondheim, NO-7491 Trondheim, Norway)

Computer modeling of rooms is most commonly done by some calculation technique that is based on decomposing the sound field into separate reflection components. In a first step, a list of possible reflection paths is found and in a second step, an impulse response is constructed from the list of reflections. Alternatively, the list of reflections is used for generating a simpler echogram, the energy decay as function of time. A number of geometrical acoustics-based methods can handle specular reflections, diffuse reflections, edge diffraction, curved surfaces, and locally/non-locally reacting surfaces to various degrees. This presentation gives an overview of how reflections are handled in the image source method and variants of the ray-tracing methods, which are dominating today in commercial software, as well as in the radiosity method and edge diffraction methods. The use of the recently standardized scattering and diffusion coefficients of surfaces is discussed. Possibilities for combining edge diffraction, surface scattering, and impedance boundaries are demonstrated for an example surface. Finally, the number of reflection paths becomes prohibitively high when all such combinations are included as demonstrated for a simple concert hall model. [Work supported by the Acoustic Research Centre through NFR, Norway.]

5:05

4pAAb6. A largely diffuse small room. George Massenburg (Blackbird Studios, 2806 Azalea Pl., Nashville, TN 37204) and Peter D'Antonio (RPG Diffusor Systems, Inc., Upper Marlboro, MD 20774)

An 8 m×10 m rectangular room is described that will have broad bandwidth diffusion completely covering 5 of its 6 surfaces, i.e., all walls and ceiling surfaces. The dimensions of the room were optimized for minimal standard deviation of the modal response. The wall surfaces are treated with a large prime, single period, number theoretic diffusor 1.2 m deep that wraps around the entire room. The ceiling consists of a 12×13 low frequency diffusor 2 m deep, which is further treated with mid-high frequency diffusors to form a nested, diffractal surface. These surfaces will be described further and illustrated. The room is intended to be deployed as a monitor room for mixing surround sound. It is hoped that the unique combination of a reduced number of specular surfaces and very neutral, wide-band ambience will improve localization (particularly for virtual sources) and offer greater support for (and speed up) the balancing/mixing of multitrack sources. The perceived effectiveness of the diffusor in the ceiling, particularly in the 2 low octaves will be described. The room will also be used as a massively diffuse recording room for various musical formats.

Session 4pAB

Animal Bioacoustics, Psychological and Physiological Acoustics and ASA Committee on Standards: Frequency Weighting For Animal Species

Larry L. Pater, Cochair

U.S. Army Engineering Research and Development Ctr., 2909 Farber Dr., Champaign, IL 61822

Ann E. Bowles, Cochair

Hubbs Sea World Research Inst., 2595 Ingraham St., San Diego, CA 92109

Chair's Introduction—1:00

Invited Papers

1:05

4pAB1. The 60-dB rule for birds: An example of the application of a weighting function in environmental impact mitigation. Ann E. Bowles (Hubbs-SeaWorld Res. Inst., 2595 Ingraham St., San Diego, CA 92109) and Sheyna Wisdom (URS Corp., San Diego, CA 92108)

Over the last decade U.S. Fish and Wildlife Service managers in California have required millions of dollars in added expenditure for NEPA consultation, mitigation barriers, and project delays to reduce the effects of noise from construction activities on endangered passerine birds when the hourly A-weighted Leq is expected to exceed 60 dB. The rule was originally intended to prevent masking of species-typical songs of endangered birds such as the Coastal California Gnatcatcher. However, no research is available to demonstrate the effectiveness of the rule for any noise-related impact. Although A-weighting is probably a conservative estimator of bird exposure in the range from 125 Hz to 8 kHz, it may underestimate exposure at very low frequencies. Its utility as a weighting function has not been tested against other possible weighting procedures, such as use of the species-typical auditory threshold function. Additionally, where sources are intense but intermittent, Leq is unlikely to be a useful metric. These issues should receive more technical scrutiny before the 60-dB rule becomes entrenched in law. It is in widespread use for NEPA consultations, and is already being extended to other species, including large mammals.

1:35

4pAB2. The problem of frequency weighting functions and standards for birds. Robert Dooling, Elizabeth Brittan-Powell, Amanda Lauer (Dept of Psych., Univ. of Maryland, College Park, MD 20742), Micheal Dent (Univ. at Buffalo, Buffalo, NY 14260), and Isabelle Noirot (Univ. of Maryland, College Park, MD 20742)

Frequency weighting functions in humans are widely used as a single-figure guess to assess noise problems and aid in making decisions with regard to noise limitations when no other data exist. However, this use of frequency weightings invariably results in a loss of precision in assessing the likelihood of a sound to produce hearing damage or sound annoyance. There is a growing interest in developing frequency weighting functions in animals presumably to assist in judging the risk of hearing damage, interference with acoustic communication, or habitat suitability. Laboratory studies reveal many parallels between humans and animals on a variety of psychoacoustic measures, such as equal loudness contours. However, differences between humans and animals on specific tests argue against using standards developed for humans to gauge the effect of noise on animals. Here we review data which show this same problem exists among birds. That is, the differences in the effects of noise among bird species can be as large as the differences between humans and birds. These results suggest that whereas frequency weighting functions and acoustic standards for a specific species might be useful, generalizing across species is likely not practical.

2:05

4pAB3. Source levels of northern elephant seal vocalizations in-air. Stephen J. Insley (Long Marine Lab., UCSC, 100 Shaffer Rd., Santa Cruz, CA 95060, sinsley@ucsc.edu) and Brandon L. Southall (NOAA Fisheries Acoust. Program, Silver Spring, MD 20910)

Accurate measurements of vocalization sound-pressure levels are necessary to determine the acoustical active space of animals in natural and human-altered ambient noise conditions. Despite this basic need, such data are limited or nonexistent for most species. Our study characterized aerial ambient noise and vocalization source levels for northern elephant seals during the breeding season. Subjects were adult males, lactating females, and dependent offspring (pups) at Año Nuevo State Reserve. Source level measurements were made using a Type 1 sound level meter and calibrated microphones on-axis: (1) at 1 m; (2) at several known distances (laser measured); and (3) simultaneously at 1 m and a second known distance. Concurrent ambient noise conditions were measured *in situ* (non-weighted 5 min Leq integrated averages) and recorded for later spectral analysis. Measurements were made at two sites, one relatively noisy and the other relatively quiet, to determine whether animals compensate for higher noise conditions by increasing source levels (Lombard effect). Results indicate a wide range in signal strength, particularly for adult males whose vocalization source levels appear to be correlated with dominance rank and related to ambient noise conditions. The Lombard effect was not observed for adult females or elephant seal pups.

4pAB4. Strategies for weighting exposure in the development of acoustic criteria for marine mammals. James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett Bay Campus Narragansett, RI 02882, miller@uri.edu), Anne E. Bowles (Hubbs-Sea World Res. Inst., San Diego, CA 92109), Roger L. Gentry (NOAA Acoust. Prog., Silver Spring, MD 20910-6233), William T. Ellison (Marine Acoust., Inc., Litchfield, CT 06759), James J. Finneran (Space and Naval Warfare Systems Ctr., San Diego, CA 92152-5000), Charles R. Greene Jr. (Greeneridge Sci., Inc., Santa Barbara, CA 93110), David Kastak (Long Marine Lab., Univ. of California at Santa Cruz, Santa Cruz, CA 95060), Darlene R. Ketten, Peter L. Tyack (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Paul E. Nachtigall (Hawaii Inst. of Marine Biol., Kane'ohe, HI 96744), W. John Richardson (LGL Ltd. Environ. Res. Assoc., King City, ON, Canada L7B 1A6), and Jeanette A. Thomas (Western Illinois Univ., Moline, IL 61265)

The Noise Exposure Criteria Group has been developing noise exposure criteria for marine mammals. Although the primary focus of the effort is development of criteria to prevent injury, the Group has also emphasized the development of exposure metrics that can be used to predict injury with accuracy and precision. Noise exposure metrics for humans have proven to be more effective when they account for psychophysical properties of the auditory system, particularly loudness perception. Usually noise is filtered using the A-weighting function, an idealized curve based on the human 40-phon equal loudness function. However, there are no empirical studies to show whether a comparable procedure for animals will improve predictions. The Noise Exposure Criteria Group panel has proposed to weight noise data by functions that admit sound throughout the frequency range of hearing in five marine mammal groupings—low frequency cetaceans (mysticetes), midfrequency cetaceans, high-frequency cetaceans, pinnipeds in air, and pinnipeds in water. The algorithm for the functions depends only on the upper and lower frequency limits of hearing and does not differentially weight frequencies based on sensitivity within the range. This procedure is considered conservative. However, if the human case may be taken as a model, it is not likely to produce precise predictions. Empirical data are essential to finding better estimators of exposure.

3:05–3:20 Break

Contributed Papers

3:20

4pAB5. Experimental validation of a species-specific behavioral impact metric for underwater noise. Jeremy R. Nedwell, Joe Lovell, and Andrew W. H. Turnpenny (Subacoustech Ltd, Chase Mill, Winchester Rd., Bishop's Waltham, Hampshire SO32 1AH UK)

The regulation of noise from offshore activities in the UK requires a metric allowing the behavioral effects on underwater animals of man-made underwater noise on a wide range of species to be objectively assessed. The dBht(species) metric is a pan-specific metric incorporating the concept of "loudness" by using a frequency-weighted curve based on the species' hearing threshold as the reference unit for a dB scale. A large number of controlled laboratory measurements have been made of the avoidance of a range of idealized noises, using fish with greatly different hearing as a model. Additional data, of many thousands of individuals, has been obtained by re-evaluation of fish avoidance of a large acoustic fish deflection system at an estuarine power station. All data, irrespective of source or species, indicate a dependence of avoidance reaction on the dBht(species) level. The data indicates three regions, "no reaction" below 0dBht (i.e., below the species' threshold of hearing), a "cognitive avoidance" region where increasing numbers of individuals will avoid the noise from 0 to 90 dBht, and "instinctive reaction" at and above 90 dBht where all animals will avoid the noise. This probabilistic model allows the behavioral impact of any noise source to be estimated.

3:35

4pAB6. Frequency weighting in the feature extraction process: Effects of parameter choice on generalized perceptual linear prediction coefficients. Patrick J. Clemins and Michael T. Johnson (Speech and Signal Processing Lab., Marquette Univ., P.O. Box 1881, Milwaukee, WI 53201-1881)

The generalized perceptual linear prediction (gPLP) feature extraction model incorporates information about the perceptual abilities of the species under study to generate features relevant to that species. The gPLP feature extraction model is based on the source-filter model of vocalization production and quantifies the general shape of the spectral envelope. gPLP coefficients remove excitation information for vocalizations with a low fundamental frequency, but capture harmonic information for vocaliza-

tions with a higher fundamental frequency through the use of a filter bank analysis component. Frequency warping, frequency masking, and equal loudness normalization, or frequency weighting, are some of the psychophysical phenomena modeled in the gPLP model. The effects of accounting for these phenomena in the feature extraction process are explored using perceptual spectrograms, statistical tests, and classification tasks. Experiments show that using this perceptual information can provide insights during the analysis of vocalizations and improve classification accuracies. The contribution of other feature extraction parameters including the number of cepstral coefficients and number of filters in the filter bank is also examined. [Work supported by the National Science Foundation under Grant No. IIS-0326395.]

3:50

4pAB7. Detection sensitivity of frequency modulated (FM) and constant frequency (CF) signals in temporal masking conditions in the Mongolian gerbils. Tomofumi Taketani, Yasunari Sasaki, and Hiroshi Riquimaroux (Grad. School of Eng., Doshisha Univ., 1-3, Miyakodani, Tatara, Kyotanabe, Kyoto 610-0321, Japan, dtf0757@mail4.doshisha.ac.jp)

The auditory systems would be sensitive to temporal changes in frequency of sounds. We hypothesized that frequency modulation would play an important role in perception. And in this study, the difference in detection sensitivity between frequency-modulated signals (FM) and constant frequency signals (CF) in temporal masking conditions was examined in three Mongolian gerbils (*Meriones unguiculatus*) by GO-NOGO procedure. We focused on their detection sensitivity of three types of signals; upward FM, downward FM, and CF. The frequency range of each FM was from 4 to 7 kHz, and CF was 6 kHz. We set $S+$ that signals were presented between white noise bursts (BW: 0 to 17.5 kHz), and then signals were temporally masked by white noise. In contrast, we set $S-$ in which silence was presented instead of signals. Subjects were trained to detect whether signals were presented or not. After subjects reached criterion (80%), each stimulus was presented randomly and attenuated in amplitude. We investigated their response rate to each $S+$. Result showed that their detection sensitivity to FM was higher than that of CF. It suggests that FM is superior to CF in recognition for subjects. [Research was supported with financial assistance from MEXT.]

Session 4pMU

Musical Acoustics and Architectural Acoustics: Acoustics of Choir Singing II

Sten O. Ternstrom, Cochair

Kungliga Tekniska Hogskolan, Speech Music and Hearing, Lindstedtsvagen 24, S-10044 Stockholm, Sweden

K. Anthony Hoover, Cochair

Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776

Chair's Introduction—2:00

Invited Papers

2:05

4pMU1. Choir singing in Subsaharan Africa: Acoustic factors of a regional style in southern Mozambique. Joao Soeiro de Carvalho (Universidade Nova de Lisboa, Av. de Berna 26c, 1069-061 Lisboa, Portugal, jsoeiro@fcsh.unl.pt)

Choir singing is a most prominent form of expressive behavior in Subsaharan Africa. A vast majority of expressive modes involves multipart singing, both within the framework of European tonal system as well as other structured ways of combining vocal sounds of different frequencies. Vocal improvisation stands as an important process for the course of performance; individual voice ranges, as well as issues of social status and musical competence, determine the ways musicians participate in performance. Aesthetic validation is often expressed by the use of a nonverbal expressive mode, "kulungwani," a vocal technique involving the action of the lower maxillae and tongue in order to produce a low-frequency interruption of sound emission. Choral singing intonation processes seem to rely on harmonic results, rather than melodic. A regional choral style in southern Africa seems to have developed, where a particular distribution of formant frequencies and an emphasis on low-frequency energy play a significant role.

2:30

4pMU2. Listener perception of and acoustic differences between girl and boy choristers in an English cathedral choir. David Howard (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK), and Graham Welch (Univ. of London, London, WC1H 0AL, UK)

For centuries, boy choristers have been singing the top (treble) line in English cathedrals. Girl choristers were first admitted in 1991, and there is a long-running debate as to whether they can carry out this role appropriately. This paper will detail the results from two listening experiments designed to establish whether or not listeners can tell the difference between girl and boy choristers singing the top line in cathedral music. In the first experiment, 189 listeners took part and on average they were able to tell the difference 60% of the time; this was statistically significant over chance. The results suggested that repertoire played a significant part in this ability, and the second experiment was carried out in which the boys and girls sang the same repertoire. Nearly 170 listeners have completed this experiment and, on average, they are making guesses (correct 52% of the time). The paper will discuss the acoustic differences between the stimuli with respect to the singing of boy and girl choristers, while placing the discussion in the context of the English cathedral tradition.

2:55

4pMU3. Making an anechoic choral recording. Ron Freiheit (Wenger Corp., 555 Park Dr., Owatonna, MN 55050, ron.freiheit@wengercorp.com), John Alexander (3M Ctr., St. Paul, MN 55144-1000), and John Ferguson (St. Olaf College, Northfield, MN 55057)

The utilization of auralization as a tool for acoustic analysis continues to grow and develop. An important element for successful auralization listening experiences is the selection of anechoic source material. In researching the current library of anechoically recorded source material, it was discovered that choral material was not readily available. The Wenger Corporation, St. Olaf College,

and 3M undertook a joint project to create an anechoic choral recording. The paper describes the challenges of this recording project—from the technological, logistical, and musical standpoints—and the solutions that were successfully implemented.

3:20–3:50

Panel Discussion

After the session at the Hilton, the session will move to the Central Lutheran Church which is a short walk from the hotel.

4:10–4:30

The session will continue with a panel discussion "Acoustical issues relevant to choral singing" with Drs. Anton Armstrong and John Ferguson, who direct the St. Olaf Cantorei, and the ASA session orgainizers.

It will end with a short concert by the St. Olaf Cantorei from St. Olaf College in Northfield, Minnesota. The 90-voice liturgical Choir will be accompanied by organ, brass quartet, handbells, and percussion. The performance, at Central Lutheran Church, which is also open to the public, will feature a hymnsing in which the audience will be invited to participate.

The St. Olaf Cantorei, which regularly performs in the very reverberant St. Olaf College Chapel, recently made an anechoic recording under the sponsorship of the Wenger Company.

THURSDAY AFTERNOON, 20 OCTOBER 2005　　　　　　　　　　　　　CONRAD A, 2:00 TO 4:15 P.M.

Session 4pPA

Physical Acoustics: Topics in Seismic Acoustics

William C. K. Alberts, II, Chair
Univ. of Mississippi, National Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

2:00

4pPA1. Technique for measurement of characteristic impedance and propagation constant for porous materials. Ki Won Jung and Anthony A. Atchley (Grad. Program in Acoust., Penn State Univ., University Park, PA 16802)

Knowledge of acoustic properties such as characteristic impedance and complex propagation constant is useful to characterize the acoustic behaviors of porous materials. Song and Bolton's four-microphone method [J. Acoust. Soc. Am. **107**, 1131–1152 (2000)] is one of the most widely employed techniques. In this method two microphones are used to determine the complex pressure amplitudes for each side of a sample. Muehleisen and Beamer [J. Acoust. Soc. Am. **117**, 536–544 (2005)] improved upon a four-microphone method by interchanging microphones to reduce errors due to uncertainties in microphone response. In this paper, a multiple microphone technique is investigated to reconstruct the pressure field inside an impedance tube. Measurements of the acoustic properties of a material having square cross-section pores is used to check the validity of the technique. The values of characteristic impedance and complex propagation constant extracted from the reconstruction agree well with predicted values. Furthermore, this technique is used in investigating the acoustic properties of reticulated vitreous carbon (RVC) in the range of 250–1100 Hz.

2:15

4pPA2. Vibration signature of human footsteps on the ground and in buildings. James Sabatier and Alexander Ekimov (NCPA, The Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Vibration and acoustic responses of the ground and floor in a building to human footsteps were experimentally investigated in the broad frequency range, starting from 0.5 Hz up to ultrasonic frequencies. It was experimentally shown that human footsteps have low- and high-frequency components. The low-frequency component is well known in the literature and generated by the force component normal to the ground/floor. This force is a function of a person's weight and a manner of motion (walking, running), or gait. The second high-frequency vibration and acoustical components of human footsteps are generated by the tangential to the ground/floor force from footstep and ground reaction, or friction force. The interactions of these two forces produce the friction noise. The low-frequency forces due to tangential motion between two contacted surfaces generate the high-frequency friction noise. The parameters of this friction noise, such as frequency band and vibration, and sound magnitudes as functions of a distance and walking style (shoes or barefoot) and ground/floor structures were studied. The results of indoor and outdoor tests are presented and discussed. [Work supported by Department of the Army, Army Research Office Contract W911NF-04-1-0190.]

2:30

4pPA3. Nonlinear acoustic behaviors of soils. Zhiqu Lu (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Coliseum Dr., University, MS 38677, zhiqulu@olemiss.edu)

Soil, like other porous materials such as rock, sand, sandstone, sedimentation, and concrete, exhibits remarkable nonlinear behaviors. In this work, the nonlinear acoustic behaviors of soils are studied by measuring both amplitude and phase frequency responses as a function of dynamic strain. Typical nonlinear phenomena such as resonant frequency shift and harmonics generation are observed. It is also found that the phase difference between the received and reference signals changes with strain. New methods for measuring the nonlinearity parameters, based on resonant frequency shift, phase shift, and phase slope measurements, are developed and compared. This suggests that phase could be a very useful parameter for determining the nonlinearity of porous materials. [This work is supported by the USDA Agricultural Research Service under Specific Cooperative Agreement 58-6408-0-108.]

4p THU. PM

2:45

4pPA4. Nonlinear acoustic experiments involving landmine detection: Connections with mesoscopic elasticity and slow dynamics in geomaterials, Part III. Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402) and James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

In nonlinear acoustic detection schemes, airborne sound at two primary tones, f_1, f_2 (closely spaced near an 80-Hz resonance) excites the soil surface over a buried landmine. Due to soil wave interactions with the landmine, a scattered surface profile can be measured by a geophone. Profiles at f_1, f_2, $f_1-(f_2-f_1)$ and $f_2+(f_2-f_1)$ exhibit single peaks; those at $2f_1-(f_2-f_1)$, f_1+f_2 and $2f_2+(f_2-f_1)$ involve higher order mode shapes for a VS 2.2 plastic, inert, anti-tank landmine, buried at 3.6 cm in sifted loess soil [J. Acoust. Soc. Am. **116**, 3354–3369 (2004)]. Near resonance, the bending (softening) of a family of increasing amplitude tuning curves, involving the vibration over the landmine, exhibits a linear relationship between the peak particle velocity and corresponding frequency. Results are similar to nonlinear mesoscopic/nanoscale effects that are observed in granular solids like Berea sandstone. New experiments show that first sweeping up through resonance and then immediately sweeping back down result in different tuning curve behavior that might be explained by "slow dynamics" where an effective modulus reduction persists following periods of high strain. Results are similar to those described by TenCate *et al.* [Phys. Rev. Lett. **85**, 1020–1023 (2000)]. [Work supported by U.S. Army RDECOM CERDEC, NVESD.]

3:00

4pPA5. Nonlinear acoustic landmine detection: Comparison of "off target" soil background and "on target" soil-mine nonlinear effects. Murray S. Korman (Phys. Dept., U.S. Naval Acad., Annapolis, MD 21402)

When airborne sound at two primary tones, f_1, f_2 (closely spaced near a resonance) excites the soil surface over a buried landmine, soil wave motion interacts with the landmine generating a scattered surface profile which can be measured over the "target." Profiles at f_1, f_2, and $f_1-(f_2-f_1)$, $f_2+(f_2-f_1)$, $2f_1-(f_2-f_1)$, f_1+f_2 and $2f_2+(f_2-f_1)$ (among others) are measured for a VS 1.6 plastic, inert, anti-tank landmine, buried at 3.6 cm in sifted loess soil. It is observed that the "on target" to "off target" contrast ratio for the sum frequency component can be ~20 dB higher than for either primary. The vibration interaction between the top-plate interface of a buried plastic landmine and the soil above it appears to exhibit many characteristics of the mesoscopic/nanoscale nonlinear effects that are observed in geomaterials like sandstone. Near resonance, the bending (softening) of a family of increasing amplitude tuning curves, involving the vibration over the landmine, exhibits a linear relationship between the peak particle velocity and corresponding frequency. Tuning curve experiments along with two-tone tests are performed both on and off the mine in an effort to understand the nonlinearities in each case. [Work supported by U.S. Army RDECOM CERDEC, NVESD.]

3:15

4pPA6. Nonlinear acoustic detection of weathered, low compliance landmines. James M. Sabatier, W. C. Kirkpatrick AlbertsII (NCPA, Univ. of Mississippi, University, MS 38677), and Murray S. Korman (U.S. Naval Acad., Annapolis, MD 21402)

Two potential impediments to acoustic landmine detection are soil weathering processes and low compliance landmines. To bury landmines, the soil within a mine diameter is removed and replaced such that bulk density, compression, and shear strength all decrease, leaving an acoustic scar detectable with the linear acoustic measurement technique. After a few soil wetting and drying cycles, this contrast is reduced. Linear acoustic mine detection measurements were made on a low impedance contrast landmine before the first rainfall on several occasions over the subsequent 5 years. During this period of time, both the spatial and frequency resolution had to be increased to maintain an on/off target velocity ratio that allowed detection. In some cases, the landmine remains undetectable. To address this, two-tone nonlinear acoustic measurements have been made

on these landmines. When the landmine is detectable with linear acoustics, two tones are broadcast at the frequency where the on/off target velocity ratio is the largest. For the cases when the landmine is undetectable, a two-tone sweep is performed and the operator observes the real-time velocity FFT, noting nonlinear sidebands. Next, two-tone tests are conducted at these sidebands to determine nonlinear velocity profiles. [Work supported by U.S. Army RDECOM, NVESD.]

3:30

4pPA7. A study of the acoustic behavior of a plastic, blast-hardened, antitank landmine. W. C. Kirkpatrick Alberts II, Roger Waxler, and James M. Sabatier (NCPA, Univ. of Mississippi, University, MS 38677)

Mechanical resonances in landmines are often exploited by some acoustic-to-seismic detection schemes. As such, a study of an acoustically active landmine has been performed. Measurements of the mechanical resonances of the pressure plate reveal seven modes below 1.6 kHz and a large frequency shift of the first symmetric mode of the pressure plate, such that it occurs higher in frequency than the first asymmetric mode. The pressure plate is considered to act as an elastically supported thin plate, and the observed perturbation of the first symmetric mode is believed to be caused by the landmine's blast-hardening mechanism. This is confirmed by a lumped acoustic element model of the system of pressure plate and blast-hardening mechanism. Mass-loading experiments in water show a decrease in resonance frequency with increasing depth that saturates at shallow depths as predicted by Velea *et al.* [J. Acoust. Soc. Am. **115**, 1993 (2004)]. Similar mass-loading measurements in sand reveal an interesting phenomenon; flush burial of the landmine effectively removes the effect of the blast-hardening mechanism on the pressure plate. A qualitative theoretical model of this phenomenon will be presented.

3:45

4pPA8. Effect of particle size on the vibration of plates loaded with granular material. Joseph A. Turner, Liyong Yang, and Wonmo Kang (Dept. of Eng. Mech., W317.4 Nebraska Hall, Univ. of Nebraska, Lincoln, NE 68588)

Acoustic methods of landmine detection are emerging as reliable techniques that are especially well suited for non-metallic mines. These methods rely on the vibrations of the top plate of the mine as it responds to the acoustic excitation. The plate response is complicated by the interaction with the surrounding soil. If the loading soil is granular (e.g., sand), it is expected that particle size will influence the mine response. Experimental results related to this hypothesis are presented here. The first resonant frequency of a sand-loaded plate is measured as a function of sand mass for dry sand of various sizes covering a range from hundreds of microns to a few millimeters. For low values of sand mass, the plate resonance decreases and eventually reaches a minimum. In this regime, there is no dependence on particle size. After the minimum, the frequency increases with additional mass. In this regime, a particle size effect is observed. Models are also presented to support and explain experimental results. The results are expected to impact quantitative aspects of mine detection by acoustic means. [Work supported by ARL.]

4:00

4pPA9. Ultrasound sensing of ground displacement in the presence of surface roughness for landmine detection. Purnima Ratilal, Ninos Donabed, Mark Andrews, Carey Rappaport (Ctr. for Subsurface Sensing and Imaging Systems, Northeastern Univ., 409 Dana Res. Cnt., Boston, MA 02115), and Douglas Fenneman (U.S. Army RDECOM CERDEC Night Vision and Electron., Fort Belvoir, VA 22060)

The presence of surface roughness often degrades the performance of the ultrasound sensor. Here, Green's theorem is applied to analytically model the field reflected off the vibrating ground in the presence of surface

roughness for a pair of ultrasound transducers. In the presence of surface roughness, the total field measured by the sensor is the sum of the reflected wave from the vibrating ground and that scattered off the roughness elements. The scattered field from the rough surface randomizes the phase of

the received field. The performance of the sensor is examined as a function of the area of ground insonified, angle of incidence and frequency of the transducer, measurement time, correlation lengths, and amplitude of surface roughness.

THURSDAY AFTERNOON, 20 OCTOBER 2005   SALON G, 2:00 TO 5:00 P.M.

Session 4pSC

Speech Communication: Speech Production: Models and Processes (Poster Session)

Peter J. Watson, Chair

Univ. of Minnesota, Dept. of Speech Language Hearing Sciences, 164 Pillsbury Dr., SE, Minneapolis, MN 55455

Contributed Papers

All posters will be on display from 2:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 2:00 p.m. to 3:30 p.m. and contributors of even-numbered papers will be at their posters from 3:30 p.m. to 5:00 p.m.

4pSC1. Reliablity of measurements from ultrasound images. Sarah M. Hardin and Stefan A. Frisch (Univ. of South Florida, 4202 East Fowler Ave., PCD1017, Tampa, FL 33704)

As ultrasound imaging gains popularity in phonetic and speech science research, examining the reliability of measures taken from ultrasound images becomes important. This study assesses the reliability of hand measures of ultrasound data collected by graduate student researchers at the University of South Florida ultrasound imaging lab. Speech production data from two different experiments, ''Ultrasound analysis of velar fronting'' (Wodzinski, 2004) and ''Ultrasound study of errors in speech production'' [Frisch, (2003)] were analyzed by two different researchers to obtain inter-rater reliability measures. In addition, one data set was measured twice by the same researcher, once when inexperienced with ultrasound analysis and 7 months later after considerable experience had been gained. The study compared researcher's choice of image to analyze, the measures of the location of articulatory landmarks, and the measures used to quantify articulatory postures. Overall, hand measures of ultrasound images were found to be reliable. There were some differences in the absolute measures obtained, however, different researcher's measures of the same data led to the same conclusions about articulation. In addition, it was found that the measurements of different researchers became more similar to one another with experience.

4pSC2. Midsagittal tongue motion patterns in English: More than pivots and arches. Melissa A. Epstein and Maureen Stone (Biomed. Sci., Univ. of Maryland Dental School, 666 W. Baltimore St., Rm. 5A12, Baltimore, MD 21201, maepstein@alumni.upenn.edu)

The transitional motion patterns of the tongue in the midsagittal plane appear to consist primarily of pivoting and arching. A pivot is a region of minimal motion and an arch is a region of maximal motion. It has been found that the location and number of pivots and arches is dependent on the phonemic content of the articulation. The aim of the present study is to systematically investigate these phonemic effects by constraining the transitions between consonants and vowels using CVCVs. Early results of this work [Epstein and Stone, J. Acoust. Soc. Am. **117**, 2575 (2005)] indicate that these two patterns are extreme descriptions of a continuum of shape patterns in which arches and pivots occur simultaneously. This is because the tongue is a nonrigid body and therefore tongue motions are not homogeneous. Canonical tongue shape alone was not enough to indicate which motion pattern would occur. Therefore, this study considers the effect of

tongue height, place, and shape on transitional patterns. The study will track locations of minimum (pivot) and maximum (arch) motions in the vocal tract to determine the effects of physiological and linguistic constraints.

4pSC3. Muscular hydrostat mechanism for lip protrusion in speech. Emi Z. Murano, Maureen Stone (Dept. Biomed. Sci., Univ. of Maryland Dental School, 666 West Baltimore St., 5A-12, Baltimore, MD 21201, ezm001@dental.umaryland.edu), and Kiyoshi Honda (ATR Human Information Sci., Kyoto 619-0288, Japan)

The lip is an organ consisting mostly of muscle similar to the tongue. While the tongue is known as a muscular hydrostat, it is unclear whether the lip is also. In this paper the muscular hydrostat issue was explored from the anatomical and functional point of view using high-resolution static MRI (hr-MRI; 0.125 mm/pixel) and tagged-cineMRI (t-MRI). A 3-D reconstruction of the lips and its muscles was obtained from hr-MRI during sustained vowels /i/ and /u/. The muscular geometry of the orbicularis oris, mentalis, and depressor labii inferior muscles were superimposed onto the principal strains that depicts compression and expansion of the internal tissue obtained from t-MRI. It is shown that (1) orbicularis oris muscle shape can predict both the borderline of glabrous and hairy skin and the manner in which the lips are protruded; (2) the lips volume is almost identical for both speech tasks; and (3) direction and intensity of compression of orbicularis oris and mentalis muscle bundles imply the role of these muscles in the protrusion appearance. These results indicate that the muscular architecture and volume preserving characteristics of the lips are consistent with a muscular hydrostat. [This work was supported by NIH (USA) and NiCT (Japan).]

4pSC4. Smoothing spline analysis of variance (ANOVA) for tongue curve comparison. Lisa Davidson (Dept. of Linguist., New York Univ., 719 Broadway, 4th Fl., New York, NY 10003)

Ultrasound imaging of the tongue is an increasingly common technique in speech production research. One persistent issue regarding ultrasound data is how to quantify them. Researchers often want to determine whether the tongue shape for an articulation under two different conditions (e.g., consonants in phrase-initial versus phrase-medial position) is the same or different. To address this issue, a method for comparing tongue

curves using a smoothing spline ANOVA has been developed (SSANOVA) [Gu, 2002, Smoothing spline ANOVA models]. The SSANOVA is a technique for comparing curve shapes (splines) for two sets of data to determine whether there are significant differences between the curve types. Data sets contain 8–10 repetitions of each of the relevant tongue curves being compared. If the interaction term of the SSANOVA model is statistically significant, then the groups have different shapes. Since the interaction may be significant even if only a small section of the curves is different (i.e., the tongue root is the same, but the tip of one group is raised), Bayesian confidence intervals are used to determine which sections of the curves are statistically different. SSANOVAs are illustrated with some data comparing obstruents produced in word-final and word-medial coda position.

4pSC5. A local vector coding for high-quality voice analysis/synthesis.

Masashi Ito and Masafumi Yano (Res. Inst. of Elec. Commun., Tohoku Univ., 2-1-1 Katahira, Aoba-ku, Sendai 980-8577, Japan, itojin@riec.tohoku.ac.jp)

Line-type spectrum is observed in frequency responses for voiced sound. The spectrum can be characterized by physical parameters: instantaneous amplitude, frequency, and phase for each component. It is difficult to estimate these parameters for natural utterances accurately by power spectrogram because the sound is usually unstationary. A new method, termed local vector coding (LVC), has been proposed to analyze these sounds. LVC assumes that the time-varying parameters for the input sound can be approximated by simple quadratic functions in a short analysis window. Utilizing the phase responses, LVC can estimate not only instantaneous amplitude and frequency for each component of the input but also their time derivatives. The validity of LVC method is examined by using naturally uttered voiced speech. The averaged estimation errors, defined by the differences between the input and resynthesized signals, are lower than 30 dB of the input energy. It indicates that LVC method is very useful for analyzing natural sounds. In addition, since the parameters of each component obtained by LVC method characterize the vowel quality, any kind of voice can be synthesized/transformed by changing each parameter independently, such as a voice of a male adult to a female voice.

4pSC6. The effect of speaking rate on supersegmentals: An acoustic and perceptual analysis.

Hsin-Huei Chiou and Peter Watson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Shevlin 115, Minneapolis, MN 55455)

Rate manipulation has been used to study change in prosodic contrasts such as emphatic stress. Timing contrasts in stressed words are reduced or eliminated when speaking rate is increased. However, reports of intonation and rate change are mixed. Some studies have reported an increase of $F0$ movement [M. Steppling and A. Montgomery, J. Phonetics **64**, 451–461 (2002)], and other reports have found that $F0$ movement is decreased at faster speaking rates [C. Fougeron and S. Jun, Percept. Psychophys. **26**, 45–69 (1998)]. This study examined the effect of speaking rate on $F0$ and duration in sentences produced with emphatic stress in different sentential position and in declarative and interrogative forms. Essentially, durational contrasts were reduced at faster speaking rates and were more pronounced at slower speaking rates. Intonation, on the other hand, was more pronounced for the fast rate and somewhat reduced for the slow rate. A perceptual component will also be reported that examines a listener's ability to recognize stressed words and mode of sentence production (declarative and interrogative) at different speaking rates.

4pSC7. Tongue movements in vowel-consonant-vowel (VCV) sequences: The effect of consonant length.

Anders Lofqvist (Haskins Labs., 300 George St., New Haven, CT 06511, lofquist@haskins.yale.edu)

This study examined the effect of consonant duration on the tongue movement from the first to the second vowel in VCV sequences, where the consonant is a short or long labial nasal consonant. Lip, tongue, and jaw movements were recorded in native speakers of Japanese using a magnetometer system. Measurements were made of the duration, path, and speed of the tongue movement trajectory between the two vowels. The coordination of the onsets of the lip closing and tongue movements was also studied, as well as the relative part of the trajectory that occurred during the consonant and the vowels. Preliminary results show a robust difference in duration between the long and short consonants, with the long ones about twice as long. The duration of the tongue movement was longer in the long than the short consonants. Both the peak and average speed of the tongue movement were slower in the long consonants. The tongue movement path was slightly longer in the long consonants. These results suggest that speakers adjust the tongue movement trajectory so that a similar relationship between the movement and the consonant closure is maintained in both the long and the short consonants. [Work supported by NIH.]

4pSC8. Laryngeal adjustments for devoicing of /h/: A within-speaker study.

Laura Koenig (Haskins Labs., 300 George St., New Haven, CT 06511)

Past work has investigated cross-speaker and cross-gender differences in voicing of /h/ in English speakers. The purpose of this study was to see whether a phonetically sophisticated speaker could intentionally alter his /h/ voicing patterns, and, if so, how he would effect any changes. One adult male speaker of American English, a trained phonetician and dialectologist, produced approximately 500 repetitions of intervocalic /h/ in short carrier phrases, with differing vowel contexts and loudness levels. In the first block, the speaker produced the utterances normally (i.e., without specific instructions on /h/ production); in the second, he was explicitly asked to devoice his /h/'s. Results indicated that the incidence of devoiced /h/ increased from 2% in the first block to 69% in the second block. On average, the /h/'s in the second block were produced with higher baseline airflows, indicating more extreme laryngeal abduction. This alone did not account for the speaker's devoicing behavior, however, since the soft condition, which had the lowest peak airflows in the second block, had the most devoicing. Voice source measures will be compared between the two blocks to clarify how the speaker altered his laryngeal setting to achieve more devoicing. [Work supported by NIH.]

4pSC9. Simulating and understanding the effects of velar coupling area on nasalized vowel spectra.

Tarun Pruthi and Carol Y. Espy-Wilson (Dept. of Elec. Eng. and Inst. of Systems Res., Univ. of Maryland, College Park, MD 20742)

MRI-based area functions for the nasal cavity of one speaker were combined with the area functions for the vowels /iy/ and /aa/ to study nasalized vowels. The oral cavity was compensated for the falling velum by decreasing the oral cavity area by an amount equal to the increase in the nasal cavity area. Susceptance plots were used along with the simulated transfer functions to understand the effects of velar coupling on nasalized vowel spectra. Susceptance plots of $-(B_p+B_o)$ and B_n suggested significant deviation from the rules suggested by O. Fujimura and J. Lindqvist [J. Acoust. Soc. Am. **49**(2), 541–558 (1971)]. In particular, the plots showed that: (1) the frequency of zero crossings of the susceptance plots changes with a change in the coupling area, and (2) formant frequencies need not shift monotonically upward with an increase in coupling area. Further, as a consequence of (1), and the fact that an increase in the coupling area results in a shift of B_n to the right and $-(B_p+B_o)$ to the left, it is postulated that zero crossings of the two plots can cross each other. [MRI data from Brad Story. Work supported by NSF Grant No. BCS0236707.]

4pSC10. Vowel production in Korean, Korean-accented English, and American English. Jimin Lee and Gary Weismer (Waisman Ctr. and Dept. of Commun. Disord., Univ. of Wisconsin-Madison, 1500 Highland Ave., Madison, WI 53705)

The current study compares vowel formant frequencies and durations produced by ten native speakers of Korean, those same speakers producing American English vowels, and ten native speakers of American English. The Korean speakers were chosen carefully to have a minimum of 2 years, and maximum of 5 years residence in the United States; all speakers were between the ages of 22 and 27. In addition, the native speakers of Korean were chosen, by means of a small-scale dialect-severity experiment, from a larger pool of speakers to achieve some homogeneity in their mastery of English phonetics. The full vowel systems of both languages were explored, and a rate condition was included (conversational versus fast) to test the hypothesis that the English vowel space is modified by rate differently for native speakers of Korean who produce English, versus native speakers of English. Results will be discussed in terms of language- and rate-induced adjustments of the vowel systems under study.

4pSC11. Common modes of vocal tract articulation for vowels. Brad Story (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721-0071)

The purpose of this study was to investigate the spatial similarity of vocal tract shaping patterns across speakers and the similarity of their acoustic effects. Vocal tract area functions for eleven American English vowels were first obtained from six speakers, three female and three male, using magnetic resonance imaging (MRI). Each speaker's set of area functions was then decomposed into mean area vectors and representative modes (eigenvectors) using principal components analysis (PCA). Three modes accounted for more than 90% of the variance in the original data sets for each speaker. The general shapes of the first two modes were found to be highly correlated across all six speakers. To demonstrate the acoustic effects the modes, both isolated and combined, a mapping between the mode scaling coefficients and $[F1,F2]$ pairs was generated for each speaker. The mappings were unique for all six speakers in terms of the exact shape of the $[F1,F2]$ vowel space, but the general effect of the modes was the same in each case. The results tend to support the idea that the modes provide a common system for perturbing a unique underlying neutral vocal tract shape. [Work supported by NIH R01-DC04789.]

4pSC12. Production of Mandarin tone contrasts by musicians and non-musicians. Terry L. Gottfried and Grace Yin-Hwei Ouyang (Dept. of Psych., Lawrence Univ., Appleton, WI 54912-0599)

Previous research has shown that native speakers of American English who are musicians perform better than non-musicians when identifying and producing the four phonemic tones of Mandarin. The present study corroborates these findings and analyzes acoustic properties of non-natives' tonal imitations. Listeners imitated Mandarin two-syllable word phrases that varied the vowel (/li/, /la/, /lu/) and tone (high-level, mid-rising, low-dipping, high-falling) of the first syllable. Four native Mandarin speakers rated the musicians, on average, to be better in their imitation of Mandarin tone 4 (high-falling) than non-musicians. There were no significant differences between groups in how they were rated on the other three tones. Acoustical analyses revealed that non-natives failed to match native speakers both in differences in initial $F0$ and in $F0$ contour (change from initial to final $F0$) across tones. Imitations by musicians did not show significant acoustic differences from non-musicians, except for tone 4, where musicians' imitations showed a significant decrease in $F0$ from initial to final portions of the syllable; the decrease in $F0$ for non-musicians was smaller and not significant. Creaky voice (often present in natives' tone 3 and 4) was observed in many non-native imitations, but was not restricted to tones 3 and 4.

4pSC13. An experimental investigation of velocity fields in divergent glottal models of the human vocal tract. Byron D. Erath and Michael W. Plesniak (Purdue Univ., 500 Allison Rd., West Lafayette, IN 47907, plesniak@ecn.purdue.edu)

In speech, sound production arises from fluid-structure interactions within the larynx as well as viscous flow phenomena that is most likely to occur during the divergent orientation of the vocal folds. Of particular interest are the flow mechanisms that influence the location of flow separation points on the vocal folds walls. Physiologically scaled pulsatile flow fields in 7.5 times real size static divergent glottal models were investigated. Three divergence angles were investigated using phase-averaged particle image velocimetry (PIV). The pulsatile glottal jet exhibited a bimodal stability toward both glottal walls, although there was a significant amount of variance in the angle the jet deflected from the midline. The attachment of the Coanda effect to the glottal model walls occurred when the pulsatile velocity was a maximum, and the acceleration of the waveform was zero. The location of the separation and reattachment points of the flow from the glottal models was a function of the velocity waveform and divergence angle. Acoustic analogies show that a dipole sound source contribution arising from the fluid interaction (Coanda jet) with the vocal fold walls is expected. [Work funded by NIH Grant RO1 DC03577.]

4pSC14. Blind inversion and the perception/production link. John Hogden (CCS-3, Los Alamos Natl. Lab., Los Alamos, NM 87545) and Philip Rubin (Yale Univ. School of Medicine, New Haven, CT 06511)

Whether humans recover articulator positions from acoustics in the course of speech perception has been debated for years. Some of the arguments against recovering articulator positions are relevant for machine speech recognition as well. However, techniques for blind inversion, in which the mapping from acoustics to articulation is learned without known articulator trajectories, were not considered when some of these arguments were being developed. We discuss the impact of blind inversion on the debate, and conclude that there are pragmatic reasons to recover articulation from acoustics, particularly when blind inversion is used and the mapping from articulation to acoustics is many-to-one.

4pSC15. Some articulatory details of emotional speech. Sungbok Lee, Serdar Yildirim, Murtaza Bulut, Abe Kazemzadeh, and Shrikanth Narayanan (Speech Anal. & Interpretation Lab., USC, Los Angeles, CA 90089, sungbokl@usc.edu)

Differences in speech articulation among four emotion types, neutral, anger, sadness, and happiness are investigated by analyzing tongue tip, jaw, and lip movement data collected from one male and one female speaker of American English. The data were collected using an electromagnetic articulography (EMA) system while subjects produce simulated emotional speech. Pitch, root-mean-square (rms) energy and the first three formants were estimated for vowel segments. For both speakers, angry speech exhibited the largest rms energy and largest articulatory activity in terms of displacement range and movement speed. Happy speech is characterized by largest pitch variability. It has higher rms energy than neutral speech but articulatory activity is rather comparable to, or less than, neutral speech. That is, happy speech is more prominent in voicing activity than in articulation. Sad speech exhibits longest sentence duration and lower rms energy. However, its articulatory activity is no less than neutral speech. Interestingly, for the male speaker, articulation for vowels in sad speech is consistently more peripheral (i.e., more forwarded displacements) when compared to other emotions. However, this does not hold for female subject. These and other results will be discussed in detail with associated acoustics and perceived emotional qualities. [Work supported by NIH.]

4p THU. PM

4pSC16. Recognition of voice onset time for use in pronunciation modeling. Abe Kazemzadeh, Sungbok Lee, and Shrikanth Narayanan (Speech Analysis and Interpretation Lab., Univ. of Southern California, University Park Campus, Ronald Tutor Hall 320, Los Angeles, CA 90089)

This study examines methods for recognizing native and accented voiceless stops based on voice onset time (VOT). These methods are tested on data from the Tball corpus of early elementary school children, which includes both native English speakers and Spanish speakers learning English, and which is transcribed to highlight pronunciation variation. We examine the English voiceless stop series, which have long VOT and aspiration, and the corresponding voiceless stops in Spanish accented English, which have short VOT and little aspiration. The methods tested are : (1) to train hidden Markov models (HMMs) based on native speech and then extract the VOT times by post-processing phone-level alignments, (2) to train HMMs with explicit aspiration models, and (3) to train, for each phoneme, different HMMs for native and accented variants. Error rates of 23%–53% for distinguishing phone VOT characteristics are reported for the first method, 5%–57% for the second method, and 0%–36% for the third. The error rates varied depending on the different phones examined. In general, the /p/ and /k/ phones had results that varied more than /t/. These results are discussed in light of each method's usefulness and ease of implementation, and possible improvements are proposed.

4pSC17. Influence of prosodic factors on segment articulations and acoustics in English. Patricia Keating (UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu) and Taehong Cho (Hanyang Univ., Seoul, Korea)

This study investigated effects of three prosodic factors—prosodic boundary, lexical stress, and accent—on articulatory and acoustic realizations of two CV syllables, /nE/ and /tE/. These syllables occurred at the beginning of trisyllabic English nonwords; their position in the larger phrase (prosodic boundary conditions), and whether they were lexically stressed and/or accented (prominence conditions) were varied. Articulatory measurements included linguopalatal contact (by electropalatography) for both C and V, stop consonant seal duration, and C-to-V contact difference; acoustic measurements include nasal duration and energy for /n/; VOT, burst energy and spectral center of gravity for /t/; and F1, vowel duration and vowel amplitude for /E/. We tested whether domain-initial strengthening occurs in the C and/or the V segments independently of stress or accent conditions. We found that the effects of position and of stress/accent can be distinguished in the production and the acoustics of these syllables. One domain-initial effect (greater consonant contact domain-initially) was complementary to one stress/accent effect (greater vowel opening with stress/accent); in other cases the effects overlapped (greater vowel energy and tendency to longer consonant both domain-initially and with stress/accent); in one case they conflicted (less consonant energy domain-initially, more consonant energy with stress/accent).

4pSC18. Mucosal wave velocity. David A. Berry, Zhaoyan Zhang, and Juergen Neubauer (The Laryngeal Dynam. Lab., UCLA Div. of Head and Neck Surgery, 1000 Veteran Ave., Rm. 31-24, Los Angeles, CA 90095-1794)

Mucosal wave velocity correlates with the elastic properties of the laryngeal mucosa, and can often discriminate normal phonation from pathology. In the past, mucosal wave velocity has been measured in the laboratory using two-point measurements. However, more recent laboratory investigations using human excised hemi-larynges have yielded quantitative vibratory data along a large portion of the medial/superior surface of the vocal fold. These data will be used to probe mucosal wave velocity as a function of (1) location on the medial and surface surfaces of the vocal fold, (2) subglottal pressure, (3) elongation force, and (4) adductory force. Such data reveal significant aspects of the mucosal wave, and the impact of this wave on phonation.

4pSC19. Observations of the near-field structures of the glottal flow. Juergen Neubauer, Zhaoyan Zhang, and David Berry (The Laryngeal Dynam. Lab., UCLA Div. of Head and Neck Surgery, 1000 Veteran Ave., Los Angeles, CA 90095)

The three-dimensional glottal jet was measured downstream from a self-oscillating physical model of the vocal folds using several techniques, including grid-based hotwire anemometry, high-speed flow visualization, and particle image velocimetry. Coherent flow structures were extracted from the spatio-temporal data using principal component analysis (PCA). The glottal jet resembled planar jets through static slitlike orifices. The direction of the jet core axis oscillated in the medial-lateral direction. Also, there was a positive correlation between the dynamics of shear layers around the laminar jet core and flow structures in the transitional region. This finding was similar to the dynamics of planar jets where jet flapping was observed previously. This can be understood in terms of an antisymmetric array of large-scale vortices being convected downstream. This antisymmetric mode of coherent vortex structures acts as an acoustic dipole that could effect the dynamics of the shear layers in the laminar core region. Thus, our observations help to understand the dynamics of the acoustic near-field of the glottal oscillator and suggest a possible feedback mechanism for source-tract coupling.

4pSC20. Effect of intraoral air pressure on the release of an alveolar stop closure. Lan Chen (Speech Commun. Group, Res. Lab. of Electron., MIT, Cambridge, MA 02139)

The effect of air pressure on the release of an alveolar stop consonant closure is simulated with a 2-D finite-element model of the front part of the midsagittal tongue in the vocal tract. Active movement of the tongue is estimated from x-ray microbeam recordings, and intraoral air pressure is also applied to the surface of the tongue as an external force. Air flow in the vocal tract is then coupled to the tongue movement by representing the upper tongue surface as part of the flow boundary. At release, the upward movement of the tongue tip due to the decrease in air pressure introduces an oscillation at the release trajectory of the supraglottal constriction. This oscillation is realized in a quasiarticulatory speech synthesizer HLsyn and causes the appearance of a plateau in the time course of the output airflow. This hesitation in tongue tip movement results in an increase in the duration of the frication noise by about 10 ms, consistent with acoustic data on the stop burst. The simulation shows how the mechanical properties of the tongue and the intraoral air pressure buildup can influence the duration of the frication noise. [Supported by Grant NIDCD DC00075 from NIH.]

4pSC21. An experimental study of the aeroacoustic sound generated by vortex shedding in static and dynamic models of the glottis. Bogdan R. Kucinschi, Kenneth J. DeWitt, Ronald C. Scherer (Dept. of Commun. Disord., Bowling Green State Univ., 200 Health Ctr., Bowling Green, OH 43402, bkucins@bgnet.bgsu.edu), and Terry T. Ng (Univ. of Toledo, Toledo, OH 43606)

Flow-visualization with smoke particles illuminated by a laser sheet, performed with a high-speed camera, was used to obtain the flow patterns in two models of the human larynx, both scaled by a factor of 7.5. The first model was static, while the second was a driven dynamic mechanical model, which simulated the motion of the vocal folds during phonation. For the dynamic model, the glottal motion frequency was of the order of the average fundamental frequency of males. It was found that the vortical flow structures were shed at different locations within and downstream the glottis, depending on the transglottal pressure, geometrical configuration and glottal motion frequency. The vortex shedding frequency, which was much higher than the motion frequency of the vocal folds, was measured optically based on the high-speed visualization images, and compared to the frequency peaks in the sound spectra, determined by microphone measurements. Generally, it was observed that the vortex sheddeing frequency corresponds to the main peak in the sound spectrum. The intensity of noise, however, was found to be dependnet of the location of vortical

structures relative to the glottal walls, indicating the presence of quadrupole- or dipole-type sound sources, due to ordered flow structures.

4pSC22. Identification of final fall in subglottal pressure contours of speech utterances. Helen M. Hanson, Janet Slifka, Stefanie Shattuck-Hufnagel (Speech Commun. Group, MIT RLE, 36-511, 77 Massachusetts Ave., Cambridge, MA 02139, helen.hanson@alum.mit.edu), and James B. Kobler (Ctr. for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, MA 02114)

Subglottal pressure (P_s) contours for speech are described as having three phases: initial rise, constant or declining working phase, and final fall. The current work is part of a project to relate characteristics of the P_s contour to prosodic events. To that end, one must identify the three phases in a P_s contour. In past work, it was found that the initial phase is relatively easy to identify, but the transition from the working phase to final fall is less clear [J. Slifka (2000)]. Confounding issues could include segmental impedance, pitch accents, and phrase and boundary tones, all of which can have local effects on P_s. In this work, it is attempted to control tones and segments at the ends of utterances in order to better identify final fall. Lung pressure is estimated from esophageal pressure (corrected for lung volume). Pilot data from one subject indicate that the beginning of final fall is easier to identify when the phrase and boundary tones are low than when they are high. Results will be presented for additional subjects and it will be attempted to relate them to the distribution of pitch accents. [NIH grant DC04331.]

4pSC23. The articulatory and acoustical characteristics of the "apical vowels" in Beijing Mandarin. Wai-Sum Lee (Dept. of Linguist., Univ. of Hong Kong, Pokfulam Rd., Hong Kong, Hong Kong, wsleeba@hku.hk)

The study investigates the articulatory and acoustical characteristics of the two so-called "apical vowels" in Beijing Mandarin, which have been referred to as "apical anterior vowel" and "apical posterior vowel" by the linguists in China. The "apical posterior vowel" has also been described as a retroflex. The results of an EMA (electromagnetic articulograph) analysis show that both vowels are apical, with the tip of tongue approaching the alveolar region for the "anterior vowel" and the postalveolar region for the "posterior vowel." The "posterior vowel" is pharyngealized, as the body of tongue in particular the posterodorsal portion is pulled backward toward the pharynx. Acoustical data obtained using the CSL4400 speech analysis software show that the two "apical vowels" have similar $F1$ value. The $F2$ value is slightly larger for the "posterior vowel" than "anterior vowel." Thus, the correlation between a larger $F2$ and the advanced tongue position is not applicable to these "apical vowels." The main difference between the two "apical vowels" is in $F3$, where the value is much smaller for the "posterior vowel" than "anterior vowel." It is assumed that the smaller $F3$ value for the "posterior vowel" is due to pharyngealization.

4pSC24. Vocal tract length development during the first two decades of life: A magnetic resonance imaging study. Houri K. Vorperian (Waisman Ctr., Univ. of Wisconsin-Madison, Madison WI 53705), Moo K. Chung (Univ. of Wisconsin-Madison, Madison, WI 53705), Lindell R. Gentry (Univ. of Wisconsin Hospitals and Clinics, Madison, WI 53792), Ray D. Kent, Celia S. Choih, Reid B. Durtschi and Andrew J. Ziegert (Univ. of Wisconsin-Madison, Madison, WI 53705)

As the vocal tract length (VTL) increases more than twofold from infancy to adulthood, its geometric proportions change. This study assesses the developmental changes of the various hard and soft tissue structures in the vicinity of the vocal tract (VT), and evaluates the relational growth of the various structures with VTL. Magnetic resonance images from 327 cases, ages birth to age 20, were used to secure quantitative

measurements of the various soft, cartilaginous and bony structures in the oral and pharyngeal regions using established procedures [Vorperian *et al.* (1999), (2005)]. Structures measured include: lip thickness, hard- and soft-palate length, tongue length, naso-oro-pharyngeal length, mandibular length and depth, and distance of the hyoid bone and larynx from the posterior nasal spine. Findings indicate: (a) ongoing growth of all oral and pharyngeal structures with changes in growth rate as a function of age; (b) a strong interdependency between structure orientation and its growth curve; and (c) developmental changes in the relational growth of the different VT structures with VTL. Findings provide normative data on the anatomic changes of the supra-laryngeal speech apparatus, and can be used to model the development of the VT. [Work supported by NIH-NIDCD Grants R03-DC4362 R01-DC006282, and NIH-NICHHD P30-HK03352.]

4pSC25. Improving automatic speech recognition via better analysis and adaptation. Douglas O'Shaughnessy, Wayne Wang, William Zhu, Vincent Barreaud, T. Nagarajan and R. Muralishankar (INRS-EMT, 800 de la Gauchetiere West, Ste. 6900 Montreal, QC, Canada H5A 1K6)

One way to improve automatic speech recognition (ASR) systems is to reduce the mismatch between system training and operating conditions, as such mismatch seriously degrades performance. We have developed model adaptation techniques able to adapt to various speech environments without modifying ASR systems, and have developed an appropriate feature transformation scheme for the Mel-frequency cepstral coefficients (MFCC), a popular front-end feature of ASR systems. We use maximum a posteriori model adaptation and a method based on Bayesian parametric representation. Feature transformation aims to maximize the desired source of information for a given speech signal in the front-end features and to minimize undesired sources. Frequency-domain autoregressive modeling and a segmentation algorithm are being developed, e.g., to segment a speech signal into syllablelike units. We also introduce a new speech-processing front-end feature that performs better than the existing MFCC, as well as a log-energy dynamic range normalization technique for ASR in adverse conditions. In addition, we have developed a continuous ASR method that exploits the advantages of syllable and phoneme-based subword unit models. [Work supported by NSERC-Canada and Prompt-Quebec.]

4pSC26. The effects of the glottal geometry on intraglottal pressure distributions. Li Sheng (Dept. of Biomed. Eng. School of Life Sci. and Technol., Xi'an Jiaotong Univ., 28 West Xianning Rd., Xi'an, 71004R, P.R. China), C. Ronald Scherer (Bowling Green State Univ., Bowling Green, OH 43403), MingXi Wan, and SuPin Wang (Xi'an Jiaotong Univ., China)

The purpose of this study is to explore the effects of the glottal geometry on intraglottal and transglottal pressures using a Plexiglas model and a commercially computational fluid dynamics code, FLUENT. Nine glottal angles (uniform, as well as convergent and divergent 5, 10, 20, and 40 deg), 18 inferior vocal-fold angles varied from 87.5 to −10 deg, and 19 superior vocal-fold surface angles varied from −85 to 45 deg for uniform, convergen 10- and divergent 10-deg glottal angle, and a wide range of entrance radii varied from 0.26 to 0.005 cm for different divergent glottis were selected separately to examine their pressure distribution effects. The empirical data were supported by computational results using FLUENT. The results suggest that the 10-deg divergence angle may correspond to least flow resistance, the vocal-fold surface pressures are essentially independent of the inferior and superior vocal-fold surface angles realistic for human phonation, and a small glottal entrance radius tends to lower the transglottal pressure, move the minimal pressure near the glottal entrance more upstream, and make the pressure dip more negative in value. These results suggest that the glottal geometry should be well specified when using physical, mathematical phonatory models.

Session 4pSP

Signal Processing in Acoustics, Underwater Acoustics, Animal Bioacoustics, Noise, Acoustical Oceanography and Engineering Acoustics: Blind Deconvolution in Acoustics

George B. Smith, Cochair

Naval Research Lab., Code 7183 Stennis Space Center MS 39529-5004

George E. Ioup, Cochair

Univ. of New Orleans, Dept. of Physics, New Orleans, LA 70148

Juliette Ioup, Cochair

Univ. of New Orleans, Dept. of Physics, New Orleans, LA 70148

Invited Papers

3:15

4pSP1. Blind deconvolution and source separation in acoustics. Leon H. Sibul, Michael J. Roan, and Christian M. Coviello (Appl. Res. Lab., Penn State Univ., P.O. Box 30. State College, PA 16804-0040, lhs2@psu.edu)

Blind deconvolution (BDC) and blind source separation (BSS) are active research topics with many important applications in acoustics. The goal of deconvolution is to recover original input signal from the output of a convolution filter. In blind deconvolution details of the convolution filter and input signals are not known. The fundamental assumption in BDC is that the input signal is a non-Gaussian stochastic process. A topic closely related to BDC is BSS. BSS is a process that is an inverse operation to a mixing process. In BSS it is assumed that inputs to the mixing systems are statistically independent stochastic processes, where only one input may be Gaussian, others must be non-Gaussian. Standard criterion functions for BDC and BSS are reviewed. Limitations of the second-order statistics and need for higher-order statistics (HOS) or information theoretic criteria that lead to nonlinear optimization algorithms are pointed out. Advantages of various information theoretic criteria for BDC and BSS are discussed. Because gradients of these criteria are nonlinear, resulting optimization algorithms are nonlinear. Linear and non-linear algorithms for BDC and BSS are examined. [Work supported by ONR Codes 321US and 333.]

4:00

4pSP2. Blind deconvolution in ocean sound channels via artificial time reversal. David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109-2133), Karim G. Sabra (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238), and Mark R. Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

A signal that travels through a sound channel is typically distorted when recorded by a remote listener because of interference arising from multiple propagation paths. If the distorted signal is recorded with a transducer array, a novel passive acoustic technique, artificial time reversal (ATR), can be used for blind deconvolution to remove sound-channel-induced distortion and recover the original signal. The technique is based on artificial backpropagation of the recorded signals using source-to-array Green's functions synthesized directly from the recorded signals. The technique relies on separating out the contribution of a propagating mode or ray path having a phase that depends linearly on frequency. Analysis of the technique shows that it is simultaneously a matched filter and an inverse filter in the limit of high signal-to-noise ratio. Computational example's employing a single source are presented and compared with results obtained from oceanic measurements. The temporal correlation between an ATR-compressed signal and the original signal approaches 100% in the computational examples, and commonly exceeds 90% for the oceanic data. Potential extension of this technique to source–array-range estimation, and to multiple sources emitting simultaneously in the same frequency band are also presented. [Work supported by ONR.]

4:30

4pSP3. Maximum-length-related sequences and a specialized deconvolution for multiple acoustic channel estimation. Ning Xiang (Architectural Acoust. Program, School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180) and John N. Daigle (Univ. of Mississippi, University, MS 38677)

A wide variety of acoustic systems have multiple sources and receivers. This paper discusses a deconvolution technique for making acoustic measurements simultaneously with multiple sources. The deconvolution is based on a collection of known excitation signals of the maximum-length sequence (MLS) related classes, widely used in the spread spectrum communication, but less well known in acoustics applications. This technique is particularly valuable when characterization of multisource, multireceiver system must be accomplished in a limited time period. Since lengths of periodic MLS-related source signals are not suitable for FFT-based decovolutions, and the fast MLS transform based on fast Hadamard transform is not applicable to the MLS-derived combined signals, our effort has been given to an efficient deconvolution algorithm. The simulation and experimental results demonstrate the feasibility, fidelity, and potential applications of the deconvolution techniques in characterizing acoustic systems.

4pSP4. Mechanisms proposed for spectrogram correlation and transformation deconvolution in FM bat sonar. James A. Simmons (Dept. of Neurosci., Brown Univ., Providence, RI 02912)

Big brown bats use time/frequency distributions to represent FM biosonar pulses and echoes as a consequence of reception through frequency tuned channels of the inner ear and subsequent processing by similarly tuned neural channels in the auditory pathway. Integration time is 350 μs, yet delay resolution is $2-10$ μs, which must be based on detecting changes in the echo spectrum caused by interference between overlapping reflections inside the integration time. However, bats perceive not merely the echo interference spectrum but the numerical value of the delay separation from the spectrum, which requires deconvolution. Because spectrograms are the initial representation, this process is spectrogram correlation and transformation (SCAT). Proposed SCAT deconvolution mechanisms include extraction of echo envelope ripples for time-domain spectrometry, cepstral analysis of echoes, use of coherent or noncoherent reconstruction with basis functions, segmentation of onsets of overlapping replicas at moderate to long time separations, and localization of the occurrence of spectral interference ripples at specific times within dechirped spectrograms. Physiological evidence from single-unit recordings reveals a cepstral-like time–frequency process based on freqlets, both single-unit and multiunit responses reveal which may prove to be time-domain basis functions, and multiunit responses exhibit modulations by onset and envelope ripple. [Work supported by NIH and ONR.]

Contributed Paper

4pSP5. Deconvolution for interference cancellation in underwater acoustics using Gibbs sampling. Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102)

Loud interfering sources often hinder efforts to localize a quiet source in shallow-water environments. Interference cancellation methods are sought that remove the field component from the extraneous sources, leaving the contribution of the source of interest intact as much as possible; this latter contribution can then be used for source localization using matched-field methods. In this work we develop a method that estimates locations and spectral characteristics (amplitude and phase) of multiple sources; the number of sources is estimated as well. The procedure is optimized with Gibbs sampling and is, in essence, a deconvolution process, since source amplitude and phase information is recovered. We find that the parameters of the strongest sources can be reliably estimated within a few iterations. The field components attributed to those interfering sources are then estimated and removed from the total acoustic field; the remaining field component can be readily used for localization of the quiet source(s) using a simple Bartlett processor. [Work supported by ONR.]

4p THU. PM

Session 5aAA

Architectural Acoustics, Engineering Acoustics and ASA Committee on Standards: ISO 3382 Acoustics and Electroacoustics Measurement Workshop

Peter A. Mapp, Cochair

Peter Mapp Associates, 5 Worthington Way, Colchester, Essex CO3, 4JZ, U.K.

Robert C. Coffeen, Cochair

Univ. of Kansas, School of Architecture and Urban Design, Marvin Hall, Lawrence, KS 66045

This workshop will discuss the room acoustics measures defined in ISO 3382 and their application, and it will also investigate how the current generation of software and 'computer with sound card' based instrumentation can be used to make these measurements. The measurement of the potential intelligibility of rooms and electroacoustic systems will also be covered.

After a brief introduction of the various ISO and intelligibility parameters, the workshop will discuss the wide variety of test signals and stimuli that are now available and give guidance as to which type is best suited for each type of measurement.

This will be followed by a brief tutorial on signal and time windows that again will aim to suggest which types of forms of window should be used for a given application. Signal acquisition and sound card/system calibration will then be tackled and practical advice will be given with respect to choosing and using sound cards.

The final section of the workshop will be devoted to a discussion of what the measurements actually mean, and a discussion of other diagnostic tests that can be undertaken to solve rather than just quantify a problem. The workshop will include a number of live measurement and calibration demonstrations.

Contributed Paper

11:45

5aAA1. Alternative methods for assessing the omnidirectionality of sources for room acoustics measurements. Timothy Leishman (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

International standards provide guidelines for qualifying omnidirectional sources for use in room acoustics measurements. This paper presents additional methods that may be used for this purpose and explains their similarities and differences. In some cases the standard methods produce results that are at odds with the alternative methods. The paper explains the discrepancies and provides suggestions for their resolution.

FRIDAY MORNING, 21 OCTOBER 2005   CONRAD D, 8:00 TO 11:55 A.M.

Session 5aMU

Musical Acoustics and Signal Processing in Acoustics: Music Information Retrieval

James W. Beauchamp, Chair

Univ. of Illinois Urbana-Champaign, School of Music and Dept. Electrical Computer Engineering, 2136 Music Bldg., Urbana, IL 61801

Invited Papers

8:00

5aMU1. MPEG-7—Standardized tools for music information retrieval. Jürgen Herre (Audio & Multimedia Depts., Fraunhofer Inst. for Integrated Circiuts (IIS), Am Wolfsmantel 33, 91058 Erlangen, Germany)

Today, many applications in Music Information Retrieval (MIR) employ audio features which have been tailored individually by the algorithm developers. For a broader use also in commercial applications, MIR technology can benefit significantly from a ''common language'' in audio signal description that can be used to annotate any type of multimedia assets in order to facilitate search and retrieval according to a wide range of conceivable criteria in an interoperable way. The audio part of the ISO/MPEG-7 ''Multimedia Content Description Interface'' provides such a common signal description language by defining a rather comprehensive set of standardized features [called ''low level descriptors'' (LLDs)], application-centric subsets, and a unified way of exchanging this data based on XML. The talk provides an overview of the MPEG-7 Audio tool chest, including existing and forthcoming extensions. While the idea is clearly to create a universal platform for any conceivable MIR task, some of the initially conceived applications of MPEG-7 Audio are illustrated.

5aMU2. Distributed digital music archives and libraries. Ichiro Fujinaga (Faculty of Music, McGill Univ., 555 Sherbrooke W., Montreal, Canada)

The main goal of this research program is to develop and evaluate practices, frameworks, and tools for the design and construction of worldwide distributed digital music archives and libraries. Over the last few millennia, humans have amassed an enormous amount of musical information that is scattered around the world. It is becoming abundantly clear that the optimal path for acquisition is to distribute the task of digitizing the wealth of historical and cultural heritage material that exists in analogue formats, which may include books and manuscripts related to music, music scores, photographs, videos, audio tapes, and phonograph records. In order to achieve this goal, libraries, museums, and archives throughout the world, large or small, need well-researched policies, proper guidance, and efficient tools to digitize their collections and to make them available economically. The research conducted within the program addresses unique and imminent challenges posed by the digitization and dissemination of music media. The are four major research projects in progress: development and evaluation of digitization methods for preservation of analogue recordings; optical music recognition using microfilms; design of workflow management system with automatic metadata extraction; and formulation of interlibrary communication strategies.

5aMU3. Music to knowledge: A visual programming environment for the development and evaluation of music information retrieval techniques. Andreas F. Ehmann (Dept. of Elec. and Computer Eng., Univ. of Illinois Urbana-Champaign, 1406 W. Green St., Urbana, IL 61801) and J. Stephen Downie (Grad. School of Library and Information Sci., Univ. of Illinois Urbana-Champaign, Champaign, IL 61820)

The objective of the International Music Information Retrieval Systems Evaluation Laboratory (IMIRSEL) project is the creation of a large, secure corpus of audio and symbolic music data accessible to the music information retrieval (MIR) community for the testing and evaluation of various MIR techniques. As part of the IMIRSEL project, a cross-platform JAVA based visual programming environment called Music to Knowledge (M2K) is being developed for a variety of music information retrieval related tasks. The primary objective of M2K is to supply the MIR community with a toolset that provides the ability to rapidly prototype algorithms, as well as foster the sharing of techniques within the MIR community through the use of a standardized set of tools. Due to the relatively large size of audio data and the computational costs associated with some digital signal processing and machine learning techniques, M2K is also designed to support distributed computing across computing clusters. In addition, facilities to allow the integration of non-JAVA based (e.g., C/C++, MATLAB, etc.) algorithms and programs are provided within M2K. [Work supported by the Andrew W. Mellon Foundation and NSF Grants No. IIS-0340597 and No. IIS-0327371.]

5aMU4. Speech-recognition interfaces for music information retrieval. Masataka Goto (Natl. Inst. of Adv. Industrial Sci. and Technol. (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, m.goto@aist.go.jp)

This paper describes two hands-free music information retrieval (MIR) systems that enable a user to retrieve and play back a musical piece by saying its title or the artist's name. Although various interfaces for MIR have been proposed, speech-recognition interfaces suitable for retrieving musical pieces have not been studied. Our MIR-based jukebox systems employ two different speech-recognition interfaces for MIR, *speech completion* and *speech spotter*, which exploit intentionally controlled nonverbal speech information in original ways. The first is a music retrieval system with the *speech-completion* interface that is suitable for music stores and car-driving situations. When a user only remembers part of the name of a musical piece or an artist and utters only a remembered fragment, the system helps the user recall and enter the name by *completing* the fragment. The second is a background-music playback system with the *speech-spotter* interface that can enrich human-human conversation. When a user is talking to another person, the system allows the user to enter voice commands for music playback control by *spotting* a special voice-command utterance in face-to-face or telephone conversations. Experimental results from use of these systems have demonstrated the effectiveness of the *speech-completion* and *speech-spotter* interfaces. (Video clips: http://staff.aist.go.jp/m.goto/MIR/speech_if.html)

5aMU5. Using music structure to improve beat tracking. Roger B. Dannenberg (Computer Sci. Dept., Carnegie Mellon Univ., Wean Hall 4212, Pittsburgh, PA 15213)

Beats are an important feature of most music. Beats are used in music information retrieval systems for genre classification, similarity search, and segmentation. However, beats can be difficult to identify, especially in music audio. Traditional beat trackers attempt to (1) match predicted beats to observations of likely beats, and (2) maintain a fairly steady tempo. A third criterion can be added: when repetitions of musical passages occur, the beats in the first repetition should align with the beats in all other repetitions. This third criterion improves beat tracking performance significantly. Repetitions of musical passages are discovered in audio data by searching for similar sequences of chroma vectors. Beats are "tracked" by first locating a sequence of likely beats in the music audio using high frequency energy as an indicator of beat likelihood. This beat sequence is then extended by searching forward and backward for more matching beats, allowing slight variations in tempo, and using a relaxation algorithm to optimize the proposed beat locations with respect to the three criteria. Other high-level music features may offer further improvements in beat identification.

5a FRI. AM

10:15

5aMU6. Representation and alignment of sung queries for music information retrieval.   Norman H. Adams and Gregory H. Wakefield   (Dept. of Elec. Eng. and Computer Sci., Univ. of Michigan, 1101 Beal Ave., Ann Arbor, MI 48109-2110)

The pursuit of robust and rapid *query-by-humming* systems, which search melodic databases using sung queries, is a common theme in music information retrieval. The retrieval aspect of this database problem has received considerable attention, whereas the front-end processing of sung queries and the data structure to represent melodies has been based on musical intuition and historical momentum. The present work explores three time series representations for sung queries: a sequence of notes, a ''smooth'' pitch contour, and a sequence of pitch histograms. The performance of the three representations is compared using a collection of naturally sung queries. It is found that the most robust performance is achieved by the representation with highest dimension, the smooth pitch contour, but that this representation presents a formidable computational burden. For all three representations, it is necessary to align the query and target in order to achieve robust performance. The computational cost of the alignment is quadratic, hence it is necessary to keep the dimension small for rapid retrieval. Accordingly, iterative deepening is employed to achieve both robust performance and rapid retrieval. Finally, the conventional iterative framework is expanded to adapt the alignment constraints based on previous iterations, further expediting retrieval without degrading performance.

10:40

5aMU7. Automatic detection of the dominant melody in acoustic musical signals.   Anssi P. Klapuri   (Inst. of Signal Processing, Tampere Univ. of Technol., Korkeakoulunkatu 1, 33720 Tampere, Finland)

An auditory-model based method is described for estimating the fundamental frequency contour of the dominant melody in complex music signals. The core method consists of a conventional cochlear model followed by a novel periodicity analysis mechanism within the subbands. As the output, the method computes the salience (i.e., strength) of different fundamental frequency candidates in successive time frames. The maximum value of this vector in each frame can be used to indicate the dominant fundamental frequency directly. In addition, however, it was noted that the first-order time differential of the salience vector leads to an efficient use of temporal features which improve the performance in the presence of a large number of concurrent sounds. These temporal features include particularly the common amplitude or frequency modulation of the partials of the sound that is used to communicate the melody. A noise-suppression mechanism is described which improves the robustness of estimation in the presence of drums and percussive instruments. In evaluations, a database of complex music signals was used where the melody was manually annotated. Use of the method for music information retrieval and music summarization is discussed.

11:05

5aMU8. Using instrument recognition for melody extraction from polyphonic audio.   Jana Eggink [a)]   (Sony Deutschland GmbH, European Technol. Ctr., Hedelfinger Strasse 61, 70327 Stuttgart, Germany) and Guy J. Brown   (Univ. of Sheffield, Sheffield S1 4DP, UK)

A system is proposed that identifies the solo instrument in accompanied sonatas and concertos, and uses this knowledge to extract the melody line played by this instrument. The approach uses a feature representation based solely on the spectral peaks belonging to the harmonic series of a fundamental frequency ($F0$). Based on an initial approximate $F0$ estimation, this representation proved to be sufficient for instrument classification even in the presence of highly unpredictable background accompaniment. Once the solo instrument is known, a more accurate estimation of the melody line is carried out based on so-called melody models, which are trained on instrument-specific training material. In every time frame multiple $F0$ candidates are extracted and their likelihood is evaluated according to the chosen melody model. Additional temporal constraints take the form of frame-to-frame transition probabilities, and are obtained from the same training material. The two knowledge sources are combined in a statistical search for the overall most likely melody line. When evaluated on realistic recordings of classical sonatas and concertos, the system was able to find the correct $F0$ in 72% of frames, an improvement of over 20% compared to a simple salience based approach. [a)] Previously at University of Sheffield.

11:30

5aMU9. Using string alignment in a query-by-humming system for real world applications.   Christian Sailer   (Fraunhofer IDMT, Langewiesener Str. 22, 98693 Ilmenau, Germany)

Though query by humming (i.e., retrieving music or information about music by singing a characteristic melody) has been a popular research topic during the past decade, few approaches have reached a level of usefulness beyond mere scientific interest. One of the main problems is the inherent contradiction between error tolerance and dicriminative power in conventional melody matching algorithms that rely on a melody contour approach to handle intonation or transcription errors. Adopting the string matching/alignment techniques from bioinformatics to melody sequences allows to directly assess the similarity between two melodies. This method takes an MPEG-7 compliant melody sequence (i.e., a list of note intervals and length ratios) as query and evaluates the steps necessary to transform it into the reference sequence. By introducing a musically founded cost-of-replace function and an adequate post processing, this method yields a measure for melodic similarity. Thus it is possible to construct a query by humming system that can properly discriminate between thousands of melodies and still be sufficiently error tolerant to be used by untrained singers. The robustness has been verified in extensive tests and real world applications.

Session 5aSC

Speech Communication: Laboratory Phonology (Poster Session)

Stefan A. Frisch, Chair

Univ. of South Florida, Communication Sciences and Disorders, 4202 E. Fowler Ave., Tampa, FL 33620-8100

Contributed Papers

All posters will be on display from 8:45 a.m. to 11:45 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:45 to 10:15 and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 11:45 a.m.

5aSC1. The lexical accent effects on the perception of fundamental frequency peaks in Japanese. Takahito Shinya (Dept. of Linguist., 226 South College, Univ. of Massachusetts, Amherst, MA 01003)

A perceptual experiment was conducted to examine the effects of lexical accent on the perceptual prominence of fundamental frequency peaks ($F0$) in Japanese. Listeners judged the relative prominence of two peaks (P1 and P2) in sentences in which the peak values and accentedness were varied. When the accentedness of the two peaks did not differ [accented–accented (*aa*) and unaccented–unaccented (*uu*)], the results confirmed earlier observations that P2 needs to be lower than P1 to be perceived to be as equally prominent. Moreover, P2 in *aa* needs to be even lower than P2 in *uu* to sound equal to P1, due to the perceptual compensation for downstep. However, when accentedness of P1 and P2 differed, to sound equally prominent P2 had to be lower in *au* than in *aa*, and P2 had to be higher in *ua* than in *uu*. These patterns can be accounted for if the perceived height of an accented word is discounted. The experimental results revealed that lexical accent plays an important role in the perception of intonational prominence in Japanese, and provide evidence for hitherto unknown compensation processes involved in it: downstep compensation and accent discounting.

5aSC2. Discrimination of phoneme length differences in word and sentence contexts. Norimune Kawai and Thomas Carrell (Dept. of Special Education and Commun. Disord., Univ. of Nebraska at Lincoln., 206 Barkley Memorial Ctr., Lincoln, NE 68583-0731)

The ability of listeners to discriminate phoneme duration differences within word and sentence contexts was measured. This investigation was part of a series of studies examining the audibility and perceptual importance of speech modifications produced by stuttering intervention techniques. Just noticeable differences (jnd's) of phoneme lengths were measured via the parameter estimation by sequential testing (PEST) task, an adaptive tracking procedure. The target phonemes were digitally manipulated to vary from normal (130 m) to prolonged (210 m) duration in 2-m increments. In the first condition the phonemes were embedded in words. In the second condition the phonemes were embedded within words, which were further embedded in sentences. A four-interval forced-choice (4IAX) task was employed on each trial, and the PEST procedure determined the duration at which each listener correctly detected a difference between the normal duration and the test duration 71% of the time. The results revealed that listeners were able to reliably discriminate approximately 15-m differences in word context and 10-m differences in sentence

context. An independent t-test showed a difference in discriminability between word and sentence contexts to be significant. These results indicate that duration differences were better perceived within a sentence context.

5aSC3. Allophonic reassignment in dialect adaptation. James German, Janet Pierrehumbert (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208-4090, j-german@northwestern.edu), and Katy Carlson (Morehead State Univ., Morehead, KY 40351)

This study explored speakers' success in acquiring an unfamiliar dialect. Twenty-four speakers of American English attempted to learn Glaswegian, a dialect in which /r/ is produced as a tap. Its closest counterpart is the flap allophone of /t/. The question posed is whether an allophone of one consonant can be remapped to serve as the realization of another consonant. Subjects were trained and tested with materials in which /r/ appeared only in the last word. Each block contained 12 initial and 12 medial /r/'s. A training block was presented twice followed by a generalization block. A week after training, subjects returned for retesting on the original training and generalizations blocks, and a new generalization block. The block order was counterbalanced. Subjects were highly successful at imitating medial /r/ (71%), a striking example of fast categorical remapping in production. Success rates were lower (46%) for /r/ in initial position, where flaps do not appear in English. Generalization and retest results were only slightly worse than original training results, indicating that subjects acquired long-term general knowledge about Glaswegian pronunciation. The slight superiority on familiar materials in the retest phase points to an interaction of lexical learning with general phonological learning.

5aSC4. Turn exchange rhythm in English dialogues. Janice Fon (Natl. Taiwan Univ., No.1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan, yjfon@ntu.edu.tw)

This study looked at the relationship between rhythm and exchange type in British English, a stress-timed language, and Singaporean English, a syllable-timed language, using a spontaneous speech corpus. Exchange intervals (EIs), or the time difference between the end of one speaker and the beginning of another, were measured and exchanges of different types were labeled. Results showed that, in a dialogue, EIs were generally limited to a narrow range. However, within this range, EIs had at least four functions. First, EIs were reflective of the cognitive load and functioned as a way to differentiate various exchange types. Those requiring more cognitive resources, such as question-and-answer pairs, generally needed longer EIs than those not as cognitively loaded, such as backchanneling pairs. Second, EIs were indicative of linguistic rhythm. Singaporean English tended to have shorter EIs than British English. Third, EIs were

reflective of politeness. The degree of politeness correlated negatively with EI. Shorter EIs showed a higher degree of respect. Finally, EIs were also indicative of the level of insecurity of a speaker, which was best reflected by gender differences. Females in general had longer EIs than males.

5aSC5. Spectral correlates lexical prosody. Anthony Okobi, Jr. (Speech Commun. Group, RLE-MIT, 50 Vassar St., Rm. 36-549, Cambridge, MA 02139, tokobi@mit.edu)

The purpose of this study is to derive a quantitative acoustic model of lexical-prosodic characteristics of stressed vowels by looking at several spectral properties associated with the articulatory mechanisms used in speech production. Native speakers of American English were asked to name disyllabic visualizable nouns. Words containing liquids or glides were not used in this study because of their effect on the spectra of adjacent vowels. Subjects uttered short phrases in which the target word was pitch accent half of the time and unaccented the other half. Results show that within the category of full vowels, unstressed and stressed vowels can be distinguished by syllable/vowel durations and spectral tilt. Spectral tilt (SpT) is an acoustic measure related to the degree of glottal spreading. Stressed full vowels had longer duration and less SpT. Distinction between unaccented and accented stressed vowels can be made by amplitude of voicing (AV), F0 (pitch), and intensity contour differences. Accented stressed vowels have higher pitch, and greater AV and intensity. These results suggest that there are acoustic correlates to lexical stress that can be used to determine the stressed syllable of a word, regardless of whether or not it is pitch accented. [Work supported by NIH T32-DC00038.]

5aSC6. The time course of speaking rate specificity effects in spoken word recognition. Conor T. McLennan and Paul A. Luce (Dept. of Psych., Univ. at Buffalo, Buffalo, NY 14260, mclennan@buffalo.edu)

Specificity effects in spoken word recognition were previously examined by examining the circumstances under which variability in speaking rate affects participants perception of spoken words. The word recognition and memory literatures are now replete with demonstrations that variability has representational and processing consequences. The research focuses on one of the conditions expected to influence the extent to which variability plays a role in spoken word recognition, namely time course of processing. Based on previous work, it was hypothesized that speaking rate variability would only affect later stages of spoken word recognition. The results confirmed this hypothesis: Specificity effects were only obtained when processing was relatively slow. However, previous stimuli not only differed in speaking rate, but also in articulation style (i.e., casual and careful). Therefore, in the current set of experiments, it was sought to determine whether the same pattern of results would be obtained with stimuli that only differed in speaking rate (i.e., in the absence of articulation style differences). Moreover, to further generalize time course findings, the stimuli were produced by a different speaker than the speaker in the earlier study. The results add to the knowledge of the circumstances under which variability affects the perception of spoken words.

5aSC7. Speaker specificity in speech perception: the importance of what is and is not in the signal. Delphine Dahan (Dept of Psych., Univ. of Pennsylvania, Rm. 302C, 3401 Walnut St., Philadelphia, PA 19104-6228, dahan@psych.upenn.edu) and Rebecca A. Scarborough (Stanford Univ., Stanford, CA 94305-2150)

In some American English dialects, /ae/ before /g/ (but not before /k/) raises to a vowel approaching [E], in effect reducing phonetic overlap between (e.g.) "bag" and "back." Here, participants saw four written words on a computer screen (e.g., "bag," "back," "dog," "dock") and heard a spoken word. Their task was to indicate which word they heard. Participants' eye movements to the written words were recorded. Participants in the "ae-raising" group heard identity-spliced "bag"-like words containing the raised vowel [E]; participants in the "control" group heard cross-spliced "bag"-like words containing standard [ae]. Acoustically identical "back"-like words were subsequently presented to both groups. The ae-raising-group participants identified "back"-like words faster and more accurately, and made fewer fixations to the competitor "bag," than control-group participants did. Thus, exposure to ae-raised realizations of "bag" facilitated the identification of "back" because of the reduced fit between the input and the altered representation of the competing hypothesis "bag." This demonstrates that listeners evaluate the spoken input with respect to what is, but also what is not, in the signal, and that this evaluation involves speaker-specific representations. [Work supported by NSF Human and Social Dynamics 0433567.]

5aSC8. Lexical confusability and nasal coarticulation in French. Rebecca Scarborough (Linguist. Dept., Stanford Univ., Stanford, CA 94305-2150, rscar@stanford.edu)

Previous research has revealed a relationship between lexical confusability and degree of coarticulation [Brown (2001); Scarborough (2004)]. In particular, English speakers produce confusable, or "hard" words with more nasal and vowel-to-vowel coarticulation than less confusable, "easy" ones. Thus, it has been suggested that speakers produce additional coarticulation in order to increase the intelligibility of "hard" words. Here, the relation between nasal coarticulation and lexical confusability is investigated for French, a language in which vowel nasality is phonemically contrastive (at least for a subset of vowels) and might constrain such a lexical effect. Acoustic measures of nasality show that "hard" words (those with low usage frequencies and many frequent, phonologically similar neighbors) exhibit more nasal coarticulation than "easy" ones (those with high frequencies and few, low-frequency neighbors) in French as well. Interestingly, however, the effect emerges only for words containing vowels that can exhibit phonemic oral-nasal contrasts (oral vowels with nasal counterparts). Thus, where the use of nasality in phonological contrast is constrained, coarticulatory nasality is constrained, too. But the existence of phonological contrast does not itself constrain the lexical confusability effect: increased coarticulatory nasality contributes to lexically motivated phonetic enhancement in French, while even more nasality provides the basis for phonemic contrast.

5aSC9. A coupled oscillator model of intergestural timing within syllables. Hosung Nam, Louis Goldstein (Haskins Labs., 300 George St., New Haven, CT 06511 and Yale Univ., New Haven, CT 06511), and Elliot Saltzman (Haskins Labs., New Haven, CT 06511 and Boston Univ., Boston, MA 02215)

A computational model has been developed that plans the relative timing of speech gestures within a syllable. *Planning oscillators* are associated with each of the gestures in a given utterance. These oscillators are phased to one another by being coupled in a pairwise, bidirectional manner according to the structure of a *coupling graph* that is part of the lexical specification of a word. A set of motion equations for the coupled oscillator system is implemented using the task-dynamical model of relative phase control [E. Saltzman and D. Byrd, Human Movement Sci. **19**, 499–526 (2000)] extended to allow multiple, potentially competitive, couplings. In the model, consonant gestures in a syllable onset are all coupled synchronously with the vowel (and in some cases sequentially with each other), while coda consonant gestures are coupled sequentially with the vowel. The steady-state output of the coupled oscillator system is a set of limit-cycle oscillations with stabilized relative phases, from which the activation trajectories of gestural scores are derived. The gestural scores are used to generate coordinated articulatory movements and constriction trajectories via the task dynamics model, and acoustic output via a vocal-tract model and HLsyn. [Work supported by NIH.]

5aSC10. Reduction of flaps: Speech style, phonological environment, and variability. Natasha L. Warner (Dept. of Linguist., Univ. of Arizona, P.O. Box 210028, Tucson, AZ 85721-0028, nwarner@u.arizona.edu)

Normal connected speech contains a tremendous amount of variability in how a given sound is realized. The current work examines American English /t/ and /d/ in flapping environments. Flaps can be reduced to an approximant or even deleted in connected speech, and this study compares reduction of flaps to reduction of other stops (/p, b, k, g/) in the same environments. The effects of segmental environment (target followed by /l/, syllabic /r/, /∂/, word boundary, etc.), stress (target following stressed syllable versus between two unstressed syllables), and speech style (casual conversation, reading connected text, and word list reading) are analyzed. Target segments are analyzed for presence/absence of a burst and of formants during closure, for duration of the segment and of closure voicing, and for amplitude during the closure or constriction. This project documents the degree of variability in reduction both within and across speakers. Preliminary results show that flaps are reduced to approximants surprisingly often, even in the most careful speech style. The results also suggest that flaps are far more likely to reduce than /p/ or /k/, suggesting that there is a categorical, phonological process at work as well as gradient variability.

5aSC11. Syllable and phrase structure effects on consonant sequence timing. Dani Byrd (Haskins Labs. & USC Linguist., 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693) and Susie Choi (USC Linguist., Los Angeles, CA 90089-1693)

Both syllable and phrasal structure are known to influence articulatory timing in consonant sequences. For example, onset clusters have been reported as less overlapped and more stable in their intergestural timing than coda clusters [e.g., Byrd, J. Phonetics (1996)]. Also, consonants spanning a phrasal boundary have been observed to be less overlapped than those spanning only a word boundary [Byrd *et al.*, LabPhon (2003)]. However, interactions between these two types of structure are less well understood; for example, it is unclear whether the intergestural timing of word-onset clusters will be perturbed at phrase boundaries, though such perturbations have been predicted [Byrd and Saltzman (2003)]. An articulatory (EMA) investigation of /s/+stop sequences produced by three speakers in a variety of syllable and phrasal positions will present kinematic data on these structural influences. Preliminary data from one speaker indicate that word-onset consonant clusters are more sensitive to prosodic context than segmentally identical coda clusters, having less overlap at successively larger boundaries. Further, while coda and onset clusters do not show a mean difference in overlap for this speaker, onset clusters do exhibit more timing stability within each phrasal context compared to corresponding coda clusters. [Work supported by NIH.]

5aSC12. Syllabification effects on the acoustic structure of intervocalic /r/. Marie Huffman (Dept. of Linguist., Stony Brook Univ., Stony Brook, NY 11794-4376)

Imaging and modeling studies suggest that American English /r/ has a complex articulatory profile. Gick [Phonology **16**, 29–54(1999)] has proposed that dialectal differences in the presence of /r/ follow from the effects of syllable structure and prosody on component vocalic and consonantal gestures of /r/. This study presents acoustic data on word-medial, intervocalic /r/'s for speakers of two varieties of American English. Both varieties show an effect of /r/ on $F3$ and/or $F4$ of a preceding vowel. Where they differ is the acoustic properties of the constriction portion of intervocalic /r/. For one group, the intervocalic /r/ is very vocalic, with little difference in formant amplitude compared to the preceding vowel. For the other group, intervocalic /r/ is more consonantal, with clearly weaker formant structure than the preceding vowel. These differences in the acoustic profile of intervocalic /r/ co-vary with dialectal differences in production of final coda /r/. These results support a gestural account of /r/

variability, while also demonstrating the need for explicit principles of syllable organization which must be specified for each dialect. [Work supported by NSF Grant No. 0325188.]

5aSC13. Unsupervised segmentation of broad classes from waveforms: Towards a model of early phonological acquisition. Ying Lin (Dept. of Linguist., UCLA 3125 Campbell Hall, Los Angeles, CA 90095)

It is commonly held that an important aspect of early phonological acquisition is the ability to learn sound distributions, or statistical learning. Yet significant differences in lexical representations are often observed in studies of infant speech perception, suggesting a protracted process of phonological development. The goal of the current project is to develop a model that links holistic and segmental representation of spoken words, using tools from contemporary speech recognition. In the present stage, the model focuses on the pre-lexical level of phonological development, and tries to identify segmental representations from acoustic signals of isolated words. The segmental representations are based on units that correspond to acoustic phonetic classes, and learning involves updating the unit models in parallel with updating phonotactics. Starting from acoustic segmentations, the model iteratively updates knowledge of the units and phonotactics, and renews segmentation hypotheses regarding each word until convergence. The results of running this algorithm on TIMIT and infant-directed speech data suggest that the model approximately identifies segment-sized broad classes in an unsupervised manner. This statistical approach also provides a different perspective on the role of lexicon in phonological development.

5aSC14. Talker sex mediates the influence of neighborhood density on vowel articulation. Benjamin Munson (Dept. Speech Lang. Hearing Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, Munso005@umn.edu)

Words with high phonological neighborhood densities (ND) are more difficult to perceive than words with low NDs [P. Luce and D. Pisoni (1998)]. Previous research has shown that the $F1/F2$ acoustic vowel space is larger for vowels in words with high ND relative to words with low NDs [B. Munson and N.P. Solomon (2004); R. Wright (2004)]. This may represent talkers' tacit attempts to partially counter the perception difficulties associated with high-ND words. If so, then we would expect to see a larger effect of ND on vowel articulation in women, who have been observed to produce more intelligible speech than men, and to accommodate more to conversational partners than men [V. Hazan and D. Markham (2004); L. Namy *et al.* (2002)]. This hypothesis was tested by examining the influence of ND on vowel-space articulation by 22 women and 22 men. As expected, women produced overall more-dispersed vowel spaces than men, and vowel spaces associated with high-ND words were more-dispersed than low-ND words. Contrary to expectations, the influence of ND on vowel-space expansion was strongest in men. This appeared to be due a tendency for women not to produce contracted vowel spaces for low-ND words.

5aSC15. Phonological neighborhood density and vowel production in older and younger adults. Peter J. Watson, Benjamin Munson, and Erik H. Belgum (Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, pjwatson@umn.edu)

Recent work has shown that neighborhood density (ND) affects vowel production in young adults (B. Munson and N. Solomon, J. Speech. Lang. Hear. Res. **47**, 1048–1058 [2004]). Vowels in words with high ND were produced with larger $F1/F2$ acoustic vowel space; this effect was independent of word frequency and duration. This may reflect speakers' modification of vowel production to assist listeners' vowel perception. However, a *post hoc* analysis of Munson and Solomon's data found that the effect is strongest in the low-front vowel /æ/, suggesting that the effect

may be due to the ongoing sound change in which /æ/ is progressively becoming lower and more central (Hillenbrand *et al.*, J. Acoust. Soc. Am. **97**, 3099–3111 [1995]) occurring more often in high-density words. If so, one should see a larger effect of neighborhood density in younger adults than in older adults. To examine this, vowel production was measured in a group of ten older adults (mean age=76.1 years), and compared to ten younger adults from Munson and Solomon [2004]. The older adults produced vowels with lower-frequency formants than younger adults. Contrary to expectations, age-group differences were not disproportionately larger for /æ /than for other vowels, and both groups showed a robust effect of neighborhood density on vowel production.

5aSC16. Perceptual learning of a talker resolves lexical ambiguity.
Rebecca L. Piorkowski and William Badecker (Johns Hopkins Univ., 237 Krieger Hall, 3400 N. Charles St., Baltimore, MD 21218, rebecca@cogsci.jhu.edu)

Recent evidence [Allen and Miller, J. Acoust. Soc. Am. **115**, 3171 (2004)] suggests that listeners are sensitive to talker-specific acoustic-phonetic properties. The present study examines whether experience with a particular talker's realization of place of articulation can eliminate the ambiguity that arises when a word like ''hen'' assimilates in place of articulation to ''hem'' in a labial context (hen best). Using a cross-modal priming paradigm, the priming effect of words that assimilate to other words was measured in two conditions. In the first condition, listeners heard examples of the talker's assimilation style in the form of words that assimilate to nonwords in a labial context (green beer) before the critical trials were heard. In the second condition, listeners did not hear any examples of the talker's assimilation style before the critical trials were heard. Evidence will be presented showing that, without previous experience with the talker's assimilation style, words that assimilate to other words create a lexical ambiguity for the listener. Additionally, evidence will be presented showing that experience with the talker's assimilation style eliminates this lexical ambiguity and allows the listener to perceive the talker's intended lexical form.

5aSC17. Effects of rhythmic precursors on perception of stress/ syllabicity.
Christian E. Stilp and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53706)

Rhythmic structure is a common property of many environmental sounds including speech. Here, perceptual effects of preceding rhythmic context are assessed in experiments employing edited words for which perceived stress/syllabicity are assessed. A series of edited naturally spoken words varying perceptually from ''polite'' to ''plight,'' was created by deleting initial-vowel glottal pulses from a recording of ''polite.'' Words were identified following nonspeech precursor sequences having either trochaic (strong–weak) or iambic (weak–strong) rhythmic patterns. Precursors consisted of a harmonic spectrum (−6-dB/octave slope) filtered by four sinusoidally modulated single-pole filters. Trochaic (strong–weak) and iambic (weak–strong) rhythmic patterns were created by varying amplitude, pitch, and duration in successive segments (akin to beats) of the precursors. Precursors were comprised of two to six repetitions of these patterns. Following trochaic precursors, listeners were more likely to report hearing ''polite'' (iambic). This pattern of results indicates that perception did not assimilate to precursor pattern, consistent with rhythmic expectancy. Instead, perception shifted in a way that contrasts with precursor temporal pattern. Additional results with precursors that are more and less like speech are being conducted to further understand how auditory perception adjusts for temporal and spectral regularities. [Work supported by NIDCD.]

5aSC18. Perceptual effects of dialectal and prosodic variation in vowels.
Robert Allen Fox, Ewa Jacewicz, Kristin Hatcher (Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210), and Joseph Salmons (Univ. of Wisconsin, Madison, WI 53706)

As was reported earlier [Fox *et al.*, J. Acoust. Soc. Am. **114**, 2396 (2003)], certain vowels in the Ohio and Wisconsin dialects of American English are shifting in different directions. In addition, we have found that the spectral characteristics of these vowels (e.g., duration and formant frequencies) changed systematically under varying degrees of prosodic prominence, with somewhat different changes occurring within each dialect. The question addressed in the current study is whether naive listeners from these two dialects are sensitive to both the dialect variations and to the prosodically induced spectral differences. Listeners from Ohio and Wisconsin listened to the stimulus tokens [beɪt] and [bɛt] produced in each of three prosodic contexts (representing three different levels of prominence). These words were produced by speakers from Ohio or from Wisconsin (none of the listeners were also speakers). Listeners identified the stimulus tokens in terms of vowel quality and indicated whether it was a good, fair, or poor exemplar of that phonetic category. Results showed that both phonetic quality decisions and goodness ratings were systematically and significantly affected by speaker dialect, listener dialect, and prosodic context. Implications of source and nature of ongoing vowel changes in these two dialects will be discussed. [Work partially supported by NIDCD R03 DC005560-01.]

5aSC19. Lexical frequency effects and phonetic duration of English homophones: An acoustic study.
Abigail C. Cohn, Johanna Brugman, Clifford Crawford, and Andrew Joseph (Dept. of Linguist., Cornell Univ., Morrill Hall, Ithaca, NY 14853-4701)

Some current views of phonology assume a single abstract representation for each lexical item, while others assume extensive encoding of fine-grained detail. Some proponents of the latter view have claimed that differences in lexical (token) frequency are manifested as differences in phonetic duration. This claim was investigated in three experiments measuring the phonetic durations of heterographic pairs of homophonous English nouns differing in token frequency. Homophonous pairs were grouped according to magnitude of frequency difference within pairs: large difference (time ∼ thyme), medium difference (pain ∼ pane), and no difference (son ∼ sun). Four participants read (a) words in a list in a frame sentence; (b) target items in composed sentences; and (c) pairs in contrast. No systematic differences of ratio of duration (more frequent/less frequent) were found for individual speakers or across speakers in (a) or (b). Preliminary results for (c) show differences in duration correlated with contrastive focus and final lengthening, but not lexical frequency. The lack of positive correlation between duration and frequency calls into question the hypothesis that greater frequency leads to shorter duration, and underlines the need for a better understanding of the locus of frequency effects in the lexicon and speech production.

5aSC20. Evaluating gestural overlap as an influence on perception of fricative-stop clusters.
Susie Choi and Dani Byrd (USC Linguist., 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693)

This study investigates the effect of intergestural timing and prosodic context on perception. The articulation and perception of word-edge /sk/ and /sp/ clusters in words preceding or following various prosodic boundaries are considered. Excised acoustic recordings of these words provided stimuli for a perception experiment in which listeners did a phoneme-monitoring task, listening for [s]. Using magnetometer data, articulatory overlap of the two consonants was measured. Listener reaction times were evaluated to determine whether the overlap measure and/or the prosodic category in which the cluster occurred had an influence on the speed of response. This presentation will focus on the onset clusters. Results show: (1) reaction times for /s/-monitoring differed according to phrasal position, such that cluster stimuli following larger prosodic boundaries have longer

RTs; (2) responses are significantly faster when clusters have greater articulatory overlap; and (3) intergestural overlap is more important in capturing response variability than phrasal category. Similar studies examining stop clusters [Byrd, Phonetica (1992); Goldstein and Suprenant, J. Acoust. Soc. Am. **104**, 518 (1998)] have found that overlap among stops can cause perceived deletion; the study suggests listeners utilize the information encoded by increased overlap when the overlapped segments aerodynamics allows preservation, rather than loss, of information. [Work supported by NIH.]

5aSC21. The interaction between specificity and linguistic contrast. Kuniko Nielsen (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543)

Previous studies have shown listeners' ability to remember fine phonetic details [e.g., Mullennix *et al.*, 1989], providing support for the episodic view of speech perception. The imitation paradigm [Goldinger, 1998, Shockley *et al.*, 2004], in which subjects' speech is compared before and after they are exposed to target speech (= study phase) has shown that subjects shift their production in the direction of the target. Our earlier results [Nielsen, 2005] showed that the imitation effect for extended VOT was generalized to new stimuli as well as to a new segment, suggesting that the locus of the imitation effect can be smaller than individual words or segments. The current study aims to further investigate how experienced speech input interacts with linguistic representations, by testing whether the imitation effect is observed when the modeled stimuli have reduced VOT (which could introduce linguistic ambiguity). In other words, do speakers imitate and generalize shorter VOT even if the change might impair linguistic contrasts? To address this question, the study phase includes words with initial /p/ with reduced VOT, while the pre- and post-study production list includes (1) the modeled words, (2) the modeled segments /p/ in new words, and (3) the new segment /k/.

5aSC22. Monophthongs and formant movement in North American English. Robert Hagiwara (Linguist. Dept., Univ. of Manitoba, Winnipeg, MB R3T 5V5, Canada, robh@cc.umanitoba.ca)

This poster describes dynamism in vowel production using data derived from two experimental studies, one in Southern California, the other in southern Manitoba. In both studies, young adults read from a script containing multiple repetitions of target vowels in /hVd/ context. The frequencies of the first four formants were measured at three time points (25%, 50%, 75% of vowel duration) in each vowel. Analyses were conducted in both linear (hertz) and auditorily transformed (bark distance) terms. Many dialects of English clearly have diphthongal production of the tense vowels, especially /e, o/, which would be indicated by movement of the formants toward the periphery of the vowel space. For the two populations studied here, none of the tense vowels /i, e, o, u/ show net movement toward the periphery. While there are individual and group differences, simplex vowels in both dialects show very little movement between time points one and two, and only non-targeted, transitional movement between time points two and three. This suggests monophthongal basic representations, though not "flat" or steady productions, for these two dialects. Implications for North American dialectology and for dynamic vowel models will be considered.

5aSC23. Acoustic cues to vowel-schwa sequences for high front vowels. Janet Slifka (Speech Commun. Group, RLE, MIT, 77 Massachusetts Ave., Cambridge, MA 02139, slifka@speech.mit.edu)

In a landmark-based model of lexical access [K. N. Stevens, J. Acoust. Soc. Am. **111**, 1872–1891 (2002)], the presence of a vowel is marked by a peak in energy in the first formant region. However, when a vowel is followed by a schwa, the schwa frequently appears as a shoulder on the peak associated with the first vowel [W. Howitt, MIT (2000)]; two landmarks are not present. The purpose of this study is to examine duration and $F2$ movement as possible cues to the presence of a vowel-schwa sequence for [+high, +front] vowels. This subset of vowels presents at least two challenges to the detection of a vowel-schwa sequence: (1) duration is expected to contribute to the difference between /i/ and /ɪ/, and (2) an $F2$ off-glide toward schwa is expected for /ɪ/ in American English. For 613 tokens from the phonetically labeled TIMIT database, equally distributed between /iə/, /i/, and /ɪ/, a measure of $F2$ curvature is a stronger cue than duration in classifying the tokens. Using $F2$ curvature, over 93% of the tokens are correctly classified. Using duration, 67% are correctly classified. $F2$ curvature may reflect a more extreme articulation for the /i/ prior to the schwa. [Work supported by NIH Grant DC02978.]

5aSC24. A novel procedure for examining pre-lexical phonetic-level analysis. James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz (Dept. of Psych., Univ. of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201-0413)

A recorded word repeated over and over is heard to undergo a series of illusory changes (verbal transformations) to other syllables and words in the listener's lexicon. When a second image of the same repeating word is added through dichotic presentation (with an interaural delay preventing fusion), the two distinct lateralized images of the word undergo independent illusory transformations at the same rate observed for a single image [Lenz *et al.*, J. Acoust. Soc. Am. **107**, 2857 (2000)]. However, when the contralateral word differs by even one phoneme, transformation rate decreases dramatically [Bashford *et al.*, J. Acoust. Soc. Am. **110**, 2658 (2001)]. This suppression of transformations did not occur when a non-speech competitor was employed. The present study found that dichotic suppression of transformation rate also is independent of the top-down influence of a verbal competitor's word frequency, neighborhood density, and lexicality. However, suppression did increase with the extent of feature mismatch at a given phoneme position (e.g., transformations for "dark" were suppressed more by contralateral "hark" than by "bark"). These and additional findings indicate that dichotic verbal transformations can provide experimental access to a pre-lexical phonetic analysis normally obscured by subsequent processing. [Work supported by NIH.]

5aSC25. Rules for pluralization in African American English: Evidence from 4-, 5-, and 6-year-olds. Sonja A. Trent-Brown (Psych. Dept., Hope College, 35 East 12th St., Holland, MI 49423)

This study investigated the African American English (AAE) forms produced by 4-, 5-, and 6-year-old children when pluralizing words ending in final consonant clusters. Participants were 105 child and 45 adult speakers of AAE. Speakers provided verbal responses to pictures of cartoon-like creatures with nonsense word names. Responses were recorded and presented to experienced listeners for language set and singular versus plural judgments. An acoustical analysis was conducted for comparison to the perceptual findings. Results suggest that although [s] inflection is non-obligatory in AAE, there are many instances in which it does occur. Variation in plural forms across final consonant clusters is discussed as well as other markers for pluralization in the absence of the [s] morpheme.

Session 5aSP

Signal Processing in Acoustics and Underwater Acoustics: Algorithms for System Characterization: Vector Fields, Acoustic Communications and Time Reversal

David M. Fromm, Chair

Naval Research Laboratory, Code 7142, 4555 Overlook Ave., SW, Washington, DC 20375-5350

Contributed Papers

9:00

5aSP1. Estimation of acoustic vector sensor correlations in arbitrary noise field for optimal array processing. Robert J. Minniti III (Arion Systems, Inc., 15059 Conference Center Dr., Ste. 200, Chantilly, VA 20151, rminniti@arionsys.com)

Optimal array processing relies heavily upon estimates of the noise field cross-correlation at the sensor locations to define sensor weights and establish an image of the field. For vector sensor arrays, the estimation process is complicated by the need to establish the correlations between the four sensors measuring the scalar pressure and three components of the acoustic particle velocity for a single location sensor and spatially displaced sensor pairs. Estimation of the noise-field correlation is typically based upon direct measurements at the sensor locations. However, it is often the case in practice that accurate direct estimation is not possible due to the limited availability of data at the analysis frequency of interest. In this paper, the author will obtain closed-form expressions for auto- and cross correlations between the components of acoustic vector sensors for within-sensor and displaced-sensor pairs in a wideband noise field of arbitrary spatial distribution. These results can be used to estimate the covariance matrix of an arbitrarily shaped array for use in a variety of optimal array processing methods. A numerical simulation applying the estimation process in a recursive implementation of minimum variance distortionless response (MVDR) will be provided as an example.

9:15

5aSP2. Acoustic vector fields in underwater waveguides. Brian Rapids (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723)

The ability to compute the sound pressure level as well as the vectors associated with the acoustic particle motion has existed for some time. However, propagation studies and ambient noise investigations have typically focused only upon the sound pressure levels that would be observed by an omnidirectional hydrophone or array of hydrophones. Recent interest in geophones and accelerometers for use as vector and dyadic sensors should encourage the investigation and analysis of the underlying vector fields contributing to the acoustic intensity and energy density fields. The frequency domain properties of the acoustic vector field generated by monopole sources having frequencies $<$1kHz in a simple iso-velocity waveguide are presented in order to build a fundamental understanding of the related quantities. Subsequently, similar field quantities computed for more realistic environments such as downward refracting profiles and deep-water profiles supporting convergence zone propagation will be discussed. Regions and phenomena associated with perturbations in the energy flux density will be highlighted.

9:30

5aSP3. Effects of sea surface on underwater communication. Jing Luo and Mohsen Badiey (College of Marine Study, Univ. of Delaware, Robinson Hall 114, Newark, DE 19711, luojing@udel.edu)

During High Frequency Experiment in 2003 (HFX-03) various ocean parameters such as the surface wave spectrum, salinity, and temperature profiles were recorded simultaneously with the broadband acoustic signals. Including broadband probe chirps as well as several other communi-

cation sequences. Received data are collected and analyzed from different receiver arrays. The statistics of arrival rays and the performances of communication system are studied in two periods of calm and rough sea conditions. Surface reflected rays show variable coherence during different sea surface conditions. This has large influence on the structure of the received signals. Strong dependence between sea surface condition and the performance of communication system is observed. Two types of receiver structures are used and their performances are compared: (1) single channel receiver with decision feedback equalizer, (2) multichannel receiver with signal combining scheme. Similar trends of dependence are found on both receivers. [Work supported by ONR-321OA.]

9:45

5aSP4. Performance of a space-time trellis-coded multiple-input, multiple-output (MIMO) underwater acoustic communication system for the littoral. Richard F. Ormondroyd and Jasdeep S. Dhanoa (The Royal Military College of Sci. Shrivenham, Cranfield Univ., Swindon, SN6 8LA, UK)

Conventional point-to-point communication systems generally use a single transmitter communicating to a single receiver. Recently, there has been considerable interest in exploiting the spatial diversity that exists in richly scattered communication channels using space-time methods that use multiple transmitters and receivers. These multiple-input, multiple-output (MIMO) systems differ from traditional spatial diversity systems because the data or coded symbols from each cochannel transmitter element are different, whereas for spatial diversity they are the same. MIMO systems can offer improved data rates and/or bit error performance, depending upon the system, when the multipath propagation is highly decorrelated. In this paper, a geometric ray model of a horizontal underwater acoustic channel in the littoral, which includes the effect of time-delay spread and Doppler spread, is used to obtain the time-varying, frequency-selective channel parameters that are applied to a discrete signal-processing model of a space-time trellis-coded MIMO system. Key parameters of this model are the separations of the multiple transducers at both the transmitting and receiving ends of the link, because these determine the degree of decorrelation in the multipath channel. The paper examines the performance trade-offs that are possible using this approach with realistic transducer separations at the transmitter and/or receiver.

10:00

5aSP5. Preliminary investigation of acoustic bar codes for short-range underwater communications. Dennis F. Jones (Defence R&D Canada - Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, dennis.jones@drdc-rddc.gc.ca)

In March 2005, underwater acoustic communications experiments were carried out from the DRDC Atlantic research vessel CFAV QUEST. A battery-operated BATS20 transmitter and a broadband barrel-stave flextensional transducer were used to broadcast noise containing acoustic bar code (ABC) information. The ABCs are silent frequency bands of fixed duration that resemble retail bar codes when viewed in a spectrogram. Two sites were selected for the experiments. The first was a shallow-water area west of the Berry Islands in the Bahamas, and the second was a

2038   J. Acoust. Soc. Am., Vol. 118, No. 3, Pt. 2, September 2005

Joint 150th Meeting/NOISE-CON 2005   **2038**

deep-water site south of the Western Bank on the Scotian Shelf. Two receiver systems were deployed; autonomous, variable-buoyancy Stealth Buoys resting on the bottom at the shallow site, and drifting AN/SSQ-53F sonobuoys fitted with GPS at the deep site. Results from these experiments will be presented and future work will be discussed.

10:15

5aSP6. Sparse channel estimation for underwater acoustic communication channels. Aijun Song and Mohsen Badiey (College of Marine Studies, Univ. of Delaware, Newark, DE 19711)

In underwater communication channels, acoustic arrivals often form a cluster structure. The number of significant taps is limited although the length of the channel can be very large. As shown in the previously reported research, channel estimation can benefit the performance of the coherent communication and promise to support high data rate and reliable underwater acoustic communication. Since underwater communication channels are sparse in time domain, channel-estimation algorithms can have lower complexity and higher accuracy by exploiting the sparse property. Various sparse channel estimation techniques, including threshold recursive least-squares (RLS), matching-pursuit, and on-off keying detection algorithms, are investigated in underwater communication channels. Their effectiveness and complexity are compared using communication data measured from the Kauai experiment, July 2003. Their impacts on the communication performance are also presented.

10:30

5aSP7. Wideband multichannel time-reversal communications in a tunnel-like structure. James Candy, Christopher Robbins, Brian Guidry, David Chambers (Lawrence Livermore Natl. Lab, P.O. Box 808, L-156, Livermore, CA 94550), Andrew Poggio, and Farid Dowla (Lawrence Livermore Natl. Lab, Livermore, CA 94550)

The development of multichannel time-reversal (T/R) processing continues to progress rapidly, especially when the need to communicate in a highly reverberant environment is critical. One such environment is a tunnel or cave, or even a pipe with many obstructions, multipath returns, severe background noise, disturbances, path disruptions (bends) as well as a long propagation path (~120 ft.). For this environment, multichannel T/R receiver designs have been extended to include a wideband processor and modulation scheme along with designs to communicate in the highly reverberative tunnel-like environment that includes high background noise levels and disturbances. The acoustic information signals are transmitted by an 8-element host or base station array and received some distance away with a significant loss in transmitted signal levels. In this paper the results of the new wideband processor and modulation scheme coupled with the underlying T/R theory are discussed to demonstrate the overall performance for both high and low bit-level designs in the tunnel.

10:45

5aSP8. Resolution issues of time-reversal operator decomposition (DORT) in beam space. David M. Fromm and Charles F. Gaumond (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

DORT has been previously demonstrated to work using beam-space source excitations and beam-space receiver using a vertical line array in shallow water. The backpropagation images of those results are shown to display effects that arise from incomplete ensonifications of the echo re-

peater, which was at relatively close range of approximately 20 water depths. The small improvement of the backpropagation images from the use of different receiver beam spaces is shown. Greater improvements are shown when more appropriate source beams are also used. These beam spaces include narrower ranges of beams as well as combinations of plane-wave beams. These results are shown to be consistent with the theoretical space of beams related to the detection volume. [Work supported by the ONR.]

11:00

5aSP9. Eigenvalues of the time-reversal operator for a small ellipsoid. David Chambers (Lawrence Livermore Natl. Lab., P.O. Box 808, L-154, Livermore, CA 94551)

Earlier analysis of the eigenvalues of the time-reversal operator for a small sphere showed there could be up to four eigenvalues for a planar array [Chambers and Gautesen, J. Acoust. Soc. Am. **109**, 2616]. In this talk, we generalize this result to a small ellipsoid and show how the relative values of the eigenvalues depend on orientation, as well as distance and material parameters. We show specific calculations for short rods and small discs. [Work performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.]

11:15

5aSP10. An adaptive mutation breeder algorithm for determining equivalent circuit parameters of a liquid-filled piezoelectric cylinder. Yao-Ting Chang, Jevon Davies (CIR, Eng. Bldg., Tennant St., Cape Town, 8000, South Africa, lewischang79@yahoo.com.tw), Bjorn Prenzlow, and Jonathan Tapson (UCT, Rondebosch 7701, Cape Town, South Africa)

Many acoustic systems require an equivalent electromechanical model to describe their behavior. As a result, a variety of models has been described in the literature including that of Mason, KLM, and Redwood. Although these models can describe complex acoustic systems, it is often difficult to extract circuit parameters for such systems using traditional theory. Aspects of mode superposition and frequency pulling between modes cause traditional theory to become increasingly complex in predicting equivalent circuit values. As a solution to this, an adaptive breeder algorithm is utilized as a means of predicting multiple resonance modes interacting with one another. A liquid-filled piezoelectric cylinder is used as an acoustic system to represent coupled piezoelectric and liquid vibrational modes. Using a variation of Mason's transmission line model, an adaptive mutation breeder algorithm is used to predict the spectral response measured by an impedance analyzer. The acoustic model is shown to fit almost identically over the measured data for a variety of different-sized cylinders and liquids. It is further shown that the breeder algorithms may be used to predict the sound velocity and density of the liquid inside the cylinder.

Session 5aUW

Underwater Acoustics: Reverberation and Bottom Interaction

Peter G. Cable, Chair
BBN Technologies, 11 Main St., Mystic, CT 06355-3641

Chair's Introduction—8:00

Contributed Papers

8:05

5aUW1. On shallow-water bottom reverberation angular dependence. P. G. Cable   (BBN Technologies, 11 Main St., Mystic, CT 06355), Y. Dorfman, B. Dillow, and M. Goldsmith   (BBN Technologies, Cambridge, MA 02138)

Low-frequency monostatic reverberation data obtained during East China Sea ASIAEX in 2001 using Institute of Ocean Acoustics (Beijing) explosive sources and vertical line-receiving array on ShiYan-3 were analyzed in octave bands from 100–800 Hz to determine bottom-scattering strength dependence on bottom-scattering angle. Several methods for estimating the reverberation directionality spectrum were applied and results compared, employing previously described techniques to estimate transmission loss in the extraction of integrated bottom-scattering strength from reverberation [P. G. Cable *et al.*, J. Acoust. Soc. Am. **112**, 2362 (2002)]. In the two higher octave bands, scattering angle dependence for grazing angles below critical was found to be approximately proportional to first power of the sine of scattering angle, while estimation bias errors associated with limited angular resolution obscured the behavior in the lowest octave band. [Work supported by ONR under ASIAEX program.]

8:20

5aUW2. Small-slope scattering strength for a rough stratified seafloor. Robert F. Gragg   (Naval Res. Lab., Code 7144, Washington, DC 20375)

A small-slope expression is developed for the bistatic scattering strength of a seafloor with random surface relief (characterized by a general power-law spatial spectrum) and arbitrary subsurface stratification. This is done by generalizing an existing formulation for homogeneous bottoms [Gragg *et al.*, J. Acoust. Soc. Am. **110**, 2878 (2001)]. The essential points are (i) that the earlier result can be rearranged so the geoacoustic parameters enter via R, the complex reflection coefficient for the same bottom *without* roughness; and (ii) that including stratification only requires introducing the appropriate layering into the form of R. These steps yield the scattering strength for a seafloor composed of a rough fluid sediment layer over an arbitrarily layered substrate, whose strata may involve elasticity and depth profiles. Implementation only requires an algorithm to generate R for such a layered bottom. Validation is done with Essen's model environment {fluid sediment and elastic basement, both uniform [Essen, J. Acoust. Soc. Am. **95**, 1299–1310 (1994)]} and the well-known analytic expression for R. Results are benchmarked against Essen's first-order perturbation results. A BLUG substrate is then addressed, with R obtained from a subroutine borrowed from a standard propagation package. [Work supported by ONR.]

8:35

5aUW3. A preliminary sediment model for the Shallow-Water-2006 Experimental site on the New Jersey Coast. Gopu Potty and James Miller   (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882)

A preliminary sediment model for the Shallow-Water-2006 Experiment site on the continental shelf off New Jersey is presented. This sediment model will describe the bottom as layers of sediments with different properties. The layering information will be based on geological, geophys-

ical, and stratigraphy models. This layered approach to sediment modeling coupled with the Biot–Stoll (with reduced number of parameters) or Buckingham model (to define the sediment geoacoustic properties) has the potential for an efficient parametrization to set up a 3D sediment inversion scheme. The surface sediment description will be based on the ESME (effect of sound on marine environment) sediment models for the U.S. East Coast. This sediment model, developed as part of the ESME initiative, draws from a variety of data sources, including the sediment texture database maintained by the United States Geological Survey (USGS). Other sources of data which will be used to develop the preliminary sediment model include published sediment core data from the region and inversions results. This study will provide sediment data for the preliminary experimental design efforts prior to the experiment in 2006. [Work supported by ONR.]

8:50

5aUW4. Characteristics of bottom-roughness backscattering matrix in mode space. Jinrong Wu, Tianfu Gao, and Erchang Shang   (Inst. of Acoust., No. 21, BeiSiHuan XiLu, Beijing 100080, China)

The bottom backscattering matrix of a normal-mode reverberation model describes the coupling relationship between the incident normal mode and scattering normal mode in the shallow-water waveguide. The numerical analysis of the bottom roughness backscattering matrix and the result of backscattering matrix inversion from experimental data were presented from mode space. The results show that: (1) at a close distance, the value of the backscattering matrix elements inclines toward high modes. The matrix is similar to a 2D high-pass filter; (2) at a middle distance, the value of the backscattering matrix elements inclines toward middle modes. The matrix is similar to a 2D bandpass filter; (3) at a far distance, the value of the backscattering matrix elements inclines toward low modes. The matrix is similar to a 2D low-pass filter. This phenomenon is caused by mode attenuation in the shallow-water waveguide. [Work supported by the NSFC.]

9:05

5aUW5. Geoacoustic inversion of horizontal line array data from the Barents Sea. Dag Tollefsen   (Norwegian Defence Res. Establishment (FFI), Box 115, 3191 Horten, Norway) and Michael J. Wilmut   (Univ. of Victoria, Victoria, BC V8W 3P6, Canada)

Results from matched-field geoacoustic inversion of low-frequency acoustic field data recorded on a long horizontal array deployed at the seafloor are presented. Data are taken from a recent experiment conducted by FFI in the Barents Sea. Estimates of sediment parameters were obtained by inversions of multitone data from a towed source at source–array ranges of 1–6 km. The Fast Gibbs Sampler [S. E. Dosso, J. Acoust. Soc. Am. **111**, 129–142 (2001)] was used to estimate posterior probability densities of model parameters. A number of different range-independent layered fluid seabed models were tested. Estimates of sound speed of Quaternary sediment were consistent with reference values from other

geophysical data collected in the area. Details of the upper part of the sediment column that can be resolved from the short- and long-range data are presented. Sediment sound speed was better defined than density and attenuation.

9:20

5aUW6. A numerical adjoint parabolic equation (PE) method for tomography and geoacoustic inversion in shallow water. Jean-Pierre Hermand (Dept. of Optics and Acoust., Université Libre de Bruxelles, Belgium), Mohamed Berrada (Université Pierre et Marie Curie, Paris, France), Matthias Meyer (Royal Netherlands Naval College, Den Helder, The Netherlands), and Mark Asch (Université de Picardie Jules Verne, Amiens, France)

Recently, an analytic adjoint-based method of optimal nonlocal boundary control has been proposed for inversion of a waveguide acoustic field using the wide-angle parabolic equation [Meyer and Hermand, J. Acoust. Soc. Am. **117**, 2937–2948 (2005)]. In this paper a numerical extension of this approach is presented that allows the direct inversion for the geoacoustic parameters which are embedded in a spectral integral representation of the nonlocal boundary condition. The adjoint model is generated numerically and the inversion is carried out jointly across multiple frequencies. The paper further discusses the application of the numerical adjoint PE method for ocean acoustic tomography. To show the effectiveness of the implemented numerical adjoint, preliminary inversion results of water sound-speed profile and bottom acoustic properties will be shown for the YELLOW SHARK '94 experimental conditions.

9:35

5aUW7. Geoacoustic inversion by mode amplitude perturbation. Travis L. Poole (MIT/WHOI Joint Program in Oceanogr. and Oceanogr. Eng.,Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, tpoole@whoi.edu), James F. Lynch, Allan D. Pierce (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), and George V. Frisk (Florida Atlantic Univ., Dania Beach, FL 33004)

In a shallow-water waveguide the geoacoustic properties of the seafloor have a significant effect on the way sound propagates through the water. Because of this, measurements of the pressure field in the water can be used to estimate bottom properties. In this talk a perturbative method is presented which allows one to use measurements of the modal amplitudes to estimate a set of bottom parameters. A key component of the method is an expression for the derivative of the mode functions with respect to some bottom parameter. Following from the work of Thode and Kim [J. Acoust. Soc. Am. **116**, 3370–2283 (2004)], the derivative is expressed as a weighted sum over all modes (both propagating and leaky). It is thought that this method can be used alongside eigenvalue perturbation [Rajan *et al.*, J. Acoust. Soc. Am. **82**, 998–1017 (1987)] to provide an inversion scheme more robust to measurement noise. To demonstrate its feasibility, the method is applied to synthetic and real data. [Work supported by the WHOI education office.]

9:50–10:05 Break

10:05

5aUW8. Physics-based clutter via semi-deterministic simulation with bathymetric fractal realizations. Charles Monjo and Juan Arvelo, Jr. (Johns Hopkins Univ./Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

A semi-deterministic normal-mode reverberation model has been developed that accounts for bistatic bottom micro-slopes generated via bathymetric fractal realizations to simulate clutter scattering from seafloor interface features. This clutter simulator also accounts for uncertainties of the subbottom geoacoustics and on range-dependent interface scattering that varies with the seafloor sediment type. This physics-based approach is compared against measured clutter-rich bottom reverberation from the TMAST02 experiment. [This work is sponsored by ONR.]

10:20

5aUW9. Relating the distribution of bathymetry to clutter distributions. Bruce Newhall and Juan Arvelo, Jr. (Johns Hopkins Univ./Appl. Phys. Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

One source of clutter for active sonar in shallow water is backscatter from the rough ocean bottom. Backscatter may often be dominated by Bragg scale roughness. Bathymetry measurements are not generally made at sufficient resolution to determine the Bragg scale roughness distribution directly. We examine the distribution of ocean bottom slopes and determine power law relationships. These relationships can be used to define fractals which allow extrapolations of bathymetry to finer scale. The fine-scale bathymetry is then related to the distributions of sonar clutter. The generalized gamma distribution is particularly useful to describe clutter. The upper tail of this distribution approaches a power law related to the power laws in the bathymetry. [This work is sponsored by the Office of Naval Research (ONR).]

10:35

5aUW10. Broadband shallow-water reverberation statistics. Jon C. Reeves, Robert J. Ferlez, Gregory A. Babich, Gary L. Morella (Appl. Res. Lab, The Penn State Univ., P.O. Box 30, State College, PA 16804), and Anthony J. Cuetzo (The Penn State Univ., State College, PA 16804)

Utilizing a wideband transmitting/receiving system, at-sea experimental results for broadband reverberation have been acquired for both harsh as well as moderate ocean acoustic environments. The principal contributor to the backscattered reverberation is bottom features at the shallow-water ocean sites. Replica correlation (so-called match filter) processing of linear FM (LFM) pulse codes is employed to study the degree of reverberation coherence, as measured by the replica correlation coefficient. The database covers several octaves of bandwidth. The corresponding reverberation coherence is computed in terms of the variance of the cross correlation coefficient (cc) as a function of time–bandwidth product. Of particular interest is the possible departure from Rayleigh statistics as signal bandwidth increases. Several bandwidths are compared, using the Kolmogoroff–Smirnof test as a measure of departure from Rayleigh. [This work was sponsored by A. Nucci, ONR 333.]

10:50

5aUW11. Non-Rayleigh reverberation prediction for shallow-water waveguides. Kevin D. LePage (NRL, Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

In a previous paper the second moment of monostatic reverberation intensity was derived for rough surfaces in range-independent waveguides with χ_1^2 scatterer amplitude statistics and Gaussian spatial correlation functions [J. Acoust. Soc. Am. **117**, 2611 (2005)]. Here, these results are expanded to bistatic geometries and the implications for the degree of non-Rayleighness of the resulting reverberation pressure envelope statistics are discussed. [Work supported by ONR.]

11:05

5aUW12. Nonlinear compressional waves in marine sediments. B. Edward McDonald (Naval Res. Lab, Code 7145, Washington, DC 20375, mcdonald@sonar.nrl.navy.mil)

A theory for nonlinear waves in marine sediments must account for the presence of a granular frame filled with water and possibly gas bubbles. When grains are in full contact, the stress–strain relation for the sediment contains a contribution varying as strain to the power 3/2, referred to as the Hertz force. The quadratic nonlinearity parameter derived from the second pressure derivative with respect to density thus diverges in the limit of small strain. We present a simple nonlinear wave equation model (a variant of the NPE) for compressional waves in marine sediments that avoids Taylor expansion and the problem of diverging nonlinearity parameter. An equation of state for partially consolidated sediments is derived from consolidation test results. Pressure is found to increase with overdensity to the power 5/2, indicating an increase in the number of contacts per

5a FRI. AM

grain as density increases. Numerical results for nonlinear compressional waves show agreement with analytic self-similar profiles derived from the nonlinear wave equation. [Work supported by the ONR.]

11:20

5aUW13. Nonlinear parameter estimation in water-saturated sandy and muddy sediments. Byoung-Nam Kim, Kang Il Lee, Jiehui Liu, and Suk Wang Yoon (Dept. Phys., Sung Kyun Kwan Univ., Suwon 440-746, Rep. Korea, swyoon@skku.ac.kr)

Nonlinear acoustic responses of water-saturated sediments are very important to understand nonlinear phenomena of gassy ocean sediments. Especially, the second harmonic, the sum, and the difference frequency acoustic waves in water-saturated sediments can provide practical criteria to estimate the nonlinear parameters of gassy sediments. In this paper, the difference frequency acoustic waves in water-saturated sandy and muddy sediments were observed in a water tank experiment using a pulse transmission technique. Their pressure levels were 10 dB higher than the background noise level at a maximum driving pressure of source acoustic transducer. The experimental results were compared with theoretical estimations of the parametric acoustic array. The nonlinear parameters of water-saturated sandy and muddy sediments were estimated as a few tens in their comparison. The nonlinear parameters were also compared with those obtained with the second harmonic acoustic wave. [Work supported by MOCIE (Ministry of Commerce, Industry and Energy).]

11:35

5aUW14. An autonomous underwater vehicle technique for *in situ* waveguide characterization. Jason D. Holmes, William M. Carey (Boston Univ., Boston, MA 02215), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Results from an autonomous underwater vehicle (AUV) towed hydrophone array experiment are presented as a demonstration of a new *in situ* measurement technique. In this technique, a source is deployed from a small vessel and the AUV navigates along a pre-programmed path towing the hydrophone array. The added degrees of freedom associated with an array towed by an AUV provides for rapid characterization of the shallow-water waveguide. Using a synthetic aperture, the horizontal wave number spectrum is measured and the modal eigenvalues are experimentally measured by picking off the peak locations in the wave number spectrum. The modal amplitudes are found by using the finite aperture array approximation to the Hankel transform in which the amplitude of the spectral peaks are finite and proportional to the mode excitation amplitude as shown by Frisk *et al.* [J. Acoust. Soc. Am. **86**, 1928–1939 (1989)]. By measuring the modal attenuation, one can determine the effects of bottom interaction. Another measurement technique using the same towed hydrophone system is described in which signal path bottom reflection results are obtained using a parametric source.