CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 116, No. 3, September 2004

ACOUSTICAL NEWS—USA 1291

USA Meetings Calendar 1298

ACOUSTICAL STANDARDS NEWS 1309

Standards Meetings Calendar 1309

TECHNICAL PROGRAM SUMMARY 1315

REVIEWS OF ACOUSTICAL PATENTS 1317

LETTERS TO THE EDITOR

Estimation of seabed reflection loss properties from direct blast pulse shape (L) — Mark K. Prior, Chris H. Harrison — 1341

Pseudo-surface waves above thin porous layers (L) — Jean F. Allard, Michel Henry, Vincent Gareton — 1345

Time reversal and the spatio-temporal matched filter (L) — D. H. Chambers, J. V. Candy, S. K. Lehman, J. S. Kallman, A. J. Poggio, A. W. Meyer — 1348

On the dichotomy in auditory perception between temporal envelope and fine structure cues (L) — Fan-Gang Zeng, Kaibao Nie, Sheng Liu, Ginger Stickney, Elsa Del Rio, Ying-Yee Kong, Hongbin Chen — 1351

GENERAL LINEAR ACOUSTICS [20]

Plane wave reflection coefficient from near field measurements — Chris H. Harrison, Peter L. Nielsen — 1355

A fast Fourier transform on multipoles (FFTM) algorithm for solving Helmholtz equation in acoustics analysis — Eng Teo Ong, Heow Pueh Lee, Kian Meng Lim — 1362

An integral equation approach to three-dimensional acoustic radiation and scattering problems — S. A. Yang — 1372

On Webster's horn equation and some generalizations — P. A. Martin — 1381

A super-spectral finite element method for sound transmission in waveguides — Andrew Peplow, Svante Finnveden — 1389

Notes on the Gaussian beam expansion — Desheng Ding, Yu Zhang — 1401

NONLINEAR ACOUSTICS [25]

Acoustoelastic effects on guided waves in a fluid-filled pressurized borehole in a prestressed formation — Zhengliang Cao, Kexie Wang, Zaitian Ma — 1406

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 3, SEPTEMBER 2004

CONTENTS—*Continued from preceding page*

AEROACOUSTICS, ATMOSPHERIC SOUND [28]

Leading edge noise from thick foils in turbulent flows — Jonathan Gershfeld — 1416

Method of evaluating dipole sound source in a finite computational domain — T. Takaishi, M. Ikeda, C. Kato — 1427

UNDERWATER SOUND [30]

Fredholm integral equation for shape reconstruction of an underwater object using backscattered ramp response signatures — Wei Li, G. R. Liu, V. K. Varadan — 1436

Propagation of sound through a spicy ocean, the SOFAR overture — Matthew Dzieciuch, Walter Munk, Daniel L. Rudnick — 1447

ULTRASONICS, QUANTUM ACOUSTICS, AND PHYSICAL EFFECTS OF SOUND [35]

Resonance frequency of microbubbles: Effect of viscosity — Damir B. Khismatullin — 1463

Frequency-dependent acoustic properties of a fluid/porous solid interface — A. A. Gubaidullin, O. Y. Kuchugurina, D. M. J. Smeulders, C. J. Wisse — 1474

Simplifications for the calculation of surface wave pulses generated by laser-irradiation — J. D. Achenbach — 1481

Electromagnetoacoustic surface waves on dispersive piezoelectric layered media — Maurizio Romeo — 1488

Focused array transducer for two-dimensional optoacoustic tomography — Victor Kozhushko, Tanya Khokhlova, Alexei Zharinov, Ivan Pelivanov, Vladimir Solomatin, Alexander Karabutov — 1498

Open cycle traveling wave thermoacoustics: Energy fluxes and thermodynamics — Nathan T. Weiland, Ben T. Zinn — 1507

Material discrimination using bispectral signatures — Paul A. Nyffenegger, Melvin J. Hinich, Dillon Ritter, Stephen Hansen — 1518

Spectral methods for modelling guided waves in elastic media — A. T. I. Adamou, R. V. Craster — 1524

TRANSDUCTION [38]

Thin, lightweight electroacoustic projector for low frequency underwater applications — James F. Tressler, Thomas R. Howarth, Walter L. Carney — 1536

Analysis of screw pitch effects on the performance of bolt-clamped Langevin-type transducers — Kazunari Adachi, Toru Takahashi, Hiroshi Hasegawa — 1544

STRUCTURAL ACOUSTICS AND VIBRATION [40]

A wide-angle on-surface radiation condition applied to scattering by spheroids — David C. Calvo — 1549

Implicit methods of solution to integral formulations in boundary element method based nearfield acoustic holography — Nicolas Valdivia, Earl G. Williams — 1559

NOISE: ITS EFFECTS AND CONTROL [50]

Subjective response of people to simulated sonic booms in their homes — David A. McCurdy, Sherilyn A. Brown, R. David Hilliard — 1573

ARCHITECTURAL ACOUSTICS [55]

Sound level distribution and scatter in proportionate spaces — Stephen Chiles, Mike Barron — 1585

Use of statistical energy analysis for the prediction of Sabine absorption coefficient of modally reactive panels — K. S. Sum — 1596

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 3, SEPTEMBER 2004

CONTENTS—*Continued from preceding page*

Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces — Masayuki Morimoto, Hiroshi Sato, Masaaki Kobayashi — 1607

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

A note on measurement of sound pressure with intensity probes — Peter Juhl, Finn Jacobsen — 1614

ACOUSTIC SIGNAL PROCESSING [60]

Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging — Zhisong Wang, Jian Li, Petre Stoica, Toshikazu Nishida, Mark Sheplak — 1621

PHYSIOLOGICAL ACOUSTICS [64]

Modeling the growth rate of distortion product otoacoustic emissions by active nonlinear oscillators — Renata Sisto, Arturo Moleti — 1632

Frequency specificity of distortion-product otoacoustic emissions produced by high-level tones despite inefficient cochlear electromechanical feedback — Sirley Carvalho, Thierry Mom, Laurent Gilain, Paul Avan — 1639

Otoacoustic emissions without somatic motility: Can stereocilia mechanics drive the mammalian cochlea? — M. C. Liberman, Jian Zuo, J. J. Guinan, Jr. — 1649

Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration — Yonatan I. Fishman, Joseph C. Arezzo, Mitchell Steinschneider — 1656

PSYCHOLOGICAL ACOUSTICS [66]

Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise — Frédéric Apoux, Sid P. Bacon — 1671

Spectral loudness summation and simple reaction time — Eva Wagner, Mary Florentine, Søren Buus, Joseph McCormack — 1681

Masked auditory thresholds in sciaenid fishes: A comparative study — John Ramcharitar, Arthur N. Popper — 1687

The discrimination of baboon grunt calls and human vowel sounds by baboons — Robert D. Hienz, April M. Jones, Elise M. Weerts — 1692

Localization ability with bimodal hearing aids and bilateral cochlear implants — Bernhard U. Seeber, Uwe Baumann, Hugo Fastl — 1698

SPEECH PRODUCTION [70]

Flow separation in a computational oscillating vocal fold model — Fariborz Alipour, Ronald C. Scherer — 1710

Sound generation by steady flow through glottis-shaped orifices — Zhaoyan Zhang, Luc Mongeau, Steven H. Frankel, Scott Thomson, Jong Beom Park — 1720

An acoustic description of the vowels of Northern and Southern Standard Dutch — Patti Adank, Roeland van Hout, Roel Smits — 1729

The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese — Johanna G. Barry, Peter J. Blamey — 1739

SPEECH PERCEPTION [71]

Integration across frequency bands for consonant identification — Diane Ronan, Ann K. Dix, Phalguni Shah, Louis D. Braida — 1749

Auditory discontinuities interact with categorization: Implications for speech perception — Lori L. Holt, Andrew J. Lotto, Randy L. Diehl — 1763

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 3, SEPTEMBER 2004

CONTENTS—*Continued from preceding page*

SPEECH PROCESSING AND COMMUNICATION SYSTEMS [72]

Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition	Mark D. Skowronski, John G. Harris	1774
Evaluation of formant-like features on an automatic vowel classification task	Febe de Wet, Katrin Weber, Louis Boves, Bert Cranen, Samy Bengio, Hervé Bourlard	1781

MUSIC AND MUSICAL INSTRUMENTS [75]

A nonmusical paradigm for identifying absolute pitch possessors	David A. Ross, Ingrid R. Olson, Lawrence E. Marks, John C. Gore	1793
Detection of random alterations to time-varying musical instrument spectra	Andrew Horner, James Beauchamp, Richard So	1800

BIOACOUSTICS [80]

Neuro-computational processing of moving sonar echoes classifies and localizes foliage	Roman Kuc	1811
Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues	Xiufen Gong, Dong Zhang, Jiehui Liu, Huanlei Wang, Yongsheng Yan, Xiaochen Xu	1819
Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*)	Marianne H. Rasmussen, Magnus Wahlberg, Lee A. Miller	1826
Correlation of ultrasonic scatterer size estimates for the statistical analysis and optimization of angular compounding	Anthony Gerig, Quan Chen, James Zagzebski, Tomy Varghese	1832

ERRATA

Erratum: 5aMU1. Movement amplitude and tempo change in piano performance [J. Acoust. Soc. Am. 115, 2590 (2004)]	Caroline Palmer, Simone Dalla Bella	1842

CUMULATIVE AUTHOR INDEX 1843

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

Alain de Cheveigne—For contributions to computational modeling of auditory processing of pitch and sound source segregation.

D. Keith Wilson—For contributions to atmospheric acoustics.

Eric J. Wood—For contributions to industrial noise control.

Stephen A. Fausti receives Magnuson Award

Stephen A. Fausti, Ph.D., Associate Chief of Staff and Director of the Veterans Affairs Rehabilitation Research and Development National Center for Rehabilitative Auditory Research (NCRAR) and Professor in the Department of Otolaryngology, Oregon Health and Science University, has received the 2004 Paul B. Magnuson Award for Outstanding Achievement in Rehabilitation Research and Development. The award is presented annually to a VA RR&D investigator who exemplifies the entrepreneurship, humanitarianism, and dedication to veterans displayed by Dr. Magnuson during his career. It is the highest award for rehabilitation research investigators. Dr. Patricia Dorn, Assistant Director for RR&D Service, presented the award to Dr. Fausti Wednesday, 31 March 2004, in Salt Lake City, Utah, at the American Academy of Audiology National Convention.

Dr. Fausti established the National Center for Rehabilitative Auditory Research in 1997 at the Portland VA Medical Center to conduct research to increase knowledge about hearing impairment and to develop technology to improve the care and treatment of veterans with hearing loss. The Magnuson Award consists of a one-time $5000 cash prize, $50 000 per year for up to three years to supplement ongoing peer-reviewed research, and a celebratory plaque. To read more about the 2004 Magnuson Award see http://www.va.gov/resdev/about/wn.cfm

Annual Reports of the Technical Committees

Acoustical Oceanography

Fall 2003 Meeting (Austin). Acoustical Oceanography (AO) sponsored a special session entitled ''Developments in Multibeam Sonar for Water-Column Measurements,'' which was organized by Kenneth Foote (Woods Hole-Oceanographic Institution). AO, in addition, co-sponsored a session with Underwater Acoustics entited ''Aubrey L. Anderson Memorial Session on Acoustics of Gas-Bearing Sediments,'' which was organized by Anthony Lyons (Applied Research Laboratory, Pennsylvania State University) and Michael Richardson (Naval Research Laboratory, Stennis Space Center). Darrell Jackson (Applied Physics Laboratory, University of Washington) represented AO in the Hot Topics in Acoustics session, presenting a paper entitled ''Hot topics in acoustical oceanography.'' The Best Student Paper Awards in Acoustical Oceanography were awarded to Deanelle Symonds (Massachusetts Institute of Technology), who won first place for her paper ''Fish schools are the dominant cause of long-range active sonar clutter in the New Jersey Continental Shelf: Quantitative correlations'' and Ioannis Bertsatos (Massachusetts Institute of Technology), who won second place for his paper ''Optimally resolving Lambertian surface orientaion.'' James Miller (University of Rhode Island) and Peter Worcester (Scripps Institution of Oceanography) represented AO at the Technical Program Organizing Meeting.

Spring 2004 Meeting (New York). The 2004 Acoustical Oceanography Prize in Acoustical Oceanography was awarded to Stan Dosso of the University of Victoria, ''for the development of non-linear methods for geoacoustic inversion and acoustic localization.'' He gave the Acoustical Oceanography Prize Lecture, entitled ''Ocean acoustic inversion for seabed geoacoustic properties.'' AO sponsored two special sessions. Ralph Stephen (Woods Hole Oceanographic Institution) and Robert Odom (Applied Physics Laboratory, University of Washington) organized a session entitled ''Ocean Acoustics of Earthquakes.'' Orest Diachok (Poseidon Sound) organized a session jointly with Animal Bioacoustics entitled ''D. Vance Holliday Special Session on Acoustical Measurements of Marine Organisms.'' Joshua Wilson (Massachusetts Institute of Technology) won first place in the competition for the Best Student Paper Award in Acoustical Oceanography for his paper entitled ''Experimental demonstration of accurate hurricane classification from local wind speed estimates obtained with underwater sound.'' Cristina Tollefsen (Memorial University of Newfoundland) was

 © 2004 Acoustical Society of America

awarded second place for her paper entitled ''Sources of uncertainty in Doppler sonar measurments of fish speed.'' Ellen Livingston (Office of Naval Research) and Kyle Becker (Applied Research Laboratory, Pennsylvania State University) 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 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 final one as Chair of the AO Technical Committee. N. Ross Chapman (University of Victoria) is the incoming Chair. I wish Ross well and hope that he has as much fun as Chair as I have. I want to close by thanking all of the volunteers who have made the activities of the AO Technical Committee possible during my tenure as Chair.

PETER F. WORCESTER
Chair, 2001–2004

Animal Bioacoustics

Animal Bioacoustics (AB) sponsored well-attended special sessions at both the Austin and the New York meetings. At the Austin meeting, George Pollak organized a session on ''Neurobiology of communication,'' and Darlene Ketten organized one, co-sponsored by Signal Processing, on ''Medical imaging techniques to understand auditory processing.'' In addition, we co-sponsored with Signal Processing, Underwater Acoustics, Speech Communication, Noise, and Engineering Acoustics four sessions on ''Detection and Classification in Acoustics.'' Fifteen people attended the technical committee meeting. Our main item of business was to discuss methods for electing new chairs of the Technical Committee (TC). We agreed to amend our previous procedure, based on paper ballots, to allow email ballots to be sent to all ASA members indicating AB as their primary interest.

The New York meeting was very heavily attended, with most AB sessions being standing room only. Three sessions on ''Natural acoustic behavior of animals,'' organized by James Simmons, were designated as memorial sessions for Donald R. Griffin, who died on 7 November 2003. Two papers presented at this session, one by Joshua Schwartz and another by Dianne Suggs and Andrea Simmons, were featured in the BBC radio program ''The Leading Edge'' which aired on 3 June 2004. Another session, organized by Edmund Gerstein, presented work on ''Infrasonic communication by animals: Signal propagation, generation, reception and function.'' AB co-sponsored with Acoustical Oceanography three sessions in honor of D. Van Holliday, ''Acoustical measurements in marine organisms.'' We also co-sponsored, along with all other technical committees, the special session to honor Leo Beranek on the occasion of his 90th birthday. Kelly Benoit-Bird presented a talk titled ''Acoustics will play an important role in revealing the biological secrets of the ocean,'' in the session on the Future of Acoustics during the Celebration Day ceremonies. Ann Bowles represented AB at the special session on ''Education in Acoustics: Careers in Acoustics,'' where she spoke on the topic ''What happens when you want to talk to the animals: Lessons of a career in animal bioacoustics.''

Prizes for the Best Student Paper Award were given to two students presenting their work at the New York meeting. Susan E. Parks of Cornell University won first prize for her presentation on ''The gunshot sound produced by male North Atlantic right whales and its potential function in reproductive advertisement.'' Jennifer Miksis-Olds of the University of Rhode Island won second prize for her presentation on ''The acoustic environment of the Florida manatee: Correlation with level of habitat use.'' Congratulations to both Susan and Jennifer for jobs well done.

Over 50 people attended the technical committee meeting in New York. Marla Holt is our new representative to the Student Council, replacing Kelly Benoit-Bird. AB has two new associate editors for ASA publications: Cynthia Moss, for ARLO and James Simmons for JASA. New members of the TC, with terms to 2007, are Sheryl Coombs, Richard Fay, Edmund Gerstein, Seth Horowitz, Cynthia Moss and Jeanette Thomas. AB will be joining other technical committees in co-sponsoring the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) meetings in Canada in 2005.

Dedicated volunteers continue to assist the operations of our technical committee. Ann Bowles remains as our representative to the ASA Committee on Standards (ASACOS), and has been diligent in preparing materials for standards related to animal bioacoustics. She presented a paper updating the issue of standards at the New York meeting. Robert Benson and James Simmons served as our TPOM representatives to the Austin and New York meetings, respectively. Charles Greene and Darlene Ketten continue as our representatives to the Medals and Awards Committee and the Membership Committee, respectively.

We look forward to continued participation in AB sessions for the upcoming year.

ANDREA SIMMONS
Chair

Architectural Acoustics

Members of the Technical Committee on Architectural Acoustics (TCAA) continue to be remarkably hardworking and productive, and our membership continues to grow.

Eighty-seven members attended our Technical Committee meeting in November, 2003 in Austin, TX. The committee members offered hearty congratulations to Dr. Leo Beranek, 2002 National Medal of Science Laureate, presented by President George W. Bush, at the White House. Members Ron Freiheit and Tim Foulkes were announced as new Fellows of the Acoustical Society of America. Dick Campbell was elected Vice-President of the ASA Foundation.

Special session chairs included Ken Roy, Alex Case, Yashushi Shimizu, Jack Randorff, Peter Ladefoged, Richard Lyon, William Hartmann, Sanford Fidell, and David Marsh.

An outreach by the Classroom Acoustics Task Force is being chaired by Bob Coffeen and Red Wetherill, who are directing much activity toward clients and organizations such as the National Education Association, state agencies, unions, and teachers. Three new websites on Classroom Acoustics are up and running.

The Student Council announced plans for a pilot student chapter at the University of Nebraska, and also announced plans to offer a mentoring award. Other business items included continuing work by the Concert Hall Research Group, and the publication of the Halls for Music Performance book. A committee is being organized to address product safety, primarily for nonacoustical issues such as flammability, toxicity, and other safety issues.

The second Concert Hall Acoustics Summer Institute was a great success, held August 11–14, 2003, in Saratoga Springs, NY, co-sponsored by the National Council of Acoustical Consultants. At least 20 students attended with full fellowships.

Technical Initiatives included: $750 for the Student Design Competition for the New York City ASA meeting; $600 for student receptions, co-sponsored by the National Council of Acoustical Consultants; $300 for website maintenance; $2,000 for Outreach for Classroom Acoustics, and $1,000 to work with Education in Acoustics on developing and promoting lesson plans. A technical initiative was submitted for $500 to cover travel for an invited speaker, Harold Marshall, for participation in the New York City-''Multisensory'' special session.

Lily Wang was elected the next chair of the Technical Committee on Architectural Acoustics.

Eighty-eight members attended our Technical Committee meeting in May, 2004 in New York City. TCAA held its meeting jointly with the Technical Committee on Noise, primarily because of the nature of the schedule for the overall meeting, and because so many members have deep interest in both technical committees.

Congratulations were offered to the Acoustical Society for its 75th Anniversary, and to Leo Beranek for his 90th birthday. David Lubman was awarded the Helmholtz-Rayleigh Interdisciplinary Silver Medal. Jian Kang, Keith Wilson, and Eric Woods have been elected new Fellows of the ASA. Student winners of the 2004 TCAA/NCAC Student Design Competition were announced, with results that can be found through links at the TCAA website. TCAA also offered congratulations to all those who participated in the ASA 75th Anniversary Book, especially Bill Cavanaugh and Hank Bass, who served as editors.

Special Session Chairs at the New York meeting were Rendell Torres, Lily Wang, Jerry Hyde, Bob Coffeen, Robin Glosemeyer, Yoichi Ando, Jim Jeon, Greg Miller, Neil Shaw, Gary Elko, James West, Elizabeth Cohen, Mendel Kleiner, Chris Jaffe, Neil Rolnick, and Ben Mueller.

Technical Initiatives supported a variety of activities including: $750 for the Student Design Competition; $300 for the Student Reception; $300 for the web page maintenance. George Wong presented a request for financial support toward a meeting of the International Organization for Standardization, to be held in North America.

The first pilot student chapter has been established at the Unversity of Nebraska. Templates for other student chapers will be available soon. Fifteen nominations were submitted for the Student Council Mentoring Award, and the award recipient will be announced at the San Diego plenary meeting.

Updates from the Concert Hall Research Group, the Newman Student Award Fund, the Classroom Acoustics Group, and the Acoustic Data Working Group were discussed.

Congratulations to Lily Wang, who was introduced as the new Chair of the Technical Committee on Architectural Acoustics.

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

K. ANTHONY HOOVER
Chair, 2001–2004

Biomedical Ultrasound/Bioresponse to Vibration

At the Fall meeting in Austin, Texas the main event for the Biomedical Ultrasound/Bioresponse to Vibration (BU/BV) committee was a one day Topical Meeting on Shock Wave Therapy run by Michael Bailey and Pei Zhong. This was held on Friday and consisted of presentations and panel discussions involving scientific researchers, medical doctors, and engineers from industry. The forum provided a great opportunity for challenges and opportunities for the use of shock waves for multiple therapeutic applications. BU/BV had three special sessions: ''Contrast Bubble Agents'' organized by Constantin Coussios, ''Multidisciplinary and Non-Conventional Approaches in Biomedical Ultrasound'' organized by Stas Emilianov and ''Tissue Harmonic Imaging'' organized by yours truly. In addition, we co-sponsored a session with Physical Acoustics entitled ''Nonlinear Acoustics in Honor of David Blackstock.'' Our Technical Program Organizing Meeting (TPOM) representative to the Austin meeting was Glynn Holt. At the Technical Committee meeting on Thursday evening there were 43 people in attendance.

The New York meeting was focused around the 75th birthday party of the Society. BU/BV had three special sessions, ''High frequency imaging organized'' by Jeff Ketterling, ''Elasticity imaging organized'' by Elisa Konofagou, and ''HIFU'' organized by Cheri Deng, and three sessions made up of contributed papers. We also co-sponsored the Robert E. Apfel Memorial Session. There were a total of 91 BU/BV talks at the meeting which I suspect is one of the highest numbers we have had. The attendance in the sessions was outstanding—typically in excess of 50 people present and very often standing room only. The technical committee meeting was held in conjunction with the open meeting of the Physical Acoustics Technical Committee (Chair Thomas Matula) with a combined attendance of 69. Thanks to our TPOM representatives Jeff Ketterling and Elisa Konofagou for managing to organize the sessions under very constrained conditions. Jeff was also on the local organizing committee for the meeting and so played a significant role in helping with the general smooth running of this large and rather complex meeting.

Beyond the technical sessions BU/BV was delighted that one of its members, Michael Bailey, was presented with the R. Bruce Lindsay Award for contributions to the understanding of shock wave lithotripsy and nonlinear acoustics. Mike was presented with his award at the plenary session on Wednesday and introduced by Lawrence Crum.

As is customary for the spring meeting we held a student paper competition to which there were 16 entries. First prize was awarded to Paolo Zanetti of Boston University for his talk entitled ''Signal-to-noise ratio and attenuation of Optison microbubbles in blood as a function of imaging frequency'' and second prize to Jeremy Bercoff of Laboratoire Ondes et Acoustique, Paris for his talk ''Studying viscoelasticity in soft tissues with supersonic shear imaging.'' All the students who entered the competition are to be commended as the quality of presentations was the highest I have ever seen. Also on the student front our new representative to the Student Council, Neil Owen, took over from Tyrone Porter.

For the San Diego meeting in November 2004, BU/BV is planning one special session, ''Light and sound for imaging in the body,'' to be organized by Ron Roy and Todd Murray. We will be holding a one-day Topical Meeting on ''Ultrasound Characterization of Cancellous and Cortical Bone'' to be organized by Kendall Waters. For the Vancouver meeting in Spring 2005 four special sessions are planned: Imaging using audible acoustics (Tom Royston), Therapeutic ultrasound for the brain (Pierre Mourad and Carr Everbach), Gene therapy and molecular imaging (Pierre Mourad and Mike Averkiou), and Tissue therapy by sound (Mike Bailey).

The BU/BV technical committee is active and healthy and I would like to thank all the members who have helped in making it so vibrant. I would be delighted to receive suggestions on how we can improve further.

ROBIN CLEVELAND
Chair

Engineering Acoustics

The Engineering Acoustics Technical Committee (EATC) met at each of the two meetings of the Society, Austin in the Fall, and New York in the Spring.

In Austin, the EATC sponsored two special sessions and one session of contributed papers. The special sessions and their organizers were ''Honoring the Contributions of Elmer Hixson,'' Brandon D. Tinianov and John T. Post, and ''New Technologies in Active Control,'' Robert D. Corsaro and Jiri Tichy. In addition, EATC co-sponsored several other sessions with the various technical committees. The Committee offers many thanks to James H. Miller, who served as the representative to the Austin Technical Program Organizing Meeting (TPOM). The EATC met in Austin on the evening of Tuesday, 11 November.

In New York, the EATC sponsored two sessions of contributed papers and co-sponsored two special sessions. The special sessions were joint with The Underwater Acoustics Technical Committee. The topics and their organizers were ''Autonomous Underwater Vehicle Acoustics: Part 1: Concepts and Systems,'' Henrik Schmidt and Thomas R. Howarth, and ''Autonomous Underwater Vehicle Acoustics: Part 2: Hardware and Devices,'' Thomas R. Howarth and Henrik Schmidt. The Committee offers many thanks to Gary W. Elko and Ilene J. Busch-Vishniac, who served as EATC representatives to the New York TPOM.

The Committee wishes to recognize Gary W. Elko, for his work as Meeting Technical Chair for the very well attended 75th Anniversary meeting of the Society. Additionally the Committee deeply thanks Stan Ehrlich for his generation and preparation of the manuscript ''The History of Engineering Acoustics,'' which was included in the 75th Anniversary Book (Chapter 8), and for his continuing contributions to the Committee.

The EATC met in New York on the evening of Thursday, May 27. The Committee welcomes Alexandra Loubeau (Penn State) as Student Council Representative for EA, and Ken Walsh (K&M Engineering LTD) as our latest new member.

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

The Technical Committee on Musical Acoustics (TCMU) was active in 2003–2004. In addition to organizing special sessions for the Austin and New York meetings, the technical committee was involved in a number of events of wider interest to those attending the meetings as well as in activity not directly related to meetings.

There were some excellent special sessions in musical acoustics at the Austin meeting in November 2003. Peter Hoekje organized and chaired a session on wind instruments and nonlinearity. Uwe Hansen and Tom Rossing co-chaired a joint session with Education, ''Neat Acoustics Websites and Software for Teaching Musical Acoustics.'' Jim Cottingham chaired the session, ''Where are they now?'' in which past winners of the Best Student Paper Award in musical acoustics reported on their current research and other activities. A very special pair of sessions honored the contributions of Gabriel Weinreich to musical acoustics. The two sessions, one on piano acoustics and one on bowed strings, were organized and chaired by Tom Rossing and Uwe Hansen. We were very pleased that Gabi was able to attend the Austin meeting to join us in these sessions. In the Hot Topics in

Acoustics session at Austin, Russell Pinkston of the University of Texas presented ''Gerrymander: An interactive musical composition for clarinet and computer.'' The paper consisted of a demo performance followed by a detailed description of the software. In conjunction with the wind instrument session, Jim Cottingham arranged a concert by the University of Texas Chamber Winds, conducted by Scott Hanna. This concert was attended by about 300 people.

The 75th Anniversary meeting in New York included a number of special events with a musical flavor. A barbershop quartet from our technical committee provided musical entertainment as part of the 75th anniversary banquet, and the Wednesday morning celebratory session featured a demonstration of a virtual orchestra. A high point in the Wednesday morning program for our technical committee was the awarding of the Silver Medal in Musical Acoustics to Johan Sundberg. An edited version of the history lecture given by Gabi Weinreich in 1999 appeared as the musical acoustics chapter in the ASA 75th Anniversary book.

The technical program in New York included some special sessions in musical acoustics with a historical focus. Tom Rossing and Murray Campbell co-chaired ''New Research on Pre-1929 Instruments,'' and Paul Wheeler and Ian Lindevald co-chaired ''Musical Instruments Developed During the ASA Era.'' Other special sessions were ''Digital Signal Processing Methods for Restoring, Enhancing and Manipulating Music Recordings,'' organized by Jim Beauchamp, and ''Neurophysiology of Musical Instrument Playing,'' organized by Ingo Titze.

The technical committee continued to encourage student involvement in musical acoustics through the student paper competition. The winner of the First Prize Student Paper Award at Austin was Todd W. Bigelow of Utah State University who presented ''Software tools for developing an acoustics multimedia CD-ROM.'' Second prize went to Ammon Paquette from Augustana College (Rock Island, IL) for ''Modes of vibration of air-driven free reeds in transient and steady state oscillation.'' First prize in New York was awarded to Stephanie Weisser of the Universite Libre de Bruxelles who presented ''The acoustics of the Bagana.'' The second prize winner was ''The application of cochlear audio analysis techniques to electro-acoustic music,'' presented by John A. Mills III from the University of Texas. Andrew Morrison of Northern Illinois University, our ASA Student Council Representative, keeps the technical committee informed of Student Council activities, and is currently the Student Council liaison to the ASA Online Education Committee.

Concern has been expressed several times in our technical committee meetings over the last few years that the musical acoustics PACS classifications (43.75) are in need of significant revision. This year a subcommittee consisting of Jim Beauchamp and Tom Rossing prepared a list of proposed revisions. These proposed revisions were submitted to *ASA* Editor in Chief Allan Pierce who has forwarded them to the American Insitute of Physics. Funding for two ongoing technical initiative projects has continued into 2004. One is a project being conducted by Peter Hoekje involving preparation of a set of demonstrations in musical acoustics to be available on CD-ROM and on the web. The second is a project in which selected papers of current interest in musical acoustics are being translated into English. The editorial board for this project is headed by Tom Rossing.

New ASA Fellows from Musical Acoustics in 2003–2004 are Murray Campbell and Julius Smith. The representatives to the Technical Program Organizing Meetings in 2003–2004 were Jim Cottingham assisted by Wilson Nolle for the Austin meeting and Uwe Hansen and Vic Sparrow for the New York meeting. Representation from the TCMU on ASA committees during the year included Jim Beauchamp on the Medals and Awards Committee, Uwe Hansen on the Membership Committee, and Ian Lindevald on the Committee on Standards. Neville Fletcher and Stephen McAdams served as *JASA* Associate Editors for Musical Acoustics. Members appointed or reappointed to serve on the technical committee for terms expiring in 2006 include R. Dean Ayers, Judith C. Brown, Courtney B. Burroughs, John R. Buschert, George F. Emerson, Thomas M. Huber, Bozena Kostek, Barry Larkin, Daniel O. Ludwigsen, Thomas D. Rossing, David B. Sharp, Julius O. Smith, and William J. Strong.

The TCMU web site continues to maintain current information on the technical committee, including committee minutes and annual reports as well as information on musical acoustics in general. The site (at http://www.public.coe.edu/~jcotting/tcmu) was updated and given a new look in the spring of 2004. The site includes information on upcoming meetings as well as the lists ''Musical Acousticians'' and ''Websites for Teaching Acoustics.''

JAMES P. COTTINGHAM
Chair

Noise

To start, we thank Bennett Brooks for his three years of service as the Chair of the Technical Committee on Noise. He has worked hard and effectively to support the activities of the members of Noise.

The Fall 2003 meeting of the ASA was held in Austin, Texas. The Technical Committee on Noise organized or co-organized several interesting and diverse special sessions. These were ''Forensic Acoustics'' (this session and a Forensic Acoustics Seminar were organized by Jack Randorff), ''Acoustical Design Issues for the Health Care Industry'' (Kerrie Standlee), ''Sound Quality for Engineered Products and the Environment'' (Brigitte Schulte-Fortkamp, Klaus Genuit and Patricia Davies), ''Sound Quality—When Sound is the Essential Quality'' (Richard Lyon), ''Building Code Noise Compliance'' (Daniel Raichel), ''Honoring the Contributions of Elmer Hixson'' (Brandon Tinianov), ''New Technologies in Active Control'' (Jiri Tichy and Robert Corsaro), and ''Progress in Measurement of Transportation Noise since Ted Schultz'' (Sanford Fidell). Mike Stinson had the difficult job of summarizing the breadth of current activity in noise in the ''Hot topics in noise'' lecture. Sanford Fidell presented the Architectural Acoustics and Noise Distinguished Lecture ''Analysis of community response to transportation noise a quarter century after Schultz.'' Our appreciation to Brandon Tinianov for serving as the Noise representative at the Technical Program Organizing Meeting.

The Spring 2004 New York City meeting was the Society's 75th Anniversary Celebration. Special sessions organized or co-organized by Noise were ''Noise Impact Evaluation: Old and New'' (Brigitte Schulte-Fortkamp and Klaus Genuit), ''Urban Transit Noise'' (Paul Burge and Daniel Raichel), ''Noise and Society'' (Nancy Nadler), ''Noise in Large Cities'' (Daniel Raichel), and ''On the Occasion of His 90th Birthday: Special Session to Honor the Contributions of Leo L. Beranek to Acoustics and Teaching'' (Tony Embleton). This last special session was co-sponsored by every other ASA Technical Committee. Laymon Miller presented the Noise and Architectural Acoustics Distinguished Lecture ''Noise, my 62 years of it.'' Brigitte Schulte-Fortkamp and Bennett Brooks were the Noise representatives at the Technical Program Organizing Meeting.

The contributions of quite a few Noise people were recognized at these two meetings. The winner of the Noise Young Presenter award for the Austin meeting was Meghan Saweikis for her talk ''Speech intelligibility index predictions for young and old listeners in automobile noise: Can the index be improved by incorporating factors other than absolute threshold?'' At the New York meeting, we had two winners, Alexander Streeter for his talk ''Hybrid feedforward-feedback active noise control for circumaural headsets'' and Brian Monson for his talk ''Active noise control of small axial cooling fans.'' New Fellows of the Society announced include George Augspurger, Kenneth Cunefare, Timothy Foulkes, Ronald Freiheit, Brigitte Schulte-Fortkamp, Jian Kang, Keith Wilson, and Eric Woods. The winner of the 2004 Helmholtz-Rayleigh Interdisciplinary Silver Medal is David Lubman for his ''work in noise and standards and for contributions to architectural and archeological acoustics.'' It can be noted that David, along with Lou Sutherland (the winner of the 2002 Silver Medal in Noise), have continued to champion efforts to improve classroom acoustics; at the New York meeting, they organized an informal session ''Brainstorming Strategies to Achieve Better Classroom Acoustics for America's Schools.'' And finally, we applaud the awarding of the President's National Medal of Science, one of the highest honors bestowed on scientists in the United States, to Leo L. Beranek

A special outcome of the New York meeting was the ASA 75th Anniversary Book, edited by Henry Bass and William Cavanaugh. Have a look at Chapter 10 to see the impressive contributions made by the ASA in the field of noise and its control. Leo Beranek and William Lang did a great job pulling together 75 years of information and writing the chapter.

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 is Brian Anderson. The Noise web page (http://www.nonoise.org/quietnet/tcn/) is maintained by Les Blomberg. Ken

Cunefare is the coordinator for the Noise Young Presenter Awards (with Ralph Muehleisen ably substituting at the New York 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

Well, another successful year has gone by for the Physical Acoustics Technical Committee (PATC). Although funding issues continue to dominate discussions, our sessions at ASA meetings continue to be filled with interesting speakers, important topics, and attentive audiences.

The Fall meeting was held in Austin, Texas. The big special session for Physical Acoustics (and co-sponsored by Biomedical Ultrasound/Bioresponse to Vibration) was the lifetime achievement session to honor David Blackstock for his work in nonlinear acoustics, organized by Mark Hamilton. Needless to say, this session was a huge success. The room was filled past capacity level, from the beginning to the end of the session. We want to give special thanks to the speakers, who put such a great effort into making this session so wonderful. From pictures of David, to his impact on nonlinear acoustics, it was a great way to introduce the work of a living legend in Physical Acoustics to the audience. The consensus is that Physical Acoustics sould continue to celebrate the works of ''living legends'' with more celebration sessions. Suggestions for such celebration sessions should be directed to a member of the technical committee.

In addition to the Blackstock special session, Physical Acoustics held sessions in thermoacoustics and resonators, bubbles and particles, outdoor propagation, and acoustic landmine detection. All sessions were well-attended. Special thanks go to the organizers, Technical Program Organizing Meeting representatives, and chairs. Thank you for a job well done!

The ASA celebrated its 75th year anniversary meeting in New York. Because of the anticipated number of people attending, the ASA asked each technical committee to limit the number of special sessions. Physical Acoustics held two special sessions. The Robert E. Apfel memorial session was organized by Larry Crum and Ron Roy, with help by many of Apfel's students and colleagues. There was obvious respect and admiration for Bob Apfel, the person, and his work. Nancy Apfel was in attendance, and the room was filled to capacity. The speakers made up a who's who in Acoustics: Lawrence Crum, Andrea Prosperetti, Ronald Roy, Christy Holland, Carr Everbach, Werner Lautherborn, and Phil Martson, to name just a few. One of Bob's last great efforts, starting *Acoustics Research Letters Online (ARLO)*, is now becoming a strong adjunct to JASA.

Jim Sabatier organized a special session entitled ''Recent advances in buried landmine detection.'' It was a huge success, covering a full day of invited and contributed talks. The invited talks were especially interesting, even for audience members not familiar with the field. It has been suggested that sessions such as buried landmine detection might be suitable as ''topical meetings,'' whereby the subject is discussed at length by a few experts, with moderated panel discussions to address issues and answer questions from the audience. Topical meeting sessions such as SonoFusion have been very successful in the past. If you are interested in having a topical meeting, please forward your request to a technical committee member.

Because of the special circumstances for the 75th year anniversary, Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration shared their open meeting on Thursday night. The meeting went well, and it was decided to combine the two open meetings again, possibly during the Hawaii meeting. Thanks to the students who brought in refreshoments!

Physical Acoustics takes special pride in acknowledging our award recipients: Phil Martson was awarded the Silver Medal in Physical Acoustics at the Austin meeting. Phil joins an elite list of innovators and achievers who have made Physical Acoustics a strong component of the ASA. New Fellow certificates were presented to Steve Kargl and Tom Matula at the Austin meeting. At the New York meeting, Mike Bailey was awarded the R. Bruce Lindsay Award. Mike received his award during a special celebration day award ceremony on Wednesday morning. The previous evening, Robin Cleveland and David Chambers were presented Fellow certificates. Congratulations to our new Fellows, Lindsay Award recipient and Silver Medal award recipient.

I would like to take time to acknowledge the Physical Acoustics Summer School (PASS), which is held biannually, most recently 20–27 June. It brings together experts in various fields related to physical acoustics. Topics taught at this year's summer school included nonlinear acoustics, solid state properties, transducers, measurements, bubbles, medical acoustics, molecular acoustics, and thermoacoustics, all geared towards graduate students. Not only do the students learn something about other fields, they also meet other students (future colleagues!) and learn something of each others research.

I wish to also 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 the ASA Committee on Standards (ASACOS). Associate editors also deserve accolades; their hard work helps keep JASA the premier journal for physical acoustics publications. Current and recent members serving as associate editors include Sean Wu, Mark Hamilton, Martin Ochmann, Richard Raspet, Ron Roy, Andrew Szeri, and Louis Sutherland. We also want to acknowledge associate editors of ARLO, including Mark Hamilton, Steve Kargl, Tom Matula, and Keith Wilson.

Finally, it is a pleasure to acknowledge PA members who are active in leadership roles in the ASA. Mark Hamilton is vice-president of the ASA, and the editor-in-chief is Allan Pierce. Members of the Executive Council include Anthony Atchley, Ron Roy, and Vic Sparrow. Thanks to everyone who helps make the PATC a joy!
THOMAS J. MATULA
Chair

Psychological and Physiological Acoustics

Reflecting P&P's strategy of emphasizing participation at ASA's spring meetings, the 146th Meeting of the Acoustical Society of America, the fall meeting in Austin, was sparsely attended by the members of P&P. In keeping with P&P policy, there will be no special sessions in San Diego, and P&P has asked that calls for papers provided by ASA include a brief explanation of our policy that fall meetings are deemphasized by the members of P&P. In stark contrast, the spring meeting in New York City (NYC) was buzzing with activity. There were nine sessions sponsored by P&P, one of which was co-sponsored by Speech Communication. Of the nine sessions, three were special sessions including one named session. We extend our thanks to the organizers of the special sessions: Marjorie R. Leek and Larry E. Humes (''The Perception of Complex Sounds: Honoring the Contributions of Charles S. Watson''); Sid P. Bacon (''Compression in Hearing''), and Jennifer J. Lentz (''Perceptual Organizational of Sound''). P&P also contributed as co-sponsor of three additional special sessions. Glenis R. Long accomplished the challenge of organizing the P&P sessions for the NYC meeting. Her efforts are doubly appreciated because the NYC meeting was both unusually short and unusually popular, two pressures that made scheduling very difficult.

The P&P open meeting was dominated by a discussion of the new Peer Xpress submission processes. Among the wrinkles discussed were (a) the relative lack of accountability—as an example there is no ''person'' to contact about the assignment of associate editors and (b) the clunky way in which figures must be entered into the system. There was also discussion centered on the high cost levied when JASA papers exceed 12 pages. One complaint was that the equation provided by JASA to estimate page length is too conservative; authors are unnecessarily shortening their papers in an effort to avoid high page charges. The Associate Editors (Kidd, Kohlrausch, Lonsbury-Martin, Shofner, and Viemeister) provided brief reports of their activities over the last year. The P&P representative to the Student Council, Joshua Bernstein, provided an overview of the Student Council meeting, and asked P&P members keep in mind the newly established award for mentorship.

Six new members of the TC were elected: Michael A. Akeroyd, Alain de Cheveigne, Brent W. Edwards, Michael G. Heinz, Elizabeth A. Strickland, and Lynne A. Werner. Their terms begin after the meeting in Spring 2005 (Vancouver). Thanks to Andrew Oxenham, the election chair, and members of the P&P community who so willingly added their names to the ballot. One of P&Ps Associate Editors, Neal F. Viemeister, completed his successful term this year. His efforts on our behalf are greatly appreciated; the work of our associate editors is important but also time consuming and difficult. Three new associate editors, John H. Grose, Robert A. Lutfi, and Andrew J. Oxenham, have been named. Their willingness to contribute, and

the contributions of the other associated editors, is highly valued by the community.

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

VIRGINIA M. RICHARDS
Chair

Signal Processing in Acoustics

The goals of the Signal Processing Technical Committee (SP) include organizing special sessions that bring together the various technical disciplines. In the past year, we have held special sessions with Acoustical Oceanography, Animal Bioacoustics, Architectural Acoustics, Engineering Acoustics, Musical Acoustics, Noise, Physical Acoustics, Speech Communication, Structural Acoustics and Vibration, and Underwater Acoustics. We will try harder to include Education in Acoustics, and Psychological and Physiological Acoustics in the future. We are always on the lookout for new topics that have the potential to bring new ideas and new people into the ASA. We try to schedule our technical committee meetings so that they do not conflict with any of the other technical committees; stop by and volunteer!

At the 146th meeting in Austin, TX, 11 sessions involved SP. The papers for these were sorted by Preston Wilson and Paul Gendron at the Technical Program Organizing Meeting (TPOM). Nine were special sessions. Joe Posey organized ''Signal Processing for Aircraft Noise.'' The invited speakers were Robert P. Dougherty, Daniel Sutliff, Bruce E. Walker, Tony L. Parrott, Jacob Klos, and Stephen A. Rizzi. George and Juliette Ioup organized the notably successful session ''Detection and Classification in Acoustics'' that was jointly sponsored by Underwater Acoustics, Animal Bioacoustics, Noise, and Engineering Acoustics. This remarkable special session filled four sessions with an exceptionally wide variety of invited and contributed papers. The invited speakers were Leon H. Sibul, Nicholas C. Makris, Lisa Pflug, Susan Jarvis, Dale Ellis, Hynek Hermansky, Stephen Greineder, Paul M. Baggenstoss, Whitlow Au, Adele Doser, and Patrick J. Loughlin. We thank George and Juliette for their good work bringing all of these groups together to hear the latest research in a common signal-processing problem.

The winner of the best student paper award was David Gerhard, of the Department of Computer Science, University of Regina, Regina, SK, Canada, for his presentation of ''Automatic computational models of acoustical category features: Talking versus singing,'' which was given in the special session ''Detection and Classification in Acoustics.'' There were seven entries. After being told that he was a winner, David Gerhard wrote to us that ''I would like to thank the ASA and all individuals involved in the presentation of this award. It is very respectable of the ASA to encourage participation of younger researchers such as myself. This award will encourage me to participate more fully in the ASA.''

The prize for the Student Challenge problem was awarded to Joseph Edwards. He is a doctoral candidate in the Department of Ocean Engineering at MIT. His research focuses on the signal processing issues arising from the use of multiple autonomous underwater vehicles (AUVs) for mine hunting in shallow waters. His primary interests are in sensor-adaptive behaviors and multi-sensor fusion for system optimization. He received the BSME and MSME degrees from Virginia Tech and has worked as an acoustic engineer specializing in submarine radiated noise suppression. His submission was one of ten from the USA, China, Chile, and Brazil. The quality of the submissions was very high. Jose Diaz and I would like to thank all of the student contributors. The problem dealt with the estimation of the frequency of a pure tone over a limited interval of time in the presence of noise. Brian Ferguson and Paul Gendron have volunteered to run the next Student Challenge problem, which is planned to appear on the website within the next year.

The 147th meeting in New York, NY, was very busy with many papers, social events, and celebrations marking the 75th anniversary of the founding of the Acoustical Society of America. We thank Geoff Edelson and Jim Preisig for attending the Technical Program Organizing Meeting (TPOM) and sorting all the signal processing papers. Signal Processing in Acoustics participated in nine sessions. Jeffrey Ketterling from Biomedical Ultrasound/Bioresponse to Vibration organized the session ''High Frequency Imaging.'' James Preisig organized ''Acoustic Communications—Space/Time Coding, Modulation, Processing and Propagation Effects'' jointly with Underwater Acoustics. The invited speakers were Arthur B. Baggeroer and Milica Stojanovic. Max Deffenbaugh and Alan W. Meyer organized ''Inverse Problems in Seismic Processing'' jointly with Acoustical Oceanography. The invited speakers were Leon H. Sibul, D. B. Harris, Robert Uhrhammer, James Berryman, Sean K. Lehman, and Michael E. Farrell. John Impagliazzo and Brian Ferguson organized ''Advances in Sonar and Imaging Techniques Including Interferometric, Synthetic and Tomographic Apertures'' jointly with Underwater Acoustics. The invited speakers were John Impagliazzo and Ron J. Wyber. Ning Xiang organized the Distinguished Lecture by Jens Blauert, ''Communication Acoustics,'' that was well attended. Jens Blauert is a well-known expert in binaural technology, architectural acoustics, noise engineering, product-sound design, speech technology, virtual environments, and telepresence. We especially thank Ning Xiang for organizing this very successful event.

The Signal Processing in Acoustics Technical Committee was well represented by Ning Xiang who presented ''Future of signal processing in acoustics'' at the Celebration Day program on Wednesday morning.

At this meeting, David Chambers was made a Fellow of the ASA. We all congratulate David on this achievement. We encourage other members of the Signal Processing Technical Committee to look into the process for becoming a fellow.

Stan Ehrlich has stepped down as the Signal Processing in Acoustics representative to the Awards and Medals committee. Leon Sibul is replacing him. We thank Stan for his years of service and Leon for volunteering.

Our representative to the Student Council is Brian Monson, a graduate student at Brigham Young University. Student Council is currently helping to organize Student Chapters at various locations. One is being started at the University of Nebraska and another is under consideration at the University of Rochester.

The winner of the best student paper was Mark Grammann, a graduate student from Pennsylvania State University, whose paper ''Blind deconvolution for non-stationary acoustical systems'' was the best of nine entries. Dave Chambers is in charge of judging our student papers. I hope that we continue to support Dave by volunteering to judge student presenters. Dave gives each volunteer a straightforward form to be filled out while listening to each student.

Paul Hursky and Sean K. Lehman volunteered to run the Gallery of Acoustics. We thank Preston Wilson for his previous efforts at this big job. The Gallery of Acoustics has the potential to make a big impact if the results are transmitted effectively to the wider audience. We wish Paul and Sean luck at this big job. The Gallery of Acoustics is scheduled to appear next at the Vancouver meeting in the Spring of 2005. We encourage everyone to consider interesting and scientifically evocative presentations of their data in visual stills, movies, or sound.

We thank our Associate Editor of the *Journal of the Acoustical Society of America*, Edmund J. Sullivan, and our Associate Editor for the *Acoustics Research Letters Online* (ARLO), James V. Candy. This is on-line medium offers the opportunity to publish color images, movies, sounds, and various other multi-media presentations with a rapid turn-around of a few weeks. The journal is accessible for free on the internet and is included by ISI in its citation index. The article must be short with a 6-page limit. These articles may complement a normal full-length JASA article. There is an ongoing problem of finding referees and receiving timely reviews. This is a sad state of affairs. If anyone thinks of a remedy, please discuss it with Ed, Jim, or myself.

As the new chairman, I would like to thank James Candy, the previous chair, and David Havelock, whose efforts were key in the inception and incorporation of the Signal Processing in Acoustics Technical Committee. They continue to be a source of insight and knowledge for me. Thanks to David Evans for his work with coordinating issues dealing with Standards. I also thank Alan Meyer for taking minutes at the technical committee meetings and preparing the minutes electronically. Lastly, I thank all the other volunteers, invited speakers, contributors, and attendees. I find serving as a chair to be a pleasure with so many enthusiastic members.

CHARLES F. GAUMOND
Chair

Speech Communication

The Speech Technical Committee (TC) supports the activities, meetings, publications, etc. for the largest technical area in the Society. The current members of the Committee are Jean Andruski, Lynne Bernstein, Fredericka Bell-Berti, Catherine T. Best, Ocke-Schwen Bohn, Suzanne E. Boyce, Ann Bradlow, Dani Byrd, Rene Carre, Robert D. Frisina, Bruce Gerratt, James Hillenbrand, Lori Holt, Hideki Kawahara, Jody E. Kreiman, Qiguang Lin, Diane L. Meador, Benjamin Munson, Shrikanth S. Narayanan, John J. Ohala, Dwayne Paschall, Peggy Nelson, Astrid Schmidt-Nielsen, Samuel A. Seddoh, Elaine T. Stathopoulos, Brad Story, Winifred Strange, Mario A. Svirsky, Emily A. Tobey, and Gary G. Weismer. Peter Assmann, Randy Diehl, Kenneth Grant, Anders Lofqvist, and Douglas O'Shaughnessy serve as our Associate Editors. Diane Kewley-Port helped facilitate the use of IPA fonts with the new Peer XPress online submission system. Other *ex-officio* members of the Committee are Tessa Bent and Maureen L. Stone. Thanks to our paper sorters who arrange the technical programs, Emily Tobey and Peter MacNeilage in Austin, and Freddie Bell-Berti and Laura Koenig in New York. Special thanks to Speech TC member Freddie Bell-Berti who served on the New York local organizing committee and the 75th Anniversary Committee. We gratefully acknowledge the organization and hard work of Patricia Kuhl who co-chaired the 75th Anniversary Committee with Leo Beranek. The ''Future Challenges in Speech Communications'' talk given by Dani Byrd at the City Center was exciting and insightful.

Maureen Stone has been our representative to the Membership Committee which makes decisions regarding Fellows of the Society. At the New York meeting we welcomed Abeer Alwan, Li Deng, and Alain de Cheveigne as Fellows of the Society. We are also happy to announce that Maureen Stone is the incoming Chair of the Speech Communications Technical Committee, and Abeer Alwan will become our representative to Membership.

Stetson Scholarship for Research

the Raymond H. Stetson Scholarship in Phonetics and Speech Production is administered by Speech TC. This annual Stetson Scholarship supports a student in graduate study in scientific areas related to the field of phonetics and motor speech production. This year's award of $3000 went to Marion Dohen (Institut de la Communication Parleé, Univ. Stendhal, Grenoble, France) Honorable mention was given to Kathryn Connaghan (Speech & Hearing Science, University of Washington). James Hillenbrand chaired the Stetson subcommittee composed of Betty Tuller, Laura Koenig and Jody Kreiman.

Student Activities

The Committee sponsors two student activities, a competition with a cash award for best oral or poster presentation and in Austin a well-attended evening reception with other technical committees. Over 24 students in Austin and 50 students in New York were considered for the Best Student Paper Award. In Austin, Carolyn Richie at Indiana University was awarded the first prize of $300 and Madhavi Basu at Purdue University received the second prize of $200 for the Best Paper Award. The coordinators for the judging in Austin were Helen Hanson and Suzanne Boyce, while in New York Ben Munson joined the coordinator team. Tessa Bent at Northwestern University is our Student Council representative. Their activities have now resulted in a very useful web site for students in the Society: http://www.acosoc.org/student/

Special Sessions

Speech Communication has sponsored many interesting special sessions during the past year. In Austin, Freddie Bell-Berti (SC) and Susan Blaeser (Standards) organized a session called ''Speech Intelligibility and AAC Devices.'' The primary purpose of this session was to raise interest in developing a new standard for speech synthesis. This session succeeded because Ann Syrdal and Corine Bickley have now formed a standards working group on Intelligibility in Text-to-Speech Synthesis that had its first meeting in New York. Other special sessions in Austin included ''Issues in Similarity and Distinctiveness'' organized by Lynne Bernstein and Pat Keating as well as ''Statistical Patterns in Speech'' organized by Lori Holt, Andrew Lotto, and Randy Diehl. As part of a workshop in Austin, Speech helped support a session on ''Forensic Acoustics'' chaired by Jack Randorff and Peter Ladefoged.

In New York, Anders Lofqvist and Laura Koenig organized as special session, ''Forty Years of Voice Onset Time.'' This session recognized the contributions of Arthur Abramson and Leigh Lisker, and was followed by a reception in their honor. Linda Polka and Megha Sundara organized ''Speech perception and production in bilingual acquisition.'' A joint session with Psychological and Physiological and Acoustics, ''Complex sound perception: Honoring the contributions of Charles Watson,'' was organized by Donna Neff, Marjorie Leek, Larry Humes, and Diane Kewley-Port.

DIANE KEWLEY-PORT
Chair 2001–2004

Underwater Acoustics

I want to first off thank John Perkins, who served admirably as Chair of the Underwater Acoustics Technical Committee (UWTC) for the three years up to and including the Nashville meeting (Spring 2003). For the Austin meeting (Fall 2003), Christopher Tiemann and David Knobles effectively represented the UWTC at the technical program organizing meeting (TPOM). The UWTC sponsored two special sessions: one organized by Grant Deane and entitled ''The Acoustics of Bubbles in the Marine Boundary Layer,'' which was co-sponsored by Acoustical Oceanography (AO) and Physical Acoustics, the other organized by Paul Hines and Dale Hutt and entitled ''Gradient Array Acoustics,'' which was co-sponsored by Engineering Acoustics (EA) and Signal Processing (SP). UW was also involved in co-sponsoring special sessions with six other technical committees, including the ''Aubrey L. Anderson Memorial Session on Gas-Bearing Sediments'' sponsored by AO. There were 22 papers submitted to the student paper competition for the Austin meeting under the auspices of the UWTC. Geoff Edelmann was awarded first prize for ''The stability of the time-reversal process in a fluctuating ocean,'' and Donald Outing received second prize for ''Geoacoustic characterization of a range dependent environment using towed array data.'' The chair of the UWTC judges committee was Peter Dahl. In Austin, UWTC members David Dowling, Peter Gerstoft, Hee-Chun Song, and Evan Westwood were recognized as new Fellows of the ASA.

Some changes in effect by the Austin meeting include the appointments of Eric Thorsos as UWTC representative to the Medals and Awards Committee, and Henrik Schmidt as UWTC representative to the Membership Committee; Joseph Zalesak remains our representative to the ASA Committee on Standards. After Austin, Tom Weber (ARL-PSU) was appointed by the chair as the student representative to the Student Council, replacing the graduating Geoff Edelmann (MPL-SIO). As promised at the UWTC meeting in Austin, a new web site for the UWTC was launched.

By all measures the New York City 75th anniversary (Spring 2004) meeting was a complete success, with one highlight being the presentation of the Gold Medal to Chester McKinney, a founding member of the UWTC. James Preisig and Natalia Sidorovskaia represented the UWTC at the Technical Program Organizing Meeting (TPOM), and handled well the scheduling challenges unique to this meeting.

The UWTC sponsored a special session entitled: ''Autonomous Underwater Vehicle Acoustics,'' organized by Henrik Schmidt and Thomas Howarth, and co-sponsored by Engineering Acoustics (EA). UW was also co-sponsor of three special sessions involving Acoustical Oceanography, Signal Processing, and Engineering Acoustics, and one special session involving all technical committees of the ASA, in honor of Leo Beranek.

The UWTC had 13 papers presented in the student paper competition for the New York meeting. Jed C. Wilbur was awarded first prize for ''Development of a submersible baffled impedance tube for characterization of bubbly fluids,'' and Shawn F. Johnson, received second prize for ''Resolution cell size effects on the statistics of seafloor.'' The chair of the judges committee was Kyle Becker. The ASA's 75th Anniversary compendium of history lectures was completed in time for distribution for the New York meeting. This included the UWTC history chapter entitled ''A Brief History of Underwater Acoustics'' by Ralph Goodman, with Introduction by Peter Dahl. A look ahead toward the future was provided by Purnima Ratilal, with her presentation in New York entitled ''Ocean acoustic exploration in 2029.'' In New York, UWTC members Gary Heald and Daniel Rouseff were

recognized as new Fellows of the ASA.
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.

2004

16–18 Sept.	Twelfth Annual Conference on the Management of the Tinnitus Patient. For professionals and tinnitus patients. The University of Iowa, Iowa City, IA [Contact Rich Tyler, 319-356-2471; E-mail: rich-tyler@uiowa.edu, WWW: www.uihealthcare.com/depts/med/otolaryngology/conferences/index.htm].
20–24 Sept.	ACTIVE 2004—The 2004 International Symposium on Active Control of Sound and Vibration, Williamsburg, VA [INCE Business Office, Iowa State Univ., 212 Marston Hall, IA 50011-2153; Fax: 515 294 3528; E-mail:ibo@ince.org; WWW: inceusa.org]
4–8 Oct	Short Course on Underwater Acoustics, State College PA [www.outreach.psu.edu/C&I/UnderwaterAcoustics]
15–19 Nov.	148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2005

16–19 May	Society of Automotive Engineering Noise & Vibration Conference, Traverse City, MI [Patti Kreh, SAE International, 755 W. Big Beaver Rd., Ste. 1600, Troy, MI 48084; Tel.: 248-273-2474; E-mail: pkreh@sae.org].
16–20 May	149th Meeting of the Acoustical Society of America, Vancouver, Canada [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].
18–22 July	17th International Symposium on Nonlinear Acoustics, State College PA [Anthony Atchley, The Pennsylvania State Univ., 217 Applied Science Research Lab. Bldg., University Park, PA 16802; Tel.: 814-865-6364; E-mail: INSA17@outreach.psu.edu; www: http://www.outreach.psu.edu/c&i/ISNA17/].
17–21 Oct.	Joint ASA 150th Meeting of the Acoustical Society of America and NOISE CON 2005, 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].

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 2005.

Technical Committees 2004–2005

Acoustical Oceanography

	Term to
N. Ross Chapman, *Chair* to 2004	2007
Kyle M. Becker	2007
Grant B. Deane	2007
Christopher Feuillade	2007
Peter Gerstoft	2007
Oleg A. Godin	2007
John K. Horne	2007
Bruce M. Howe	2007
Anthony P. Lyons	2007
Ralph A. Stephen	2007
Kathleen E. Wage	2007
Peter F. Worcester	2007
Mohsen Badiey	2006
Paul E. Chandler	2006
Dezhang Chu	2006

Kerry W. Commander	2006
Christian de Moustier	2006
Stan E. Dosso	2006
Andone C. Lavery	2006
Zoi-Heleni Michalopoulou	2006
Jeffrey A. Nystuen	2006
David R. Palmer	2006
Simon D. Richards	2006
Martin Siderius	2006
Daniela Di Iorio	2005
Matthew A. Dzieciuch	2005
Gary J. Heald	2005
Daniel L. Hutt	2005
Timothy G. Leighton	2005
James H. Miller	2005
Carlos J. Robinson	2005
Daniel Rouseff	2005
Emmanuel K. Skarsoulis	2005
Jerome A. Smith	2005
Gang Yuan	2005

Ex officio:
Shira L. Broschat, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
David R. Dowling, Associate Editor of JASA
Kenneth G. Foote, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Alexandra I. Tolstoy, Associate Editor of JASA
James F. Lynch, member of Medals and Awards Committee
Mohsen Badiey, member of Membership Committee
Kerry W. Commander, member of ASACOS

Animal Bioacoustics

	Term to
Andrea M. Simmons, *Chair* to	2006
Sheryl L. Coombs	2007
Richard R. Fay	2007
Edmund R. Gerstein	2007
Seth S. Horowitz	2007
Cynthia F. Moss	2007
James A. Simmons	2007
Jeannette A. Thomas	2007
Ann E. Bowles	2006
John R. Buck	2006
Christopher W. Clark	2006
William C. Cummings	2006
Charles R. Greene	2006
Mardi C. Hastings	2006
D. Vance Holliday	2006
David A. Mann	2006
Sam H. Ridgway	2006
William C. Burgess	2005
James J. Finneran	2005
Adam S. Frankel	2005
Darlene R. Ketten	2005
Larry L. Pater	2005
Peter L. Tyack	2005

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

Architectural Acoustics

	Term to
Lily M. Wang, *Chair* to	2007
Warren E. Blazier	2007
Joseph F. Bridger	2007
Norm Broner	2007
Bennett M. Brooks	2007
Steven M. Brown	2007
Todd A. Busch	2007
Richard H. Campbell	2007
F. M. del Solar Dorrego	2007
Erin L. Dugan	2007
M. David Egan	2007
Jesse J. Ehnert	2007
Donna A. Ellis	2007
Michael Ermann	2007
Adam R. Foxwell	2007
Ronald R. Freiheit	2007
Klaus Genuit	2007
Matthew V. Golden	2007
Kenneth W. Good, Jr.	2007
Brad N. Gover	2007
Mark A. Holden	2007
K. Anthony Hoover	2007
Dana S. Hougland	2007
Jerald R. Hyde	2007
David W. Kahn	2007
Martha M. Larson	2007
Gary S. Madaras	2007
Benjamin E. Markham	2007
Charles T. Moritz	2007
David L. Moyer	2007
Edward T. Nykaza	2007
Paul B. Ostergaard	2007
Dennis A. Paoletti	2007
Stephen W. Payne	2007
Benjamin C. Seep	2007
Neil A. Shaw	2007
Ryan G. Sieler	2007
Rose Mary Su	2007
Jeff D. Szymanski	2007
Richard H. Talaske	2007
Michelle Vigeant	2007
Ewart A. Wetherill	2007
George E. Winzer	2007
Michael R. Yantis	2007
Nils-Ake Andersson	2006
Walter Beamer	2006
Leo L. Beranek	2006
Sergio Beristain	2006
Sylvio R. Bistafa	2006
Leslie D. Blomberg	2006
Erica Bowden	2006
David T. Bradley	2006
David Braslau	2006
Todd L. Brooks	2006
Courtney B. Burroughs	2006
Paul T. Calamia	2006

Alexander U. Case	2006
William J. Cavanaugh	2006
Dan Clayton	2006
Elizabeth A. Cohen	2006
Chris L. Davies	2006
Damian Doria	2006
John Erdreich	2006
Robin S. Glosemeyer	2006
Robert D. Hellweg	2006
Murray R. Hodgson	2006
Ian B. Hoffman	2006
Jin Jeon	2006
James A. Johnson	2006
Bertram Y. Kinzey, Jr.	2006
Iraklis E. Lampropoulos	2006
Jerry G. Lilly	2006
Peter A. Mapp	2006
David E. Marsh	2006
Gregory A. Miller	2006
Hideo Miyazaki	2006
Richard A. Moscoso-Bullon	2006
Matthew A. Nobile	2006
Bruce C. Olson	2006
Richard J. Peppin	2006
Scott D. Pfeiffer	2006
James E. Phillips	2006
Joseph Pope	2006
Roy L. Richards	2006
Carl J. Rosenberg	2006
Kenneth P. Roy	2006
Francisco Ruffa	2006
Melvin L. Saunders	2006
Kevin P. Shepherd	2006
Yasushi Shimizu	2006
Gary W. Siebein	2006
Abigail E. Stefaniw	2006
Christopher A. Storch	2006
Jason E. Summers	2006
Louis C. Sutherland	2006
Jiri Tichy	2006
Nancy S. Timmerman	2006
Brandon D. Tinianov	2006
Gregory C. Tocci	2006
Rendell R. Torres	2006
Alfred C. C. Warnock	2006
George P. Wilson	2006
Christopher N. Blair	2005
John S. Bradley	2005
Christopher N. Brooks	2005
Angelo J. Campanella	2005
Quinsan Ciao	2005
Robert C. Coffeen	2005
Peter D'Antonio	2005
Felicia M. Doggett	2005
Timothy J. Foulkes	2005
Richard D. Godfrey	2005
Clare M. Hurtgen	2005
J. Christopher Jaffe	2005
Mendel Kleiner	2005
David Lubman	2005
Ralph T. Muehleisen	2005
Michael T. Nixon	2005
Jack E. Randorff	2005
H. Stanley Roller	2005
B. Schulte-Fortkamp	2005
Noral D. Stewart	2005
Ning Xiang	2005

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

Biomedical Ultrasound/Bioresponse to Vibration

	Term to
Robin O. Cleveland, *Chair* to 2005	2005
Constantin-C. Coussios	2007
Diane Dalecki	2007
J. Brian Fowlkes	2007
Kullervo H. Hynynen	2007
T. Douglas Mast	2007
Thomas J. Matula	2007
Robert J. McGough	2007
Douglas L. Miller	2007
Ronald A. Roy	2007
Thomas J. Royston	2007
Kendall R. Waters	2007
Pei Zhong	2007
Shira L. Broschat	2006
Ibrahim M. Hallaj	2006
Joie P. Jones	2006
Inder R. Makin	2006
Michael V. Scanlon	2006
Thomas L. Szabo	2006
Janet M. Weisenberger	2006
Junru Wu	2006
Suk Wang Yoon	2006
John S. Allen	2005
Michael R. Bailey	2005
Charles C. Church	2005
Gregory Clement	2005
Floyd Dunn	2005
E. Carr Everbach	2005
Christy K. Holland	2005
R. Glynn Holt	2005
James A. McAteer	2005
Wesley L. Nyborg	2005
Mark E. Schafer	2005

Ex officio:
Charles C. Church, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
T. Douglas Mast, Associate Editor of JASA
William P. Shofner, Associate Editor of JASA
Andrew Oxenham, Associate Editor of JASA
Lawrence A. Crum, member of the Medals and Awards Committee
Anthony J. Brammer, member of the Membership Committee
E. Carr Everbach, member of ASACOS

Engineering Acoustics

	Term to
Kim C. Benjamin, *Chair* to 2006	2006
Mahlon D. Burkhard	2007
James Christoff	2007
Fernando Garcia-Osuna	2007

Charles S. Hayden	2007
Dennis F. Jones	2007
Jan F. Lindberg	2007
Yushieh Ma	2007
Elizabeth A. McLaughlin	2007
Alan Powell	2007
Roger T. Richards	2007
Kenneth D. Rolt	2007
Neil A. Shaw	2007
James F. Tressler	2007
Kenneth Walsh	2007
Stanley L. Ehrlich	2006
Gary W. Elko	2006
Robert D. Finch	2006
Guillermo C. Gaunaurd	2006
Thomas R. Howarth	2006
Dehua Huang	2006
Sung Hwan Ko	2006
Theodore J. Mapes	2006
Victor Nedzelnitsky	2006
James M. Powers	2006
P. K. Raju	2006
Yongrae Roh	2006
Ahmet Selamet	2006
Stephen C. Thompson	2006
James E. West	2006
George S. K. Wong	2006
Steven R. Baker	2005
David A. Brown	2005
Stephen C. Butler	2005
Robert D. Corsaro	2005
Stephen E. Forsythe	2005
Brian H. Houston	2005
W. Jack Hughes	2005
Robert M. Koch	2005
L. Dwight Luker	2005
Arnie L. Van Buren	2005
Daniel M. Warren	2005
Joseph F. Zalesak	2005

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

Musical Acoustics

	Term to
James P. Cottingham, *Chair* to	2005
R. Dean Ayers	2007
Judith C. Brown	2007
Courtney B. Burroughs	2007
John R. Buschert	2007
George F. Emerson	2007
Thomas M. Huber	2007
Bozena Kostek	2007
Barry Larkin	2007
Daniel O. Ludwigsen	2007
Thomas D. Rossing	2007
David B. Sharp	2007
Julius O. Smith	2007
William J. Strong	2007
Anders G. Askenfelt	2006
James W. Beauchamp	2006
Xavier Boutillon	2006
Jonas Braasch	2006
Murray D. Campbell	2006
Rene E. Causse	2006
Antoine J. Chaigne	2006
Nicholas J. Giordano	2006
J. M. Harrison	2006
William M. Hartmann	2006
James M. Pyne	2006
Daniel A. Russell	2006
Punita G. Singh	2006
Paul A. Wheeler	2006
Shigeru Yoshikawa	2006
George A. Bissinger	2005
Annabel J. Cohen	2005
Diana Deutsch	2005
Roger J. Hansen	2005
Uwe J. Hansen	2005
Peter L. Hoekje	2005
James H. Irwin	2005
Ian M. Lindevald	2005
Stephen E. McAdams	2005
Gary P. Scavone	2005
Gabriel Weinreich	2005

Ex officio:
Neville H. Fletcher, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Allan J. Zuckerwar, Associate Editor of JASA
to be appointed, member of Medals and Awards Committee
Uwe J. Hansen, member of Membership Committee
Ian M. Lindevald, member of ASACOS

Noise

	Term to
Michael R. Stinson, *Chair* to	2006
Martin Alexander	2007
Brian E. Anderson	2007
Keith Attenborough	2007
John P. Barry	2007
Leo L. Beranek	2007
Arno S. Bommer	2007
James O. Buntin	2007
John C. Burgess	2007
Jim R. Cummins	2007
Kenneth A. Cunefare	2007
Paul R. Donavan	2007
Ronald R. Freiheit	2007
Klaus Genuit	2007
Michael L. Gross	2007
David C. Haser	2007
Gerald C. Lauchle	2007
George C. Maling	2007
Thomas R. Norris	2007
John P. Seiler	2007
Noral D. Stewart	2007
Louis C. Sutherland	2007
Jiri Tichy	2007
D. Keith Wilson	2007
Ning Xiang	2007
Jorge P. Arenas	2006
Sergio Beristain	2006
Leslie D. Blomberg	2006
Bennett M. Brooks	2006
Robert D. Bruce	2006

Angelo J. Campanella	2006
William J. Cavanaugh	2006
Gilles A. Daigle	2006
Patricia Davies	2006
Jesse J. Ehnert	2006
Donna A. Ellis	2006
Tony F. W. Embleton	2006
John Erdreich	2006
David J. Evans	2006
William M. Hartmann	2006
Robert D. Hellweg	2006
David K. Holger	2006
Patrick M. Hurdle	2006
Seong-Woo Kang	2006
William W. Lang	2006
Richard H. Lyon	2006
Alan H. Marsh	2006
Ralph T. Muehleisen	2006
William J. Murphy	2006
Kenneth J. Plotkin	2006
Joseph Pope	2006
Daniel R. Raichel	2006
Brigitte Schulte-Fortkamp	2006
Ben H. Sharp	2006
Kevin P. Shepherd	2006
Scott D. Sommerfeldt	2006
Kerrie G. Standlee	2006
George S.K. Wong	2006
James C. Yu	2006
Elliott H. Berger	2005
Ann E. Bowles	2005
Frank H. Brittain	2005
Steven M. Brown	2005
Mahlon D. Burkhard	2005
Sanford Fidell	2005
Lawrence S. Finegold	2005
Samir N. Y. Gerges	2005
Richard D. Godfrey	2005
Murray R. Hodgson	2005
Jerry G. Lilly	2005
David Lubman	2005
George A. Luz	2005
Luc Mongeau	2005
Michael T. Nixon	2005
Matthew A. Nobile	2005
Richard J. Peppin	2005
Robert A. Putnam	2005
Jack E. Randorff	2005
Stephen I. Roth	2005
Kenneth P. Roy	2005
Paul D. Schomer	2005
Nancy S. Timmerman	2005
Brandon D. Tinianov	2005
Gregory C. Tocci	2005

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

Physical Acoustics

	Term to
Thomas J. Matula, *Chair* to	2005
David T. Blackstock	2007
David A. Brown	2007
John A. Burkhardt	2007
Kerry W. Commander	2007
Bruce C. Denardo	2007
Logan E. Hargrove	2007
D. Kent Lewis	2007
Julian D. Maynard	2007
George Mozurkewich	2007
Lev A. Ostrovsky	2007
Andrea Prosperetti	2007
Neil A. Shaw	2007
Victor W. Sparrow	2007
Richard Stern	2007
Roger M. Waxler	2007
Henry E. Bass	2006
Yves H. Berthelot	2006
James P. Chambers	2006
David I. Havelock	2006
Murray S. Korman	2006
Philip L. Marston	2006
Philip S. Spoor	2006
Larry A. Wilen	2006
D. Keith Wilson	2006
Evgenia A. Zabolotskaya	2006
Robert T. Beyer	2005
Robin O. Cleveland	2005
Lawrence A. Crum	2005
Kenneth E. Gilbert	2005
Robert A. Hiller	2005
R. Glynn Holt	2005
Bart Lipkens	2005
Ralph T. Muehleisen	2005
Harry Simpson	2005
John S. Stroud	2005
Preston S. Wilson	2005

Ex officio:
Paul E. Barbone, Associate Editor of JASA
Yves H. Berthelot, Associate Editor of JASA
Anthony J.M. Davis, Associate Editor of JASA
Floyd Dunn, Associate Editor of JASA
Guillermo C. Gaunaurd, Associate Editor of JASA
Mark F. Hamilton, Associate Editor of JASA
Michael S. Howe, Associate Editor of JASA
Martin Ochmann, Associate Editor of JASA
Richard Raspet, Associate Editor of JASA
Louis C. Sutherland, Associate Editor of JASA
Andrew J. Szeri, Associate Editor of JASA
Lonny L. Thompson, Associate Editor of JASA
Sean F. Wu, Associate Editor of JASA
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

	Term to
Virgina M. Richards, *Chair* to	2005
David A. Eddins	2007
Lawrence L. Feth	2007
Hedwig E. Gockel	2007
Jennifer Lentz	2007
Mario A. Ruggero	2007
Michelle L. Hicks	2006
Lynne Marshall	2006

Christopher J. Plack	2006
Donal G. Sinex	2006
Magdalena Wojtczak	2006
John F. Culling	2005
Torsten Dau	2005
Armin Kohlrausch	2005
Robert A. Lutfi	2005
Christine R. Mason	2005
Brian C.J. Moore	2005

Ex officio:
John H. Grose, Associate Editor of JASA
Gerald D. Kidd, Jr., Associate Editor of JASA
Armin Kohlrausch, Associate Editor of JASA
Brenda L. Lonsbury-Martin, Associate Editor of JASA
Robert A. Lutfi, Associate Editor of JASA
Andrew J. Oxenham, Associate Editor of JASA
William P. Shofner, Associate Editor of JASA
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

	Term to
Charles F. Gaumond, *Chair* to 2006	2006
Max Deffenbaugh	2007
Jose A. Diaz	2007
Alireza A. Dibazar	2007
Gary W. Elko	2007
Alan W. Meyer	2007
Daniel J. Sinder	2007
David C. Swanson	2007
Robert C. Waag	2007
Preston S. Wilson	2007
Lixue Wu	2007
James V. Candy	2006
William M. Carey	2006
Geoffrey S. Edelson	2006
Stanley L. Ehrlich	2006
Brian Ferguson	2006
Paul J. Gendron	2006
Peter Gerstoft	2006
William M. Hartmann	2006
Kevin D. Heaney	2006
William S. Hodgkiss	2006
Paul D. Hursky	2006
John M. Impagliazzo	2006
Hua Lee	2006
Patrick J. Loughlin	2006
Jens M. Meyer	2006
Hassan Namarvar	2006
Joseph W. Posey	2006
James C. Preisig	2006
Brian D. Rapids	2006
Edmund J. Sullivan	2006
David H. Chambers	2005
David J. Evans	2005
Howard A. Gaberson	2005
David I. Havelock	2005
Jean-Pierre Hermand	2005
George E. Ioup	2005
Juliette Ioup	2005
Sean K. Lehman	2005
Zoi-Heleni Michalopoulou	2005
Joseph Pope	2005
Leon H. Sibul	2005
Randall W. Smith	2005
James E. West	2005
Gary R. Wilson	2005
Ning Xiang	2005

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

Speech Communication

	Term to
Maureen L. Stone, *Chair* to 2007	2007
Patrice S. Beddor	2007
Melissa A. Epstein	2007
Robert Allen Fox	2007
Helen M. Hanson	2007
Diane Kewley-Port	2007
Jody E. Kreiman	2007
Andrew J. Lotto	2007
Shrikanth S. Narayanan	2007
Dwayne Paschall	2007
Jean E. Andruski	2006
Lynne E. Bernstein	2006
Ocke-Schwen Bohn	2006
Suzanne E. Boyce	2006
Bruce R. Gerratt	2006
Benjamin R. Munson	2006
Peggy B. Nelson	2006
Brad H. Story	2006
Fredericka Bell-Berti	2005
Ann R. Bradlow	2005
Dani M. Byrd	2005
James M. Hillenbrand	2005
Lori L. Holt	2005
Hideki Kawahara	2005
Qiguang Lin	2005
Mario A. Svirsky	2005
Gary G. Weismer	2005

Ex officio:
Peter F. Assmann, Associate Editor of JASA
Randy L. Diehl, Associate Editor of JASA
Alexander F. Francis, Associate Editor of JASA
Kenneth W. Grant, Associate Editor of JASA
Anders Lofqvist, Associate Editor of JASA
Douglas D. O'Shaughnessy, Associate Editor of JASA
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

	Term to
Courtney B. Burroughs, *Chair* to 2006	2007
Stephen C. Conlon	2007
Linda P. Franzoni	2007
Robert C. Haberman	2007
Rudolph Martinez	2007
Koorosh Naghshineh	2007
Michael F. Shaw	2007
Benjamin A. Bard	2006

Alain C. Berry	2006
Kenneth A. Cunefare	2006
Joseph W. Dickey	2006
David Feit	2006
Allison B. Flatau	2006
Guillermo C. Gaunaurd	2006
Karl Grosh	2006
Sabih I. Hayek	2006
Louis A. Herstein	2006
Francis Kirschner	2006
Jean R. Nicolas	2006
Andrew N. Norris	2006
Sally J. Pardue	2006
Allan D. Pierce	2006
Mauro Pierucci	2006
Scott D. Sommerfeldt	2006
Earl G. Williams	2006
Sean F. Wu	2006
John A. Burkhardt	2005
Kenneth D. Frampton	2005
Sunil Mehta	2005
Martin L. Pollack	2005
Paul J. Remington	2005
Jeffrey S. Vipperman	2005
Richard L. Weaver	2005
Kuangcheng Wu	2005

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

Underwater Acoustics

	Term to
Peter H. Dahl, *Chair* to 2006	2006
David C. Calvo	2007
Jee Woong Choi	2007
Christian P. de Moustier	2007
Stan E. Dosso	2007
Nicholas C. Makris	2007
Zoi-Heleni Michalopoulou	2007
Tracianne B. Neilsen	2007
Robert I. Odom	2007
Marshall H. Orr	2007
Gregory J. Orris	2007
James Preisig	2007
Martin Siderius	2007
Kathleen E. Wage	2007
Jixun Zhou	2007
Ralph N. Baer	2006
Kyle M. Becker	2006
John R. Buck	2006
Nicholas P. Chotiros	2006
Dezhang Chu	2006
David R. Dowling	2006
Steven I. Finette	2006
Roger C. Gauss	2006
Paul C. Hines	2006
Anatoliy N. Ivakin	2006
Finn B. Jensen	2006
William A. Kuperman	2006
John J. McCoy	2006
B. Edward McDonald	2006
Kazuhiko Ohta	2006
John S. Perkins	2006
John R. Preston	2006
Purnima Ratilal	2006
Ralph A. Stephen	2006
Eric I. Thorsos	2006
Brian H. Tracey	2006
Alexander G. Voronovich	2006
Paul B. Baxley	2005
Shira L. Broschat	2005
Douglas H. Cato	2005
Michael D. Collins	2005
Grant B. Deane	2005
Gerald L. D'Spain	2005
Peter Gerstoft	2005
Charles W. Holland	2005
Kevin B. Smith	2005
Hee Chun Song	2005
Dajun Tang	2005
Christopher T. Tindle	2005
Alexandra I. Tolstoy	2005
Stephen N. Wolf	2005

Ex officio:
Shira L. Broschat, Associate Editor of JASA
William M. Carey, Associate Editor of JASA
Kenneth D. Foote, Associate Editor of JASA
William L. Siegmann, Associate Editor of JASA
Ralph A. Stephen, Associate Editor of JASA
Edmund J. Sullivan, Associate Editor of JASA
Alexandra I. Tolstoy, Associate Editor of JASA
Henrik Schmidt, member of Membership Committee
Eric I. Thorsos, member of Medals and Awards Committee
Joseph M. Zalezak, member of ASACOS

Administrative Committees 2004–2005

Archives and History

	Term to
Julian D. Maynard, *Chair* to 2007	2007
Henry E. Bass	2007
David T. Blackstock	2007
E. Carr Everbach	2007
William W. Lang	2007
David L. Moyer	2007
Richard Stern	2007
Rosalie M. Uchanski	2007
Jont B. Allen	2006
Ralph R. Goodman	2006
David I. Havelock	2006
Wesley L. Nyborg	2006
Richard J. Peppin	2006
William J. Strong	2006
Leo L. Beranek	2005
William J. Cavanaugh	2005
Logan E. Hargrove	2005

Audit Committee

	Term to
William A. Yost, *Chair* to 2005	2005
George V. Frisk	2007
Steven M. Brown	2006

Books[+]

	Term to
David L. Bradley, *Chair* to 2005	2005
Stanley L. Chin-Bing	2007
Robert C. Spindel	2007
Jerry H. Ginsberg	2006
Philip L. Marston	2006
Joseph Pope	2006
Robert A. Walkling	2006
Nancy S. McGarr	2005
Jeffrey A. Nystuen	2005
Neal A. Shaw	2005
Emily A. Tobey	2005
Ex officio:	
Allan D. Pierce, Editor-in-Chief	

College of Fellows Steering

	Term to
Janet M. Weisenberger, *Chair* to 2007	2007
Peter G. Cable	2007
M. David Egan	2007
Uwe J. Hansen	2007
Diane Kewley-Port	2007
Thomas D. Rossing	2007
Joseph W. Dickey	2006
Judy R. Dubno	2006
Jiri Tichy	2006
Stanley L. Ehrlich	2005
E. Carr Everbach	2005
Robert D. Frisina	2005
Daniel R. Raichel	2005
William J. Cavanaugh, *ex officio* as past Chair	
Richard H. Lyon, *ex officio* as past Chair	

Education in Acoustics

	Term to
Uwe J. Hansen, *Chair* to 2006	2006
George A. Bissinger	2007
David A. Brown	2007
Robert D. Collier	2007
Corinne M. Darvennes	2007
Margaritis S. Fourakis	2007
Carole E. Gelfer	2007
Daniel O. Ludwigsen	2007
Sharon Y. Manuel	2007
Philip L. Marston	2007
Ralph T. Muehleisen	2007
Andrew A. Piacsek	2007
Daniel R. Raichel	2007
Thomas D. Rossing	2007
Ronald A. Roy	2007
Dawn R. Schuette	2007
Scott D. Sommerfeldt	2007
William Thompson, Jr.	2007
Robert A. Walkling	2007
George S. K. Wong	2007
William A. Ahroon	2006
Takayuki Arai	2006
Anthony A. Atchley	2006
Fredericka Bell-Berti	2006
Suzanne E. Boyce	2006
Robert D. Celmer	2006
Annabel J. Cohen	2006
E. Carr Everbach	2006
Thomas B. Gabrielson	2006
Katherine S. Harris	2006
Elizabeth S. Ivey	2006
Joie P. Jones	2006
Maria B. Mody	2006
Amy T. Neel	2006
P. K. Raju	2006
Deborah M. Rekart	2006
Daniel A. Russell	2006
M. Roman Serbyn	2006
Victor W. Sparrow	2006
Emily A. Tobey	2006
Courtney B. Burroughs	2005
Robin O. Cleveland	2005
Kenneth A. Cunefare	2005
D. Michael Daly	2005
Mary Florentine	2005
Logan E. Hargrove	2005
Mardi C. Hastings	2005
Peter L. Hoekje	2005
Darrell R. Jackson	2005
Michel T. T. Jackson	2005
Murray S. Korman	2005
Luc Mongeau	2005
James M. Sabatier	2005
Neil A. Shaw	2005
Kevin B. Smith	2005
Ralph A. Stephen	2005
James E. West	2005
Wayne M. Wright	2005

Ethics and Grievances

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

Investments

	Term to
Ira Dyer, *Chair* to 2007	2007
Patricia K. Kuhl	2006
Lawrence A. Crum	2005
Richard H. Lyon	2005
David Feit, Treasurer, *ex officio*	

Medals and Awards

		Term to
David L. Bradley, *Chair* to 2005		2005
to be appointed	Musical Acoustics	2007
Fredericka Bell-Berti	Speech Communication	2007
Mahlon D. Burkhard	Engineering Acoustics	2007
James F. Lynch	Acoustical Oceanography	2007
Leon H. Sibul	Signal Processing in Acoustics	2007
Joseph W. Hall	Psychological and Physiological Acoustics	2006

Lawrence A. Crum	Biomedical Ultrasound/ Bioresponse to Vibration	2006
Robert M. Keolian	Physical Acoustics	2006
Nancy S. Timmerman	Noise	2006
Charles R. Greene	Animal Bioacoustics	2005
Mauro Pierucci	Structural Acoustics and Vibration	2005
Eric I. Thorsos	Underwater Acoustics	2005
Ewart A. Wetherill	Architectural Acoustics	2005

Meetings

May 2004–November 2004

Dana S. Hougland, *Chair* to 2005
Damian Doria, Spring 2004, New York
D. Wesley Grantham, Spring 2003, Nashville
Mark F. Hamilton, Vice President
Murray S. Hodgson, Spring 2005, Vancouver
Russell Johnson, Spring 2004, New York
William A. Kuperman, Fall 2004, San Diego
Elaine Moran, ASA Office Manager, *ex officio*
Donna Neff, Vice President-Elect
Peggy B. Nelson, Fall 2005, Minneapolis
Clark Penrod, Fall 2003, Austin
Charles E. Schmid, Executive Director, *ex officio*

November 2004–June 2005

Dana S. Hougland, *Chair* to 2005
Damian Doria, Spring 2004, New York
Mark F. Hamilton, Vice President
Murray D. 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*
Donna Neff, Vice President-Elect
Peggy B. Nelson, Fall 2005, Minneapolis
Clark Penrod, Fall 2003, Austin
Charles E. Schmid, Executive Director, *ex officio*

Membership

		Term to
Peter H. Rogers, *Chair* to 2006		2006
Mohsen Badiey	Acoustical Oceanography	2007
Steven L. Garrett	Physical Acoustics	2007
David I. Havelock	Signal Processing in Acoustics	2007
Henrik Schmidt	Underwater Acoustics	2007
Gregory C. Tocci	Architectural Acoustics	2007
Abeer Alwan	Speech Communication	2006
Leslie R. Bernstein	Psychological and Physiological Acoustics	2006
Uwe J. Hansen	Musical Acoustics	2006
Thomas R. Howarth	Engineering Acoustics	2006
Darleen R. Ketten	Animal Bioacoustics	2006
Anthony J. Brammer	Biomedical Ultrasound/Bioresponse to Vibration	2005
Courtney B. Burroughs	Structural Acoustics and Vibration	2005
Burton G. Hurdle	Foreign Members	2005
John Erdreich	Noise	2005

Public Relations

	Term to
Barbara J. Sotirin, *Chair* to 2006	2006
Kelly J. Benoit-Bird	2007
E. Carr Everbach	2007
Christy K. Holland	2007
Joe W. Posey	2007
Stephen C. Thompson	2007
Paul A. Baxley	2006
Paul Hursky	2006
Elizabeth A. McLaughlin	2006
Jack E. Randorff	2006
Kevin B. Smith	2006
James C. Yu	2006
Gerald L. D'Spain	2005
Geoffrey F. Edelmann	2005
Ellen S. Livingston	2005
James F. Lynch	2005
Duncan E. McGehee	2005
Allan D. Pierce, Editor-in-Chief, *ex officio*	
Elaine Moran, ASA Office Manager, *ex officio*	
Charles E. Schmid, Executive Director, *ex officio*	
Thomas D. Rossing, Echoes Editor, *ex officio*	

Publication Policy

	Term to
Floyd Dunn, *Chair* to 2006	2006
Charles C. Church	2007
Mark F. Hamilton	2007
Mardi C. Hastings	2007
Jont B. Allen	2006
David I. Havelock	2006
James F. Lynch	2005
James H. Miller	2005
Allan J. Zuckerwar	2005
William A. Yost, President-Elect, *ex officio*	
Allan D. Pierce, Editor-in-Chief, *ex officio*	

Regional Chapters

Elizabeth A. McLaughlin, *Chair* to 2005

Angelo J. Campanella	Central Ohio
Robert M. Keolian	Central Pennsylvania
Scott D. Pfeiffer (acting)	Chicago
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
Noral D. Stewart	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Charles F. Gaumond	Washington, DC
Thomas M. Disch	Wisconsin

Uwe J. Hansen, Chair, Education in Acoustics, *ex officio*
David Feit, Treasurer, *ex officio*
Erica Bowden, Student Council representative, *ex officio* -

Rules and Governance

Term to

	Term to
Tony F.W. Embleton, *Chair* to 2005	2005
William J. Cavanaugh	2007
Floyd Dunn	2007
Elaine Moran	2006
Charles E. Schmid	2006
William M. Hartmann	2005
Richard H. Lyon	2005

Prizes and Special Fellowships

	Term to
Wayne M. Wright, *Chair* to 2005	2005
Anthony A. Atchley	2007
Constantine Trahiotis	2007
Fredericka Bell-Berti	2006
James E. West	2006
Uwe J. Hansen	2005

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
Mark F. Hamilton, Chair of ASA Technical Council, *ex officio*
Kerry W. Commander, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
George E. Winzer, Architectural Acoustics
E. Carr Everbach, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Ian M. Lindevald, Musical Acoustics
Richard J. Peppin, Noise
Sameer I. Madanshetty, Physical Acoustics
Peggy B. Nelson, Psychological and Physiological Acoustics
David J. Evans, Signal Processing in Acoustics
Shrikanth S. Narayanan, Speech Communication
Louis A. Herstein, Structural Acoustics and Vibration
Joseph M. Zalezak, Underwater Acoustics
ASA Officers
David Feit, Treasurer, *ex officio*
Charles E. Schmid, Executive Director, *ex officio*
Past Chair of ASACOS (*ex officio*)
Tony F. W. Embleton
Associate Editors for Standards News—JASA (*ex officio*)
Susan B. Blaeser
George S.K. Wong

Student Council

David T. Bradley	Chair and Architectural Acoustics
Brian E. Anderson	Noise
Kelly J. Benoit-Bird	Animal Bioacoustics
Tessa C. Bent	Speech Communication
Joshua G. Bernstein	Psychological and Physiological Acoustics
Erica Bowden	Regional Chapters Committee Liaison
Alireza A. Dibazar	Signal Processing
Geoffrey F. Edelmann	Underwater Acoustics
Wheeler B. Howard	Physical Acoustics
Nicolas Le Dantec	Acoustical Oceanography
Lance L. Locey	Structural Acoustics and Vibration
Alexandra Loubeau	Engineering Acoustics
Andrew C.H. Morrison	Musical Acoustics
Tyrone M. Porter	Biomedical Ultrasound/Bioresponse to Vibration

Tutorials

	Term to
Yves H. Berthelot and Beverly A. Wright, Cochairs to 2005	2005
Ann R. Bradlow	2007
James V. Candy	2007
James P. Chambers	2007
Gerald D. Kidd	2006
Lily M. Wang	2006
Ann E. Bowles	2005
Gary W. Elko	2005

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

Women in Acoustics

	Term to
Peggy B. Nelson, *Chair* to 2006	2006
Sarah Hargus Ferguson	2006
Benjamin R. Munson	2006
Donna L. Neff	2006
Nancy S. Timmerman	2006
David T. Bradley	2006
Chi-Feng Chen	2006
Melissa A. Epstein	2006
Subha Maruvada	2006
Katherine Sawicki	2006
Uwe J. Hansen	2005
Mardi C. Hastings	2005
Lily Wang	2005

Donna Neff, *ex officio* as Vice President-Elect

JASA Editorial Board

	Term to
P. E. Barbone, Physical Acoustics	2007
W. W. L. Au, Bioacoustics–Animal	2007
Y. H. Berthelot, Ultrasonics, Physical Effects of Sound	2007
A. R. Bradlow, Speech Perception	2007
C. C. Church, Bioacoustics	2007
A. J. M. Davis, General Linear Acoustics	2007
F. Dunn, Bioacoustics-Biomedical	2007

A. L. Francis, Speech Perception 2007
J. H. Ginsberg, Structural Acoustics and Vibration 2007
J. H. Grose, Psychological Acoustics 2007
M. F. Hamilton, Nonlinear Acoustics 2007
M. S. Howe, Atmospheric Acoustics and Aeroacoustics 2007
M. Kleiner, Architectural Acoustics 2007
R. A. Lutfi, Psychological Acoustics 2007
M. Ochmann, General Linear Acoustics 2007
A. J. Oxenham, Psychological Acoustics 2007
D. D. O'Shaughnessy, Speech Processing 2007
R. Raspet, Ultrasonics, Physical Effects of Sound 2007
W. L. Siegmann, Underwater Acoustics 2007
L. C. Sutherland, Atmospheric Acoustics and Aeroacoustics 2007
L. L. Thompson, General Linear Acoustics 2007
R. L. Weaver, Structural Acoustics and Vibration 2007
E. G. Williams, Structural Acoustics and Vibration 2007

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

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

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.

Reviews of Acoustical Patents

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
KEVIN P. SHEPHERD, *Mail Stop 463, NASA Langley Research Center, Hampton, Virginia 23681*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,713,762

43.35.Sx ACOUSTIC ABSORPTION ELECTROMAGNETIC RADIATION SENSING WITH SIC

James D. Parsons, assignor to Heetronix
30 March 2004 (Class 250/338.1); filed 16 July 2001

The device employs SiC (silicon carbide) as a radiation sensor at wavelengths below about 10 μm to be used for infrared (IR) sensing within this range. Applications of the device include the sensing of radiation power or energy emitted from a radiation source such as a laser, contactless temperature sensing of bodies heated by IR radiation, and controlled varistors for electrical circuits. A useful acoustic absorption mechanism can be obtained if the device has a single crystal structure with a thickness of at least 200 μm. While there may be some acoustic absorption with thinner samples of SiC, the amount is so small that it has not been previously observed. IR radiation absorbed via acoustic absorption is converted directly into heat energy and causes the resistance of the SiC to increase approximately linearly with the IR energy, as opposed to impurity absorption which causes the SiC resistance to drop.—DRR

6,714,306

43.35.Sx SENSOR FOR OPTICALLY SENSING AIR BORNE ACOUSTIC WAVES

Joseph Lawrence Chovan *et al.*, assignors to Lockheed Martin Corporation
30 March 2004 (Class 356/484); filed 19 May 1986

This patent describes an optical sensor for airborne acoustic waves. The sensor produces mutually coherent optical sampling and reference beams, which may be combined to form an intermediate-frequency carrier. The sampling beam is exposed to the acoustic field in which the acoustic-wave induced density variations occur. The density variations generate a variation in the index of refraction plus a phase modulation of the sampling beam. The phase modulation can be perceived by an optical detector and a phase detector as an electrical signal representative of the acoustic signal.—DRR

6,718,076

43.35.Sx ACOUSTO-OPTIC TUNABLE FILTER WITH SEGMENTED ACOUSTO-OPTIC INTERACTION REGION

Kenneth J. Bures *et al.*, assignors to Unaxis USA, Incorporated
6 April 2004 (Class 385/7); filed 22 March 2002

This acousto-optic tunable filter incorporates a polarization beam splitter, a multisegment interaction region, and a polarization beam combiner. The polarization beam splitter generates two polarized optical signals. The multisegment optical interaction region includes an acoustic wave generator that produces acoustic waves in that region. That region also includes a second optical interaction region that is non-collinear relative to the first and a second acoustic wave generator that produces acoustic waves within the second region. Optical signals that are phase-matched to the acoustic waves are mode-converted in response to the acoustic waves. The polarization beam combiner, within the acousto-optic tunable filter, generates both a mode-converted optical signal and a non-mode-converted optical signal.—DRR

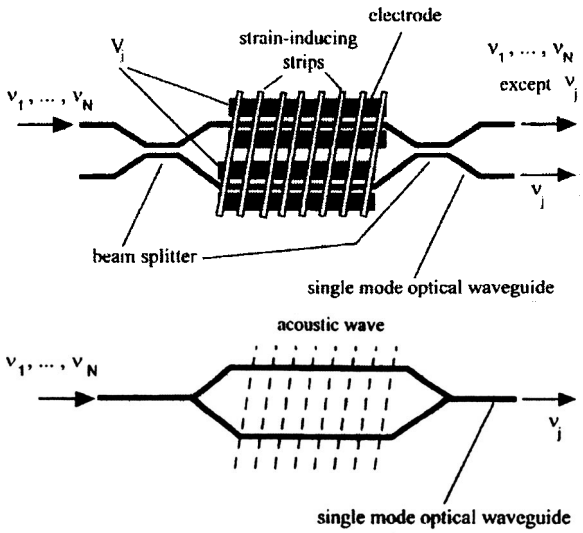
6,721,466

43.35.Sx GUIDED WAVE ELECTROOPTIC AND ACOUSTOOPTIC TUNABLE FILTER APPARATUS AND METHOD

Henry F. Taylor and Ohannes Eknoyan, both of College Station, Texas
13 April 2004 (Class 385/11); filed 14 December 2000

This is a two-port wave tunable filter in a birefringent electrooptic and/or acoustooptic substrate material that includes two three-port, symmetric, Y-branch beam splitters connected by two waveguide sections in which phase-matched polarization coupling occurs. The optical path difference between the beam splitters is one-half of an optical wavelength and the polarization coupling regions between the beam splitters are relatively displaced by an odd integral multiple of half the spatial period of the perturbation responsible for the coupling. In a second embodiment, essentially an acoustooptic tunable filter, polarization coupling arises from a surface acoustic wave and tuning is accomplished by changing the acoustic frequency. Alternatively, four-port electrooptic and acoustooptic tunable filters are created

 © 2004 Acoustical Society of America



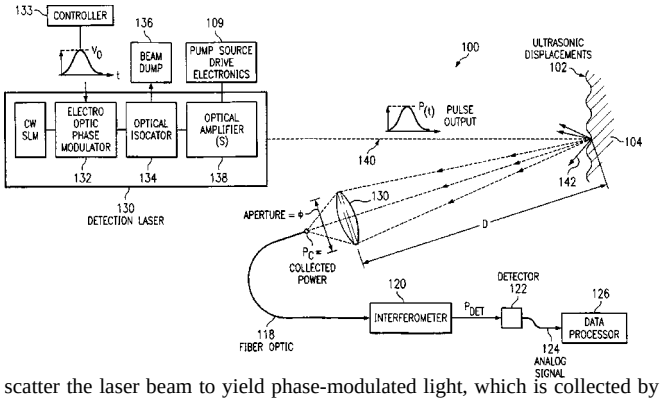
by replacing the three-port beam splitters with four-port directional couplers. In this case, in each of the directional couplers, the splitting ratio for horizontally polarized input plus the splitting ratio for vertically polarized input is substantially equal to unity.—DRR

6,711,954

43.35.Yb METHOD AND APPARATUS FOR IMPROVING THE DYNAMIC RANGE OF LASER DETECTED ULTRASOUND IN ATTENUATIVE MATERIALS

Thomas E. Drake, Jr., assignor to Lockheed Martin Corporation
30 March 2004 (Class 73/655); filed 19 January 2001

This device entails a method for detecting ultrasonic displacements in a material through the use of a time-varying, pulsed laser beam. A laser source produces a beam that is modulated by an assembly placed in the beam's path. The resultant modulated laser has a time-dependent profile. Ultrasonic displacements at a remote target further modulate, reflect, and/or



scatter the laser beam to yield phase-modulated light, which is collected by an optical system. An interferometer coupled to the optical system demodulates the phase-modulated light and yields an output signal representative of the ultrasonics at the remote target.—DRR

6,715,354

43.35.Zc FLAW DETECTION SYSTEM USING ACOUSTIC DOPPLER EFFECT

Shi-Chang Wooh, assignor to Massachusetts Institute of Technology
6 April 2004 (Class 73/598); filed 24 February 1998

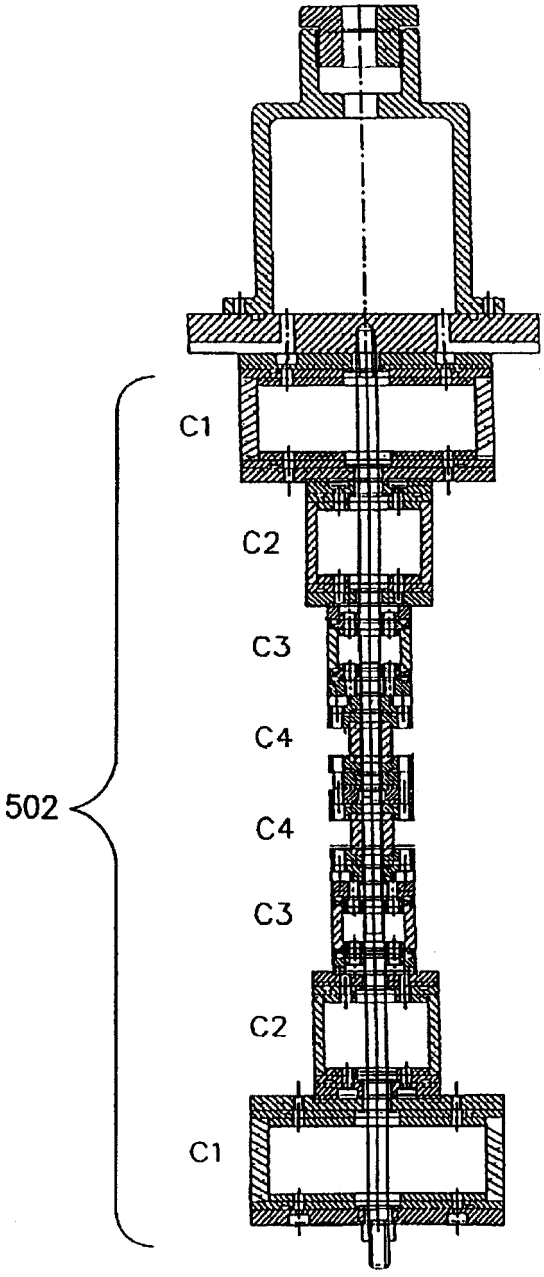
This system is intended for inspection of rails and may be used in other situations where there is relative motion between the system and the medium to be inspected. A noncontacting transducer injects a signal at a given frequency and a second such transducer senses the resulting signal. A processor extracts the Doppler-shifted signal representative of a flaw and compares it to a predetermined threshold.—EEU

6,722,003

43.38.Fx UNDERWATER WIDE-BAND ELECTROACOUSTIC TRANSDUCER AND PACKAGING METHOD

Sheng-Dong Dunn *et al.*, assignors to Chung-Shan Institute of Science and Technology
20 April 2004 (Class 29/25.35); filed 19 March 2003

A sonar projector having a relatively uniform and broad-band transmitting voltage response is realized from a collinear arrangement of piezo ceramic transducers of different sizes such that the various primary resonant frequencies of these transducers are in close frequency juxtaposition to one another. Hence, the ensemble transmitting voltage response displays an approximate band-pass characteristic extending from just below the lowest



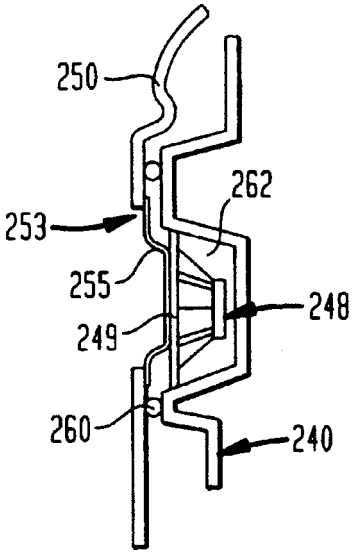
resonant frequency to somewhat above the highest resonant frequency. The figure illustrates a set of four different sized transducers denoted **C1–C4**. These are presumably piezoceramic cylinders although the patent is very vague in describing the individual transducers. This set of transducers **502** is booted with a layer of injection-molded polyurethane, not shown in the figure. Since the concept of stagger-tuning a set of transducers has been practiced since approximately the year one, it is difficult to identify the novelty of this patent.—WT

6,578,658

43.38.Ja ACCOUSTIC CHAMBER INTEGRATED WITH VEHICLE DOOR

Stephen P. Jones *et al.*, assignors to SAI Automotive USA–SAL, Incorporated
17 June 2003 (Class 181/141); filed 11 January 2001

A loudspeaker is integrated into the door of an automobile such that sealed cavity **262** is formed behind speaker diaphragm **249**. Various other



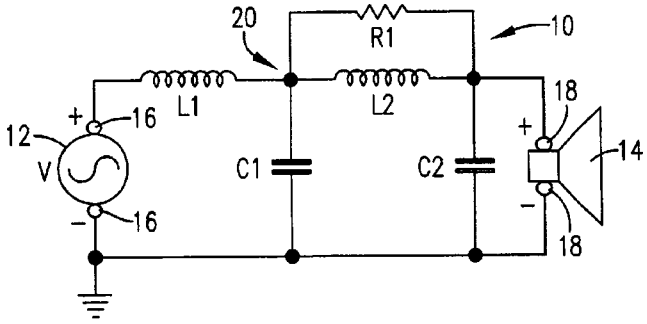
geometric arrangements are described, all of which include a small sealed cavity behind the loudspeaker. Such an arrangement is said to allow a smaller, cheaper loudspeaker to be used.—KPS

6,707,919

43.38.Ja DRIVER CONTROL CIRCUIT

John Koval and Christopher E. Combest, assignors to Multi Service Corporation
16 March 2004 (Class 381/100); filed 20 December 2000

What is shown is the low-pass section of a loudspeaker frequency



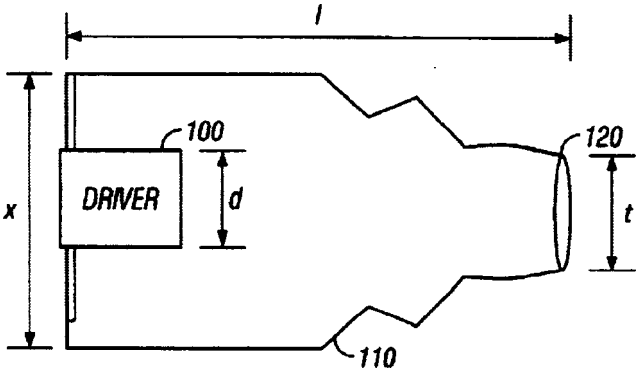
dividing network. What is patented is resistor **R1**. Seventeen claims were granted.—GLA

6,700,984

43.38.Ja NON-LINEARLY TAPERING TRANSMISSION LINE SPEAKERS

Tuomas Holberg and Marcel Gavriliu, assignors to California Institute of Technology
2 March 2004 (Class 381/338); filed 7 December 2000

According to this patent, a tapered transmission line—as opposed to a cylindrical or flared waveguide—can provide better low-frequency performance for a given overall volume, which is true. What is patented, however, is the notion that a nonlinear taper can provide lower distortion, better reso-



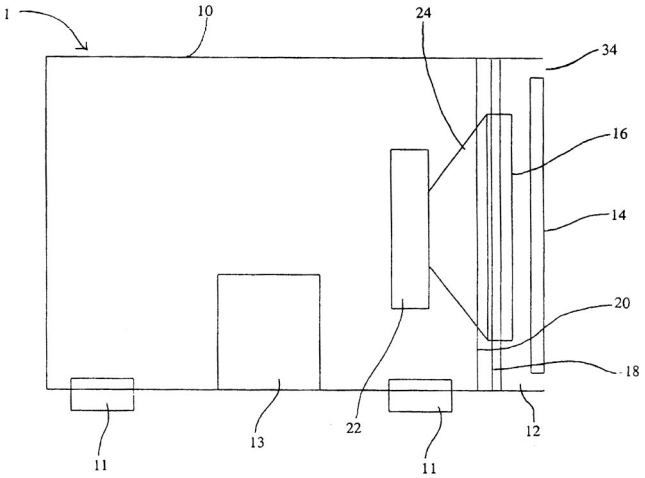
nant behavior, extended bass response, and a further reduction in size. None of these statements is true, but not worth arguing about because nonlinear tapers have been known for at least 30 years. For that matter, this reviewer has investigated and reported on a variety of pipe geometries using software developed in 1998.—GLA

6,704,425

43.38.Ja SYSTEM AND METHOD TO ENHANCE REPRODUCTION OF SUB-BASS FREQUENCIES

Jan P. Plummer, assignor to Virtual Bass Technologies, LLC
9 March 2004 (Class 381/339); filed 19 November 1999

What has been patented here is actually well-known prior art. Sound pressure waves generated by loudspeaker **24** are constrained to move outwardly and emerge from radial slot **34**. The arrangement shown was used by



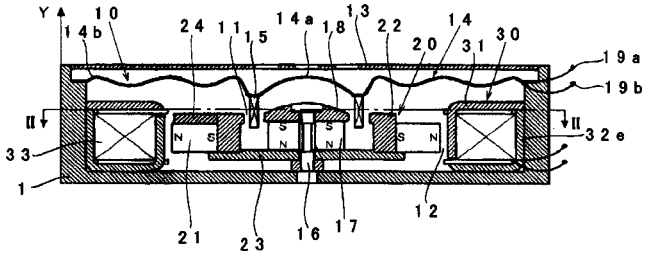
Edmond A. May in designs marketed by Frazier-May in the 1950s and later by JBL. A folded variant described in the patent was used as a subwoofer in a common bass stereo system designed by Paul Weathers, also in the 1950s.—GLA

6,711,269

43.38.Ja MULTIFUNCTION ACOUSTIC DEVICE

Takashi Kobayashi and Akira Nikaido, assignors to Citizen Electronics Company, Limited
23 March 2004 (Class 381/117); filed in Japan 13 March 2001

Those Japanese audio engineers not involved with digital circuitry are spending their time designing dual purpose speaker/vibrators for cellular telephones. In the device shown, audio signals to voice coil **15** drive



diaphragm **14a**. The unusual feature is that the voice coil magnetic circuit **23** can rotate when low frequency signals are fed to vibrator coil **33**. Thus, any combination of sound and vibration can be accommodated without interaction.—GLA

6,700,985

43.38.Kb EAR LEVEL NOISE REJECTION VOICE PICKUP METHOD AND APPARATUS

Jon C. Taenzer, assignor to GN ReSound North America Corporation
2 March 2004 (Class 381/356); filed 30 June 1998

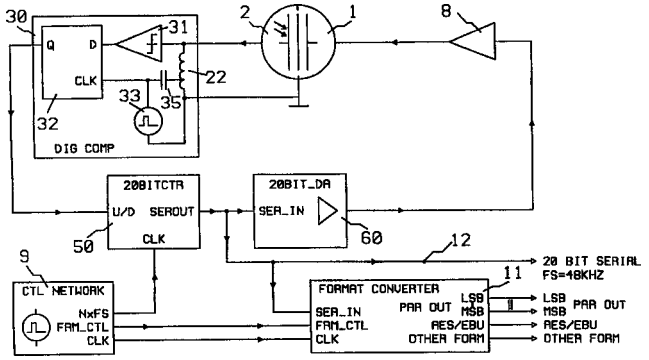
The patent describes an exemplary embodiment as follows, ''...a standard miniature pressure gradient type microphone element is provided and mounted very close to the side of the user's head, preferably near the user's ear. The microphone element is oriented so that its direction of maximum sensitivity is parallel to the side of the user's head, and pointing as much as possible toward the user's mouth.'' Makes sense, but where is the invention?—GLA

6,697,493

43.38.Kb PROCESS AND ARRANGEMENT FOR CONVERTING AN ACOUSTIC SIGNAL TO AN ELECTRICAL SIGNAL

Otmar Kern, assignor to Georg Neumann GmbH
24 February 2004 (Class 381/111); filed in Germany 27 March 1996

The patent includes a good overview of the problems encountered in designing a true digital microphone. An interesting departure from prior art is described in which a capacitor microphone diaphragm is not deflected according to acoustic signal strength but is kept almost at its rest position by



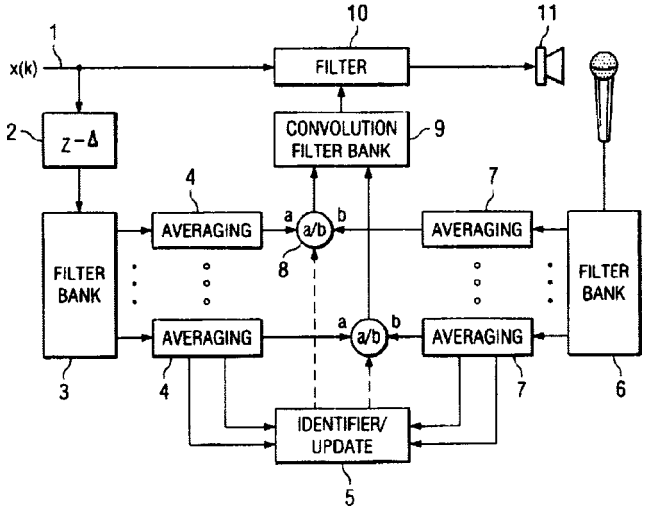
an electrical counter signal. ''It means that the comparator function as (an) elementary function of each analog/digital conversion process is carried out directly at the sound receptor, without requiring an analog intermediate signal...''—GLA

6,697,492

43.38.Lc DIGITAL SIGNAL PROCESSING ACOUSTIC SPEAKER SYSTEM

Hirohisa Yamaguchi and Yoshito Higa, assignors to Texas Instruments Incorporated
24 February 2004 (Class 381/103); filed in Japan 1 May 1998

Several self-equalizing loudspeaker systems are currently available commercially. This patent describes a practical design for such a system. By using two filter banks, it is possible to compare deviations between electrical



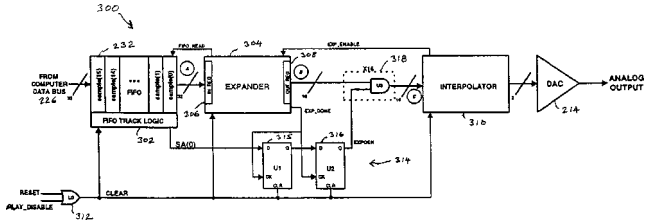
input and acoustical output in individual frequency bands using either random noise or ordinary program material. Thus, the system can automatically equalize itself during actual use. A threshold template is included to help decide if a particular band is equalizable or not.—GLA

6,710,725

43.38.Lc ACOUSTIC NOISE SUPPRESSING CIRCUIT BY SELECTIVE ENABLEMENT OF AN INTERPOLATOR

Martin P. Soques, assignor to Advanced Micro Devices, Incorporated
23 March 2004 (Class 341/61); filed 12 April 1999

Digital audio devices may be expected to handle signals encoded in several different formats. Suppose that such a device defaults to 16-bit signed format. If it now receives data encoded in 8-bit samples, the format



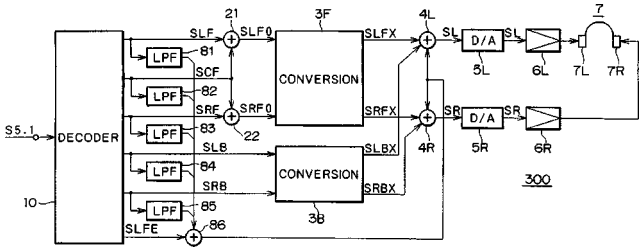
ting block may generate a nasty audio transient while reinitializing the system. As this patent explains, anticipating and suppressing such transients is not a trivial exercise. Those interested in this field will find the patent interesting and clearly written.—GLA

6,711,270

43.38.Md AUDIO REPRODUCING APPARATUS

Hiroshi Masuda, assignor to Sony Corporation
23 March 2004 (Class 381/309); filed in Japan 2 December 1998

Dolby AC-3 is a 5.1 format—left, center, right, and two rear (and the subwoofer). At issue is the automatic conversion of a 5.1 source (like a DVD) into a stereo signal (for headphones, for example). As shown in the



figure, the foreground and background channels are mixed prior to the D/A converter. The conversion blocks **3F** and **3B** reportedly use a head-related transfer function approximation.—MK

6,695,477

43.38.Md AUDIO SIGNAL REPRODUCING APPARATUS

Masayuki Nishiguchi and Yoshihito Fujiwara, assignors to Sony Corporation
24 February 2004 (Class 384/74); filed in Japan 25 October 1989

During the past few years, the bandwidth required for digital music storage has been greatly reduced by the development of efficient ''perceptual encoding'' algorithms. A 64 Mbit memory card can now be used to store almost 10 min of stereo program material. It should be practical to design a very small, portable stereo player by combining (a) a digital memory card reader, (b) decoder circuitry, (c) a D/A converter, (d) stereo headphones, (e) a small housing, (f) a miniature control panel, and (g) cables to connect everything together. That list is exactly what has been patented here.—GLA

6,718,501

43.38.Md COPY PROTECTED DIGITAL AUDIO COMPACT DISC AND METHOD AND SYSTEM FOR PRODUCING SAME

Moshe Brody and Baruch Sollish, assignors to Macrovision Europe Limited
6 April 2004 (Class 714/752); filed in Israel 30 November 1999

Here, the issue is the ease of copying a CD-audio disc by any CD recorder. The inventors propose replacing certain blocks with noise so that the end result is interpolated and sounds the same. The question therefore is how to choose these blocks at the encoder side and how to find them on the decoder side. The inventors depend on the error-correcting code of the CD player to save the audio recording and on the bit-to-bit copying of the CD recorder to invalidate the copy.—MK

6,718,044

43.38.Md FETAL COMMUNICATION APPARATUS

Neville Alleyne, La Jolla, California
6 April 2004 (Class 381/336); filed 2 June 1998

Believing that inter-utero audio stimulation ''may result in an enhanced ability to acquire language and other skills,'' this inventor proposes adding (1) a wireless connection, (2) a visual display of the sound voume, and (3) a ''comfortable'' housing for snugly attaching to the mother's swollen abdomen.—MK

6,662,027

43.38.Si METHOD OF ARBITRATING SPEAKERPHONE OPERATION IN A PORTABLE COMMUNICATION DEVICE FOR ELIMINATING FALSE ARBITRATION DUE TO ECHO

Samuel Han Siong Chia *et al.*, assignors to Motorola, Incorporated
9 December 2003 (Class 455/570); filed 16 March 2001

This brief patent describes what is probably a relatively ineffective method of speakerphone operation. There is no mention of the usual long-delay FIR filtering. The system relies solely upon timing, switching off the incoming signal during those periods when a reflection of the outgoing signal would be expected. Fortunately, the system is intended for use where a normal echo canceller is already in use by the phone company.—DLR

6,697,491

43.38.Vk 5-2-5 MATRIX ENCODER AND DECODER SYSTEM

David H. Griesinger, assignor to Harman International Industries, Incorporated
24 February 2004 (Class 381/20); filed 3 September 1998

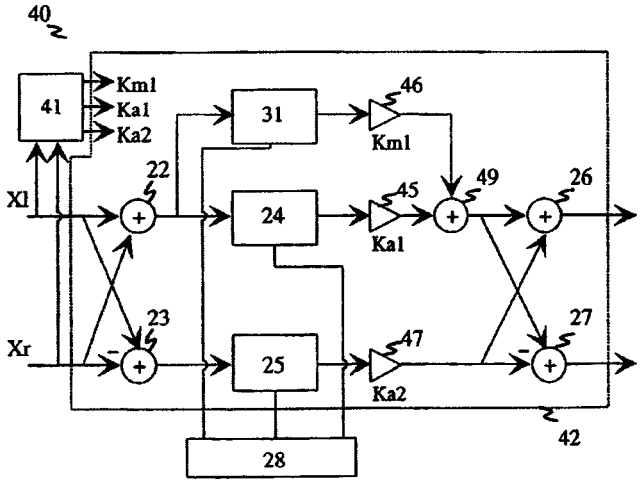
A sound reproduction system converts two-channel input signals to five- or seven-channel surround sound and vice versa. Correlated components of the input signals in the desired direction are enhanced by a matrix while strengths of signals not associated with the encoded direction are reduced.—DAP

6,700,980

43.38.Vk METHOD AND DEVICE FOR SYNTHESIZING A VIRTUAL SOUND SOURCE

Matti Sakari Hämäläinen and Jukka Holm, assignors to Nokia Display Products Oy
2 March 2004 (Class 381/17); filed in Finland 7 May 1998

This patent includes the ringing admonition, ''...monophony must be defined separately.'' Not only would this make a wonderful T-shirt slogan but it makes sense in the context of the invention. As with some earlier



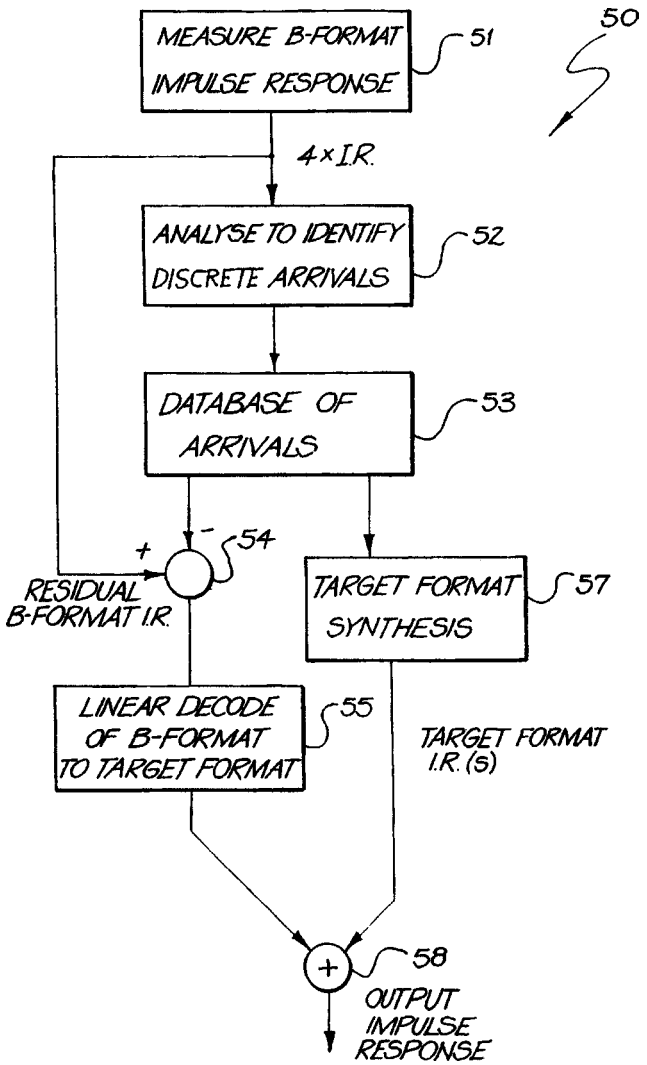
devices, a major goal is to expand the stereo image without objectionable coloration of monophonic signals. A mono/stereo estimator is described whose output is used to set gains of individual enhancement filters.—GLA

6,707,918

43.38.Vk FORMULATION OF COMPLEX ROOM IMPULSE RESPONSES FROM 3-D AUDIO INFORMATION

David Stanley McGrath and Adam Richard McKeag, assignors to Lake Technology Limited
16 March 2004 (Class 381/58); filed in Australia 31 March 1998

What is described in this short patent does not seem all that complicated at first reading. The goal is to provide improved reproduction of ambisonic-encoded program material through multiple loudspeakers. Loudspeaker/room response is first measured to derive impulse response data for all four w, x, y, and z components. Early arrivals are identified and separated out from the reverberant tails. ''A database of arrivals is deter-



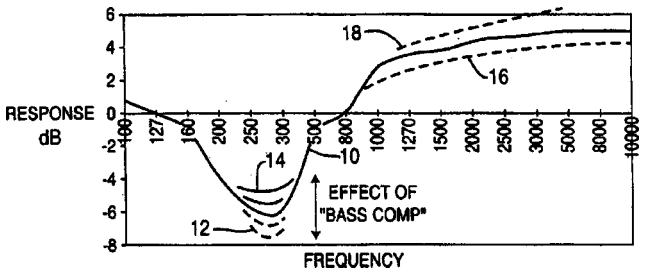
mined and utilized firstly, to subtract the arrivals out of the initially measured impulse response functions so as to form a residual B-format impulse response function which is then linearly decoded utilizing standard techniques. The database of arrivals is also separately utilized so as to synthesise the detected targets separately on the output speaker array.'' Any questions?—GLA

6,711,265

43.38.Vk CENTRALIZING OF A SPATIALLY EXPANDED STEREOPHONIC AUDIO IMAGE

Robert Edward Morris, Jr., assignor to Thomson Licensing, S.A.
23 March 2004 (Class 381/17); filed 24 October 2001

In two-channel stereo reproduction it has long been known that boosting the relative level of the difference signal expands the width of the stereo image. Such stereo expansion circuitry is usually included in single-cabinet



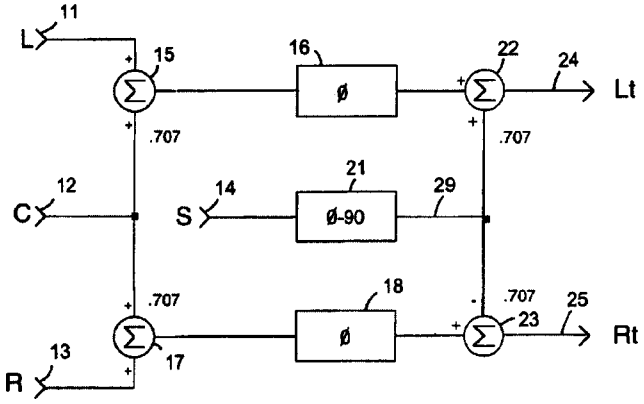
systems such as TV receivers. According to this patent, ''It has been found that ... decreasing the bass response of the (L+R) sum signal along with increasing the treble response of the (L+R) sum signal helps the listener to better localize central audio material more towards the center.''—GLA

6,711,266

43.38.Vk SURROUND SOUND CHANNEL ENCODING AND DECODING

J. Richard Aylward, assignor to Bose Corporation
23 March 2004 (Class 381/22); filed 7 February 1997

More than a dozen patented circuits attempt to increase the overall spaciousness of stereo reproduction, but suppose you want to do just the opposite? This surround sound encoding scheme includes such a feature



which ''... facilitates reducing the perceived width of the front left/right sound stage image when listening to audio-for-video sound fields reproduced in concert with a video display device...''—GLA

6,722,201

43.40.Le METHOD AND DEVICE FOR DETERMINING VIBRATION CHARACTERISTICS OF VIBRATED ARTICLES SUCH AS EGGS

Josse De Baerdemaeker *et al.*, assignors to FPS Food Processing Systems B.V.
20 April 2004 (Class 73/627); filed in the European Patent Office 5 March 2001

A device is disclosed for determining vibration characteristics of vibrated, supported, generally round, substantially ellipsoid articles, such as eggs. An elastic hammer taps and thus acoustically vibrates each article. A handle drive element reciprocates the hammer in a plane. A microphone located immediately next to the specimen being evaluated picks up acoustic vibrations generated by the specimen. A signal processor analyzes the microphone signals to determine the vibration characteristics of the specimen.—DRR

6,715,566

43.40.Tm BALANCE STRUCTURE FOR ROTATING MEMBER

Don Wallace, Smithville, Texas
6 April 2004 (Class 175/56); filed 29 April 2002

This patent relates generally to a balance system for rotating equipment, such as a well drill. In order to overcome an imbalance that may result when clay or other material adheres to the rotating member, this member is provided with an annular cavity that is partially filled with small high-density pellets in a viscous fluid medium. In the event of an imbalance, these pellets distribute themselves so as to oppose this imbalance.—EEU

6,715,746

43.40.Tm VIBRATION ISOLATION DEVICE WITH LOAD DEPENDENT STIFFNESS

Paul J. Bachmeyer *et al.*, assignors to Lord Corporation
6 April 2004 (Class 267/294); filed 20 July 2001

This device, stated specifically to relate to vibration isolation of an aircraft engine oil tank, incorporates a mechanical ''fuse.'' This fuse, which consists of a shear pin, provides a rigid connection that essentially disables the isolator as long as the vibratory load is small. Once this load exceeds a certain value, the shear pin ruptures and lets the isolator operate conventionally.—EEU

6,719,108

43.40.Tm VIBRATION DAMPING DEVICE FOR VEHICLES

Koichi Hasegawa and Rentaro Kato, assignors to Tokai Rubber Industries, Limited
13 April 2004 (Class 188/378); filed in Japan 27 March 2000

This patent describes several embodiments of impact-type dampers consisting of rubber-coated spherical masses located in somewhat oversized structural cavities. Structural vibrations result in the masses impacting against the cavity walls so as to remove energy from these vibrations.—EEU

6,719,503

43.40.Tm TUNED DAMPED ABSORBER SUPPORT

Paul E. McCalmont *et al.*, assignors to Unova IP Corporation
13 April 2004 (Class 409/141); filed 7 September 2001

A dynamic absorber (or tuned damper), here intended to suppress vibrations of a machine tool in three directions of motion, consists of a resiliently supported mass that can vibrate in three orthogonal directions. The mass consists of a rectangular block that is fastened to the tool structure by means of long bolts that are located in oversized holes. The bolt heads rest on springs or other resilient elements to provide resilience in the bolt's axial direction, while limiting the bolt's stiffness contribution in the lateral directions. O-rings or similar elastic elements, placed between the mass and the tool structure, provide stiffness in the lateral direction to augment that due to the bolts.—EEU

6,723,401

43.40.Tm VIBRATION DAMPING MEMBER AND METHOD OF MAKING SAME

Robert T. McKnight *et al.*, assignors to Ergodyne Corporation
20 April 2004 (Class 428/35.7); filed 25 August 2000

This patent describes a laminate that is intended for use in gloves, on tool handles, and on shoes, in order to reduce injuries caused by personnel exposure to vibrations. The laminate consists of a layer of a polymeric material like neoprene, a middle layer of a material like foamed polyurethane, and a third layer of a viscoelastic material such as nonfoamed polyurethane. The outer layers are intended to attenuate vibrations in different mid-frequency ranges, whereas the inner layer is intended to attenuate high-frequency vibrations.—EEU

6,554,099

43.50.Gf EXHAUST MUFFLER

Masayuki Uegane, assignor to Honda Giken Kogyo Kabushiki Kaisha
29 April 2003 (Class 181/275); filed in Japan 1 March 2000

Here is yet another muffler design that incorporates a valve to control the flow path of gases through the muffler in response to engine speed. The valve is designed with a view towards simplicity in terms of number of components and ease of assembly. Two pipes of slightly different diameters are designed to slide one inside the other. Holes in the walls of both pipes can alter the gas flow. Relative motion of the two pipes is driven by the exhaust gas pressure and acts against a coil spring. Numerous geometric variations are described.—KPS

6,557,799

43.50.Gf ACOUSTIC TREATED THRUST REVERSER BULLNOSE FAIRING ASSEMBLY

Joe E. Sternberger, assignor to The Boeing Company
6 May 2003 (Class 244/110 B); filed 9 November 2001

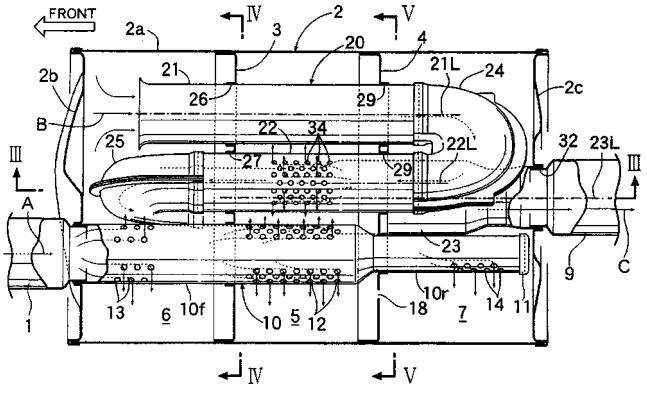
An acoustical treatment for application to the thrust reverser of a turbofan aircraft engine consists of a honeycomb core bonded between a perforated face sheet and a solid backing sheet. Various alternative material systems are described along with fastening arrangements.—KPS

6,571,911

43.50.Gf MUFFLER FOR AN ENGINE

Satoshi Watanabe *et al.*, assignors to Honda Giken Kogyo Kabushiki Kaisha
3 June 2003 (Class 181/265); filed in Japan 21 September 2000

An automobile muffler consists of inlet **1**, outlet **9**, and baffle plates **3** and **4**. The long outlet pipe consists of three straight sections **21**, **22**, and **23**, connected by elbows. The interesting aspect of this design is that the



longitudinal axes of these three straight sections are positioned on the vertices of a triangle, thus resulting in an overall muffler shape that is substantially cylindrical.—KPS

6,571,910

43.50.Gf METHOD AND APPARATUS FOR IMPROVED NOISE ATTENUATION IN A DISSIPATIVE INTERNAL COMBUSTION ENGINE EXHAUST MUFFLER

Mark Storm, assignor to Quiet Storm, LLC
3 June 2003 (Class 181/264); filed 20 December 2001

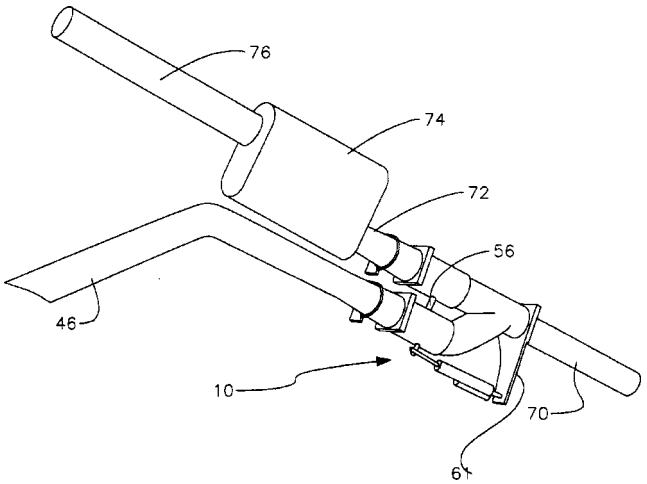
A dissipative muffler typically consists of a straight, perforated pipe surrounded by absorptive material, such as fiberglass, and enclosed in an outer, cylindrical casing. Such an arrangement is reasonably effective at high frequencies and less so at low frequencies. Replacing the perforated pipe with fiber metal having much higher flow resistance is aimed at improving low-frequency performance and mitigating problems associated with ''blow out'' of the fill material, thus yielding longer muffler life. Other arrangements in which line-of-sight through the muffler is blocked by fiber metal are also described.—KPS

6,584,767

43.50.Gf EXHAUST DIVERTER

Steve Koenig, Fairfax, South Dakota
1 July 2003 (Class 60/288); filed 9 November 2001

An exhaust system consists of two alternative flow paths, one through a muffler, the other releasing directly to the surrounding atmosphere. A valve



arrangement determines which flow path is followed. The contribution of this patent appears to be the ability to control the valve system from within the vehicle.—KPS

6,715,580

43.50.Gf GAS FLOW-THROUGH LINE WITH SOUND ABSORPTION EFFECT

Ralph Gerstner and Hans Kutter–Schrader, assignors to Stankiewicz GmbH
6 April 2004 (Class 181/224); filed in Germany 12 November 1997

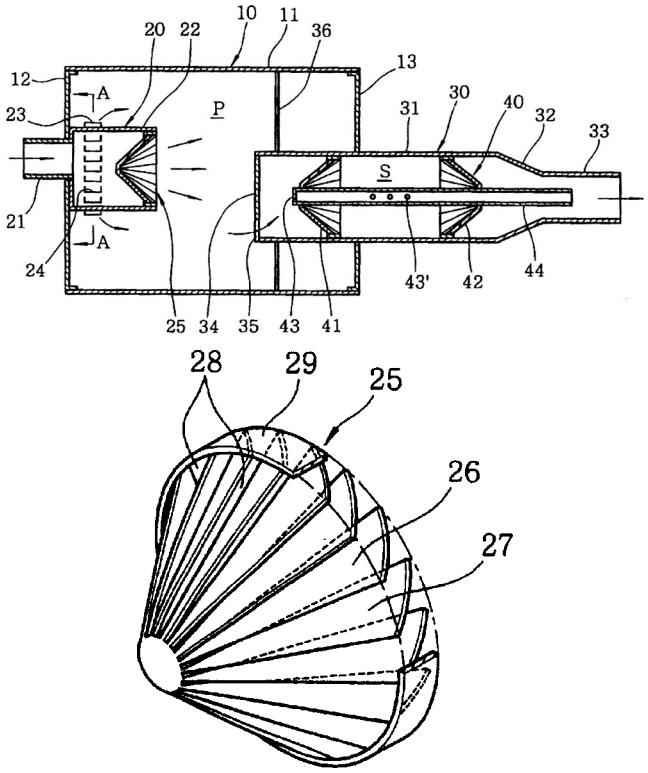
The wall of a flexible hose of polymeric or rubberlike material, such as an air suction hose of an internal combustion engine, is provided with a multitude of perforations with different diameters of less than 1 to 2 mm. A sound absorption effect is said to result from turbulent flows that arise in the regions of these holes. Sheathing in the form of a secondary hose that surrounds the perforated hose, but with some standoff, may provide attenuation via Helmholtz resonator effect.—EEU

6,588,545

43.50.Gf MUFFLER FOR INTERNAL COMBUSTION ENGINE

Ok–no Lee, assignor to Ok–no Lee
8 July 2003 (Class 181/264); filed in the Republic of Korea 5 February 1999

A muffler for an internal combustion engine consists of an inlet **21** with slots **23** and a ''vortex generator'' **22**. The vortex generator consists of twisted blades **26**, **27** with air passages between them. Additional vortex



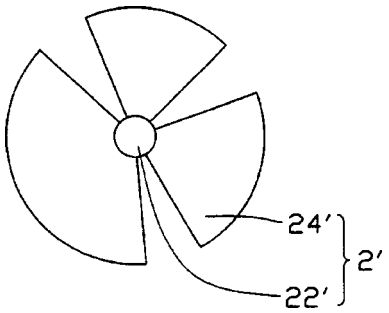
generators **40**, **41** are found in the outlet pipe. This arrangement is said to exhibit good acoustic performance and less backpressure than other designs.—KPS

6,719,530

43.50.Gf FAN INCORPORATING NON-UNIFORM BLADES

Min Chan John Chow, assignor to Hon Hai Precision Ind. Company, Limited
13 April 2004 (Class 416/203); filed in Taiwan, Province of China 12 December 2001

Fan **2′** incorporates a central hub **22′** and a number of spaced and substantially sector-shaped blades **24′** that extend radially from the central hub. While the shapes of the blades are similar, their sizes differ from one another and the spacings between any two adjacent blades differ from each



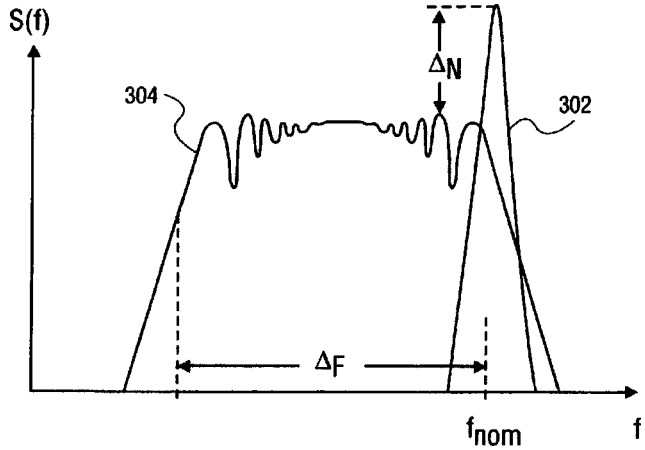
other. The patent asserts that the blades can be configured so that their natural frequency lies outside of the working frequency range of an associated heat dissipation assembly. It is also said that the resonance of the blades can be avoided, thereby minimizing blade noise and vibration when the fan is operated.—DRR

6,720,756

43.50.Gf MODIFYING ACOUSTIC EMISSIONS OF FANS

Harry G. Skinner *et al.*, assignors to Intel Corporation
13 April 2004 (Class 324/71.1); filed 5 November 2001

In this arrangement for reducing unwanted acoustic fan noise, the control signal for a computer system's cooling fan is modulated in such a



manner that the fan's acoustic noise power spectral density **304** covers a bandwidth greater than that shown at **302**, when the control signal remains constant.—DRR

6,722,466

43.50.Gf ACOUSTIC BLANKET FOR MACHINERY AND METHOD FOR ATTENUATING SOUND

Wei Tong *et al.*, assignors to General Electric Company
20 April 2004 (Class 181/200); filed 7 October 2002

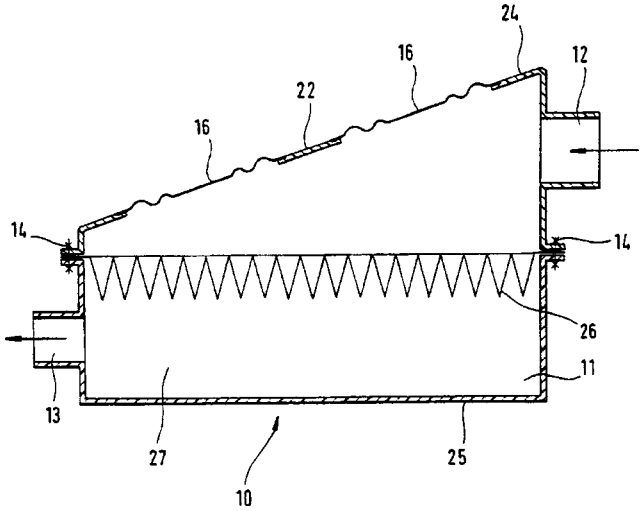
An acoustic blanket, intended for use on a noisy industrial machine, includes multiple flexible panels, each of which includes at least one layer of fiber glass material, an outer casing of a chemically resistant material, and an attachment for connecting the panel to an adjacent panel. The deployment within the blanket of these flexible panels is such that each panel accommodates a specific location and contour on the machine. In another embodiment, the blanket consists of a quilt of interconnected flexible panels in which each panel includes at least one layer of a sound absorbing material, an outer casing of chemically resistant material, and an attachment to connect to the next panel. Each panel bears an index indicating the position of the panel in the blanket and a securing device is provided to hold the blanket onto the noise-generating machine.—DRR

6,551,389

43.50.Lj SOUND TRANSMITTING AIR FILTER

Helmut Spannbauer and Juergen Freisinger, assignors to Filterwerk Mann & Hummel GmbH
22 April 2003 (Class 96/380); filed in Germany 27 August 1999

An air filter for use in an automobile is designed to radiate sound so that the driver can hear the state of the engine. Air flowing through inlet **12**



strikes housing **22**. The face of the housing has circular or oval membranes **16**, which are designed to be efficient sound radiators.—KPS

6,546,734

43.50.Nm PROCESS AND DEVICE FOR ATTENUATING THE NOISE MADE IN A TURBOMACHINE BY ROTOR/STATOR INTERACTION

Angélique Hélène Antoine *et al.*, assignors to SNECMA Moteurs
15 April 2003 (Class 60/772); filed in France 21 September 2000

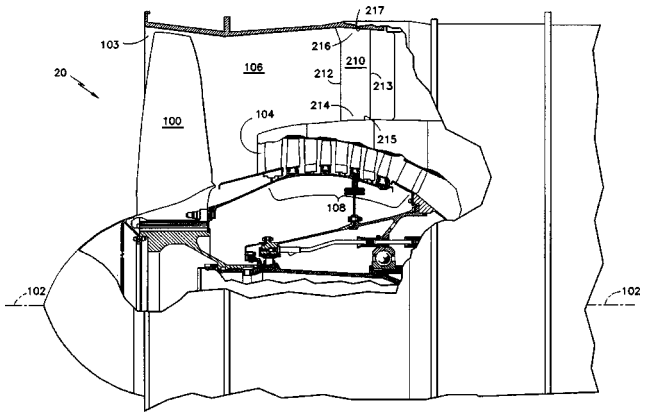
A method is described to reduce noise in a turbofan engine due to interaction between the wakes from the fan blades and the downstream stator blades. Bleed air is introduced through orifices arranged around the perimeter of the inner wall of the inlet duct, upstream of the fan face. The number of orifices is equal to the number of stator blades. This ''ring'' of orifices is able to rotate through an angle equal to that between two stator blades. Noise generated by the interaction of the air jets and the fan blades is thus adjusted to be exactly out of phase with that produced by the fan wake-stator interaction. A microphone placed upstream, in the inlet, provides the error signal for the control system.—KPS

6,554,564

43.50.Nm REDUCED NOISE FAN EXIT GUIDE VANE CONFIGURATION FOR TURBOFAN ENGINES

Wesley K. Lord, assignor to United Technologies Corporation
29 April 2003 (Class 415/119); filed 14 November 2001

Noise reduction in a turbofan aircraft engine can be achieved by sweeping the stator (exit guide) vanes **210** in the downstream, axial direction. This approach is designed to reduce tonal noise associated with interaction between the fan blade wakes and the stator vanes. This patent is aimed at reducing broadband noise. Rather than sweeping the entire blade,



only outer portion **216** is swept, either forward or backward. An alternate design consists of sweeping both the inner and outer portions of the blade, but in opposite directions, thus eliminating the need for the additional duct length required when sweeping the entire blade.—KPS

6,720,068

43.55.Ev SOUND ABSORBENT THIN-LAYER LAMINATE

William Ray Vanbemmel *et al.*, assignors to Rieter Automotive (International) AG
13 April 2004 (Class 428/317.9); filed in Switzerland 3 March 1998

This sound-absorbing, thin-layer laminate combines a backing of open cell or low-density foam and an open cell layer of compacted nonwoven microfibers for use in the automotive industry. This process provides a simple, low-cost, acoustically highly effective, ultralight, thin-layer laminate, which exhibits excellent sound-absorbing properties at lower frequencies with structural stability.—CJR

6,720,069

43.55.Ev SOUND ABSORBING STRUCTURE

Atsushi Murakami *et al.*, assignors to Nichias Corporation
13 April 2004 (Class 428/319.3); filed in Japan 14 January 1999

This sound absorbing structure consists of an open cell foam, most likely for use as an engine cover for an automobile. The foam has a specific cell structure formed by mixing open cells and closed cells of specific diameters, shapes, and depth, some of which penetrate through the foam, some of which do not.—CJR

6,720,362

43.55.Ev PERFORATED FOAMS

Chung P. Park, assignor to The Dow Chemical Company
13 April 2004 (Class 521/79); filed 12 March 2001

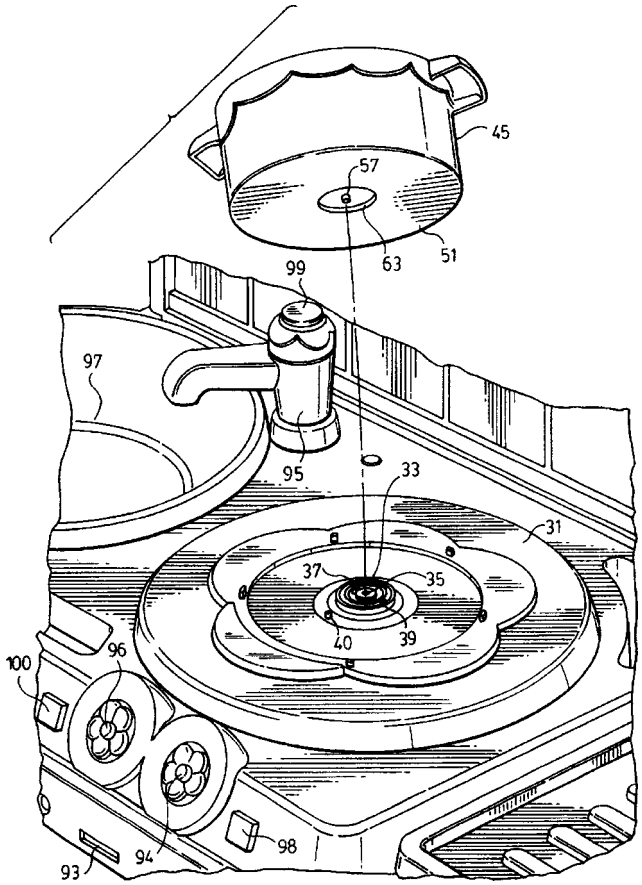
This perforated thermoplastic polymer foam uses average cell sizes greater than about 2 mm arrayed in a substantially open cell structure with relatively large pores connecting the cells. This creates a foam with relatively low airflow resistance and rather greater mechanical strength. It is economical to manufacture and is hydrolytically stable for use in sound absorption, sound insulation, and other applications.—CJR

6,719,603

43.58.Wc INTERACTIVE TOY PLAY SET WITH SENSORS

Albert Wai Tai Chan, assignor to Thinking Technology, Incorporated
13 April 2004 (Class 446/175); filed 29 August 2002

Both boys and girls of a certain age like toy stoves. Imagine augmenting a toy stove with sound: switch **37** can be turned on with pot **45** and then



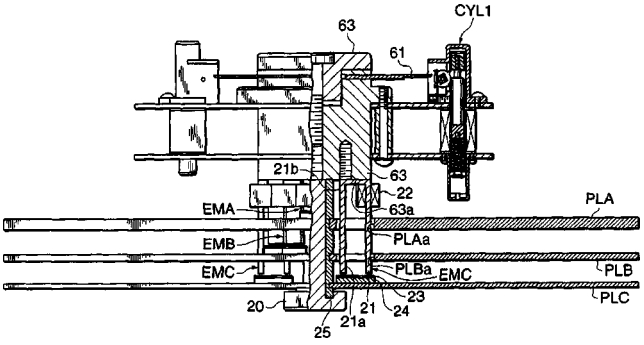
you can generate all the sounds of a kitchen (excluding obscenities). Will olfactory imitations soon follow?—MK

6,723,911

43.58.Wc PERFORMANCE APPARATUS

Shigeru Muramatsu and Yasutoshi Kaneko, assignors to Yamaha Corporation
20 April 2004 (Class 84/723); filed in Japan 24 May 2002

Yamaha proposes an electromagnetic reed instrument. The cross section is shown in the figure. The electromagnetic solenoid is **CYL1**; it strikes



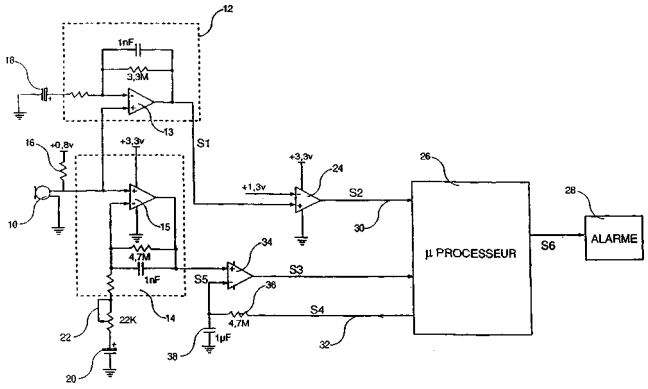
the reed **61** (held by the center block **63**). Plates **PLA**, **PLB**, and **PLC** are resonant plates.—MK

6,720,875

43.60.Bf SELF-ADJUSTING ALARM DEVICE WITH LOW ENERGY CONSUMPTION

François Philippe, assignor to F and F International S.A.R.L.
13 April 2004 (Class 340/541); filed 18 November 2002

Acoustic sensors for alarms suffer from false alarms due to ''random and fugitive atmospheric disturbances.'' The inventor notes that wind has a



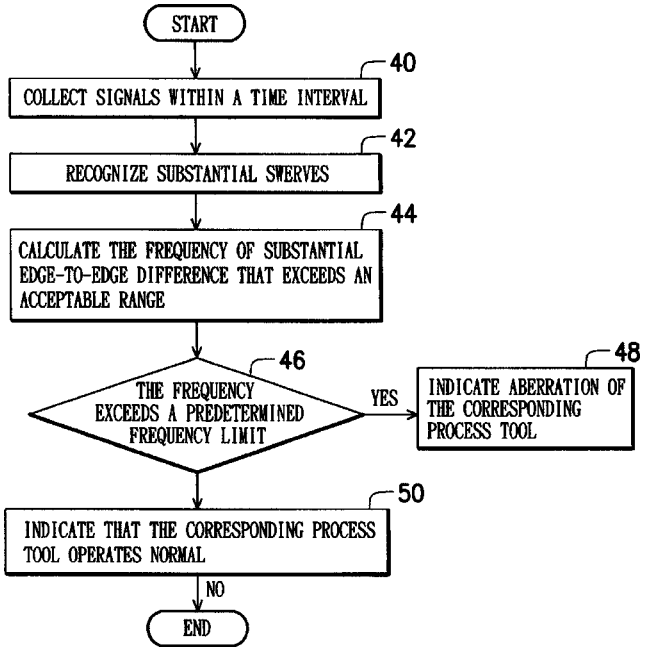
low-frequency component of 10–20 Hz and so, by suitable filtering, one can change the bias and threshold of the comparator.—MK

6,720,885

43.60.Bf METHOD FOR GENERATING A SIGNAL VIBRATION ALERT

Chin-Wen Lee and Hung-Wen Chiou, assignors to ProMOS Technologies Incorporated
13 April 2004 (Class 340/683); filed 13 November 2001

This patent pertains particularly to motion control systems used in the manufacture of integrated circuits. It describes a way of automatically generating an alert when the level and frequency of a signal vibration exceed a certain scale, without noise interference. Multiple signals generated by an apparatus are collected within a time interval. Based on the vibratory signals, substantial swerves and edge-to-edge differences are recognized. Every



three consecutive substantial swerves constitute a peak or a valley. An edge-to-edge difference is the difference between two consecutive swerves, which is tested to see if it exceeds a predetermined noise range. If the edge-to-edge difference frequency exceeds a preset acceptable frequency limit, an alert is generated to indicate functional aberration in the apparatus being monitored.—DRR

6,724,689

43.60.Bf PERSONAL IDENTIFICATION METHOD AND APPARATUS USING ACOUSTIC RESONANCE ANALYSIS OF BODY PARTS

Philip Koenig, Winchester, Massachusetts
20 April 2004 (Class 367/125); filed 8 March 2002

We have here a biometric personal identification system that employs acoustic energy to capture the supposedly unique acoustic resonance spectrum of a selected body part, such as a person's hand. The spectral information is stored in an electronic database and subsequently compared with data being scanned for identification and recognition purposes. The device captures the acoustic resonance spectrum of a person's head, feet, fingers, toes, or other appropriate parts of the body and stores the spectral information for later use, thereby forming a database library of spectra pertaining to any number of participants in the recognition process. The device also maps and compares the spectral information collected from a scanned sample against the spectra stored in the library and renders decisions as to the identification of the scanned subject.—DRR

6,721,234

43.60.Dh TEMPORAL AND SPECTRAL SPREAD SUBMARINE LOCATOR TRANSMITTER, AND METHOD AND INSTALLATION USING SAME

Hubert Thomas, assignor to Advanced Concepts & Systems Architecture
13 April 2004 (Class 367/127); filed in France 10 March 2000

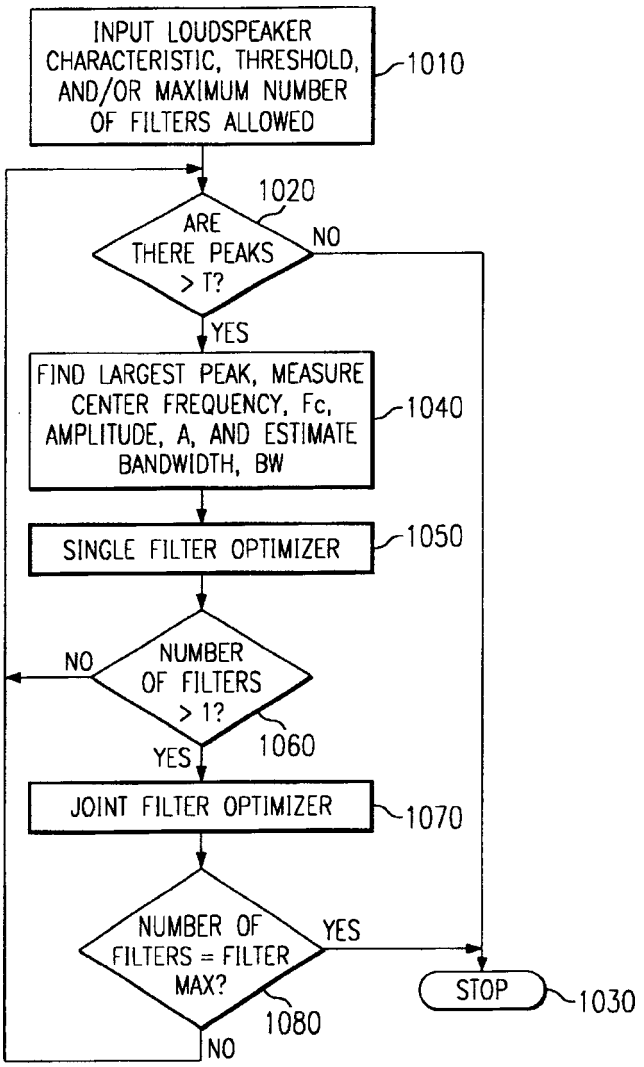
To aid in the tracking of the movement of a mobile underwater unit, such as a submarine or a torpedo, over a tracking range, the units are equipped with transmitters that radiate aperiodic, pulsed acoustic signals according to a predetermined pseudorandom sequence. The signals are received at a number of measurement points, each of which is equipped with a receiver, a time scale reference, and a means of tracking the movement of the measurement point itself. All the information gathered at the measurement points is transmitted to a central signal processing center.—WT

6,721,428

43.60.Dh AUTOMATIC LOUDSPEAKER EQUALIZER

Rustin W. Allred *et al.*, assignors to Texas Instruments Incorporated
13 April 2004 (Class 381/103); filed 13 November 1998

The device provides a method for generating digital filters for the purpose of equalizing a loudspeaker. Target and actual response curves of sound level versus frequency are determined for the loudspeaker. These two data sets are compared and it is determined whether the actual response curve is within a tolerance range of the target curve. If the actual response curve does not lie within the tolerance range, digital filters are iteratively



generated, and modified data relating to the response of the loudspeaker modified by the digital audio filters are generated. The frequency, amplitude, and bandwidth of the digital audio filters are automatically optimized until the modified data lies within the tolerance range or a predetermined limit on the number of digital audio filters has been reached, whichever occurs first.—DRR

6,718,217

43.66.Cb DIGITAL AUDIO TONE EVALUATING SYSTEM

Hironobu Shinohara *et al.*, assignors to JSR Corporation
6 April 2004 (Class 700/94); filed in Japan 2 December 1997

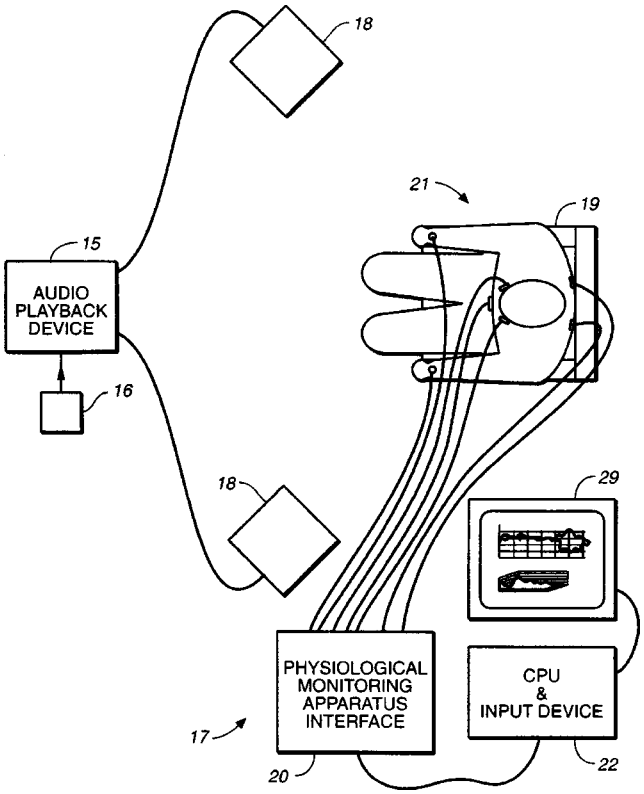
The widely used A-weighting is a rough approximation to the ear's frequency response. The increase in psychoacoustical knowledge combined with digital implementation should produce better evaluation of tonal qualities. Essentially, the inventors propose using a DSP algorithm with 24 critical bands (via the FFT) to compute loudness. What's missing is a discussion of the windowing effects (no mention is made of the short-time Fourier transform) and the evaluation algorithms are simplistic.—MK

6,584,346

43.66.Lj PROCESS AND APPARATUS FOR SELECTING OR DESIGNING PRODUCTS HAVING SOUND OUTPUTS

Ray T. Flugger, assignor to Flowmaster, Incorporated
24 June 2003 (Class 600/544); filed 22 January 2001

Determination of the most pleasing sounds associated with competing product designs (''sound quality'') is typically accomplished through a combination of objective measurements and subjective judgments. The scheme



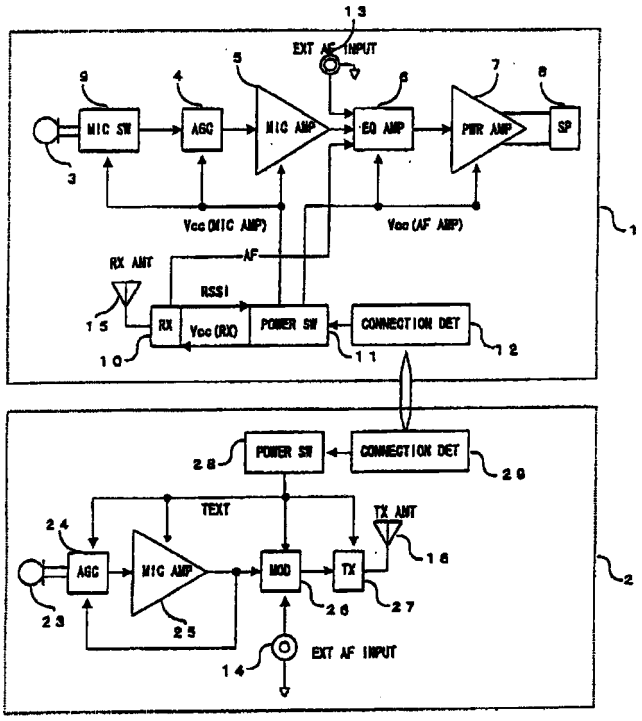
described here replaces the sound measurements and judgments with physiological measures such as pulse rate, respiratory rate, and muscle activity, which are said to be sufficient to discriminate between complex sounds such as those produced by automobile mufflers.—KPS

6,724,900

43.66.Ts HEARING AID

Shouji Fujino, assignor to Temco Japan Company, Limited
20 April 2004 (Class 381/315); filed in Japan 12 May 2000

A hearing aid with a conventional microphone also has an external microphone which may be connected via hard wire or wirelessly. The circuitry switches power automatically between the external and internal microphones when the external microphone is separated from the main hearing aid case. In the wireless mode, a received signal strength indicator detects if



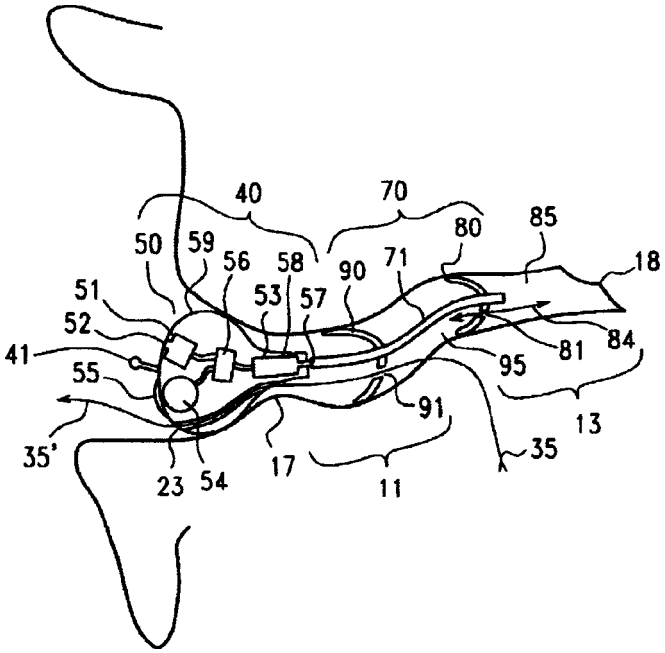
the wireless signal is too weak and switches power back to the primary microphone.—DAP

6,724,902

43.66.Ts CANAL HEARING DEVICE WITH TUBULAR INSERT

Adnan Shennib and Richard C. Urso, assignors to InSound Medical, Incorporated
20 April 2004 (Class 381/328); filed 29 April 1999

A two-part canal hearing device is said to reduce oscillatory acoustic feedback while minimizing occlusion effects. A main module is located in



the cartilaginous portion of the ear canal. Sound output is in the bony portion of the ear canal via a flexible, replaceable tubular insert.—DAP

6,719,707

43.66.Yw APPARATUS AND METHOD FOR PERFORMING MUSICAL PERCEPTION SOUND ANALYSIS ON A SYSTEM

Nathan Montgomery, Royal Oak, Michigan
13 April 2004 (Class 600/586); filed 24 August 2001

This apparatus uses musical perception analysis for analyzing a system. A transducer receives acoustic events generated by the system. An analysis module coupled to the transducer receives the acoustic events and monitors the system as a function of a set of musical perception parameters to identify the presence of a predetermined condition. In another embodiment, an apparatus determines a set of musical perception parameters of a system for use in performing musical sound analysis. This embodiment includes a transducer for sensing acoustic energy of the system under various conditions and, in response, generating acoustic event signals. The apparatus also includes an A/D converter and a computer-based analyzer, which convert the acoustic signals into corresponding musical data signals.—DRR

6,675,114

43.71.An METHOD FOR EVALUATING SOUND AND SYSTEM FOR CARRYING OUT THE SAME

Yoichi Ando and Hiroyuki Sakai, assignors to Kobe University
6 January 2004 (Class 702/75); filed in Japan 15 August 2000

This patent describes an elaborate auditory source analysis system based on certain characteristics of the human auditory system. Binaural signals are autocorrelated and crosscorrelated to obtain the basic working input data streams. Secondary goals of the system include the measurement of timbre, scale, loudness, pitch, tone color, perceived duration, subjective diffuseness, and apparent source width, all based on the correlation data streams. Certain left- and right-ear differences are also included in the model. A detailed mathematical presentation is provided.—DLR

6,675,010

43.71.Ft MOBILE COMMUNICATION SYSTEM FOR LEARNING FOREIGN VOCABULARY

Hao Ming Yeh, Taipei, Taiwan, Province of China
6 January 2004 (Class 455/422.1); filed 22 June 2000

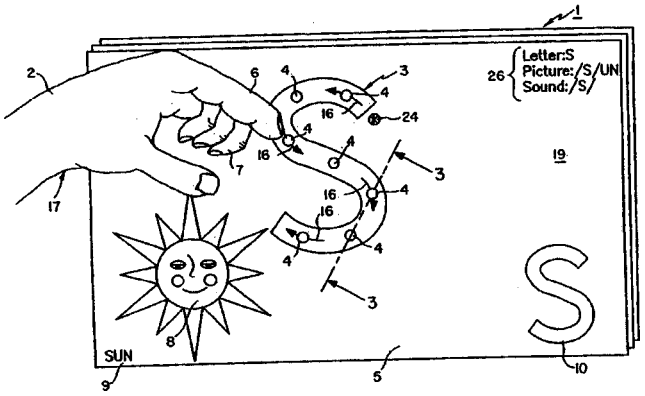
This language-learning device makes use of a variety of Internet services to provide a foreign language learning opportunity in a general mobile context, allowing vocabulary practice at the most effective moment. A large database located at a central server site includes words and usage information. A variety of wireless and other protocols provide access to this database. The user's handheld unit organizes the information into currently active items, study lists, and tests.—DLR

6,669,478

43.71.Ft METHOD AND DEVICE FOR MULTI-SENSORY LEARNING

Ann Edwards, Brooklyn, New York and Karen Goldberg, Glen Ridge, New Jersey
30 December 2003 (Class 434/159); filed 26 November 2001

This patent involves the teaching of letters and numbers and their phonetic characteristics through the use of multiple simultaneous sensory cues, including visual, auditory, and tactile inputs. The case is made that the



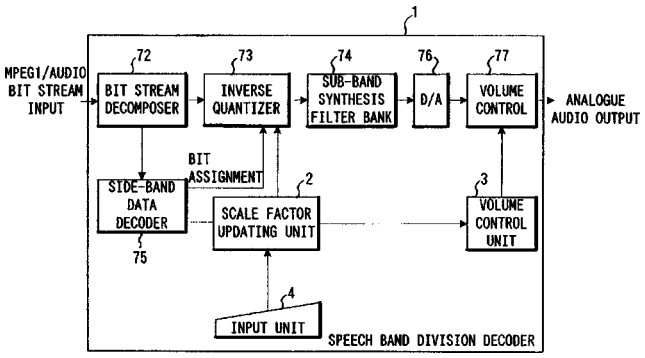
particular combination of cues described here has not been previously patented. The device in question has a raised and tactilely interesting surface for each symbol being presented.—DLR

6,718,295

43.72.Ar SPEECH BAND DIVISION DECODER

Satoshi Hasegawa, assignor to NEC Corporation
6 April 2004 (Class 704/200.1); filed in Japan 5 November 1997

This speech band division decoder incorporates a tone quality control function. The decoder produces a spectral breakdown of a band-division coded signal by dividing a speech signal into multiple subband signals. These signals are coded and synthesized into subbands and a playback scale factor is updated for a predetermined subband signal. The resultant signal is



converted into an analog speech signal having a predetermined frequency characteristic. The speech band division decoder also includes a method of speech signal gain control to control the gain of the converted analog speech signal.—DRR

6,711,546

43.72.Gy UNARY CODING SCHEME FOR DIGITAL AUDIO SIGNALS

David R. Thomas, Opio, France
23 March 2004 (Class 704/500); filed in the European Patent Office 19 July 1999

An audio data compression scheme is based on representing the change in the number of active speaker elements of a unary digital loudspeaker array in a succession of difference amplitude values which are expressed as a sign bit and several data bits. The greater the number of active speaker elements, the greater the audio output volume.—DAP

6,711,618

43.72.Gy APPARATUS AND METHOD FOR PROVIDING SERVER STATE AND ATTRIBUTE MANAGEMENT FOR VOICE ENABLED WEB APPLICATIONS

Ryan Alan Danner and Steven J. Martin, assignors to Cisco Technology, Incorporated
23 March 2004 (Class 709/228); filed 15 December 1999

Voice application control between a web browser and an application server is provided by a web-based voice messaging system with a hypertext transport (HTTP) connection on an Internet Protocol (IP) network. In response to a new web page request from a web browser, the application server initiates a web application and maintains a server-side data record identifying user attribute and session information.—DAP

6,662,161

43.72.Ja COARTICULATION METHOD FOR AUDIO-VISUAL TEXT-TO-SPEECH SYNTHESIS

Eric Cosatto *et al.*, assignors to AT&T Corporation
9 December 2003 (Class 704/260); filed 7 September 1999

This audio/video speech synthesis system simultaneously generates a video talking head image while producing an audio speech sequence. The core of the system is a dual library of audio and video segments, organized by phonemes. Referred to as a ''coarticulation library,'' this database contains sets of video frames covering mouth movement sequences, but there is no mention of any of the usual diphone or triphone sequences as commonly used in speech synthesis. The emphasis is clearly on the video, rather than the audio, aspects of the synthesis.—DLR

6,664,892

43.72.Ja DEVICE INVENTORY BY SOUND

Andrew Thomas *et al.*, assignors to Hewlett–Packard Development Company, L.C.
16 December 2003 (Class 340/505); filed in the United Kingdom 1 December 2000

When you walk into an unfamiliar room, you want to know what's there. If, for some reason, a quick glance won't do, there's a better way. How about if each device could announce its presence? On cue, such as a handclap, each device would speak an identifying phrase. Don't everybody talk at once! Each device listens and waits for a moment of silence before speaking. In case two try to talk at the same time, they use a common network collision avoidance scheme—each stops and waits a random time before trying again. Presumably, if there are any people present, they have to decide for themselves what to do.—DLR

6,665,641

43.72.Ja SPEECH SYNTHESIS USING CONCATENATION OF SPEECH WAVEFORMS

Geert Coorman *et al.*, assignors to ScanSoft, Incorporated
16 December 2003 (Class 704/260); filed 12 November 1999

This diphone synthesizer uses a large speech segment database, which allows selection of an appropriate segment using additional features, including prominence, stress, syllable position, sentence type, and boundary type, in addition to the local phoneme context. A significant portion of the patent deals with a cost function for evaluating the constructed segment sequences. The patent includes a well-written background section covering prior methods of synthesis by concatenation and has many clear and detailed tables and graphs showing how the system works.—DLR

6,651,043

43.72.Ne USER BARGE-IN ENABLEMENT IN LARGE VOCABULARY SPEECH RECOGNITION SYSTEMS

Egbert Ammicht and Allen Louis Gorin, assignors to AT&T Corporation
18 November 2003 (Class 704/275); filed 3 April 2001

This conceptual patent covers two aspects of an interactive voice recognition system, a barge-in capability and a semantic content discriminator. The brief patent contains little explicit detail as to how these goals are to be accomplished, listing, instead, some general approaches which might be of value in building such a recognizer. For example, the barge-in capability would consist of classifying input sounds as speech, traffic noise, etc., rejecting those classifications which are not of interest. The classifier technology is not addressed. The semantic discriminator would consist of checking the recognized phrase against lists of possible input sentences, some relevant to the specific, narrow application, some not. The recognizer technology is not addressed.—DLR

6,662,157

43.72.Ne SPEECH RECOGNITION SYSTEM FOR DATABASE ACCESS THROUGH THE USE OF DATA DOMAIN OVERLOADING OF GRAMMARS

Brenda Horowitz *et al.*, assignors to International Business Machines Corporation
9 December 2003 (Class 704/235); filed 19 June 2000

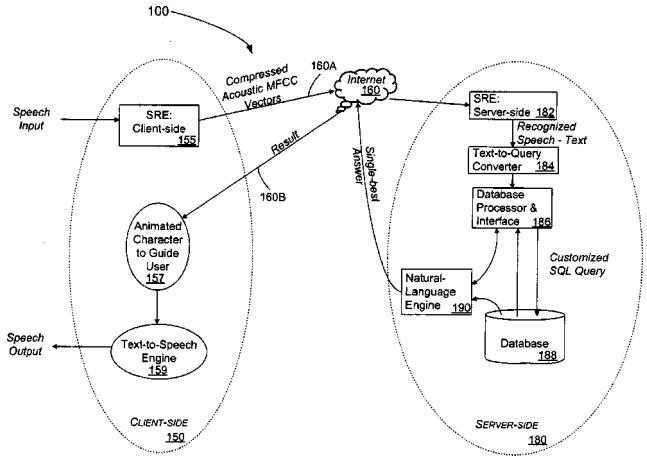
Intended for use in an application with a relatively limited semantic breadth, such as a travel reservation system, the basic idea of this speech recognition system is that the grammar would contain all relevant words and phrases and only those words and phrases relevant to the specific application area. If an input item matches the grammar, it is handled. If not, it is rejected with an error message. That's the concept. In fact, the grammar does include certain additional related items which would allow a response such as ''I'm sorry. I have no information on ...,'' rather than the more brusque, ''Please repeat.''—DLR

6,665,640

43.72.Ne INTERACTIVE SPEECH BASED LEARNING/TRAINING SYSTEM FORMULATING SEARCH QUERIES BASED ON NATURAL LANGUAGE PARSING OF RECOGNIZED USER QUERIES

Ian M. Bennett *et al.*, assignors to Phoenix Solutions, Incorporated
16 December 2003 (Class 704/257); filed 12 November 1999

This speech recognition system is organized as a study aid for students using the Internet. Speech input by the user is minimally analyzed and a packet of parameters is sent to a more elaborate HMM system at the server. Using course-oriented vocabulary and grammar, the recognized text is used



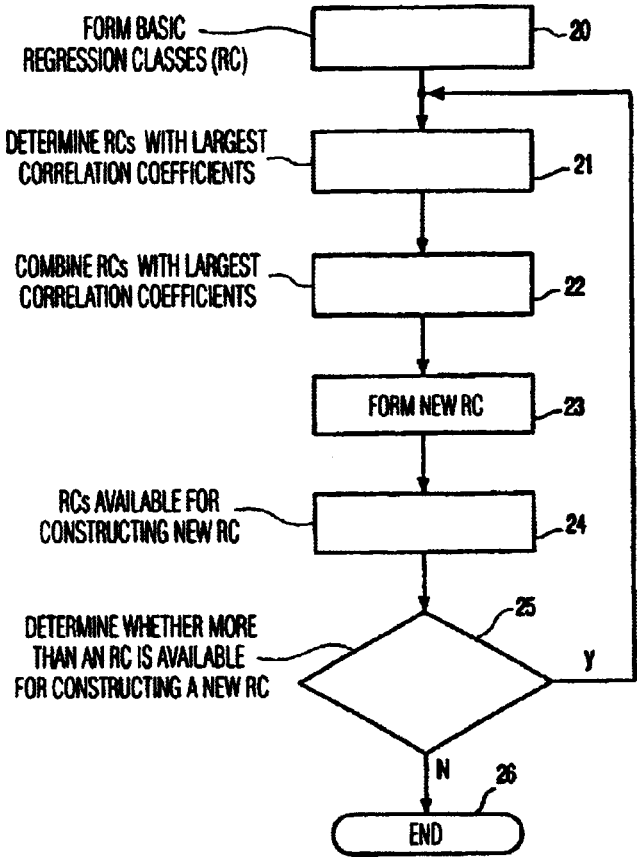
to look up relevant queries in a database of possible system interactions. The result is sent back to the student's computer, where it is used to synthesize an appropriate response.—DLR

6,718,305

43.72.Ne SPECIFYING A TREE STRUCTURE FOR SPEECH RECOGNIZERS USING CORRELATION BETWEEN REGRESSION CLASSES

Reinhold Häb-Umbach, assignor to Koninklijke Philips Electronics N.V.
6 April 2004 (Class 704/245); filed in Germany 19 March 1999

A data-driven estimate for automatic construction of a regression class tree structure is said to reduce the error rate of speech recognition to a level comparable to estimates obtained with expert knowledge of phonetic structures. The tree structure is determined recursively based on combining successive pairs of regression classes having the largest correlation parameter.—DAP

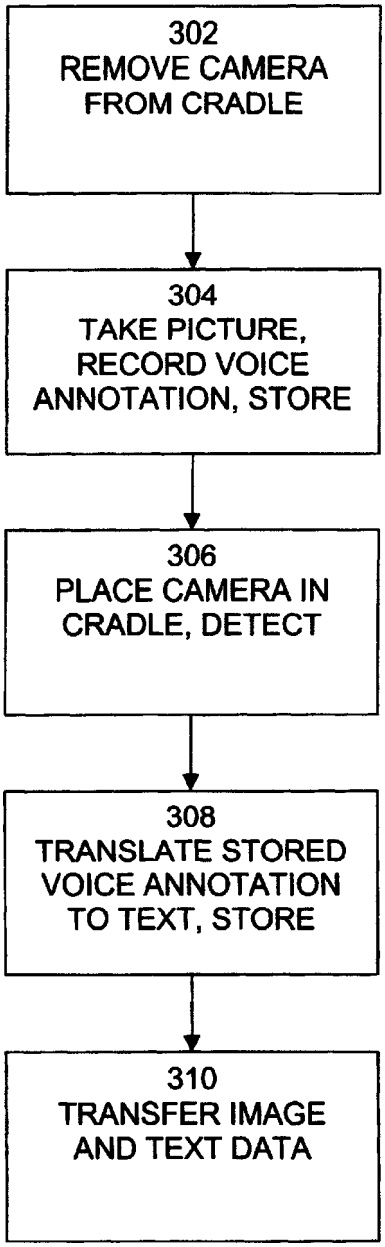


6,721,001

43.72.Ne DIGITAL CAMERA WITH VOICE RECOGNITION ANNOTATION

Viktors Berstis, assignor to International Business Machines Corporation
13 April 2004 (Class 348/231.3); filed 16 December 1998

A camera system facilitates recording voice annotations for each picture. Voice recognition translates the stored voice data into text data which may accompany the digital picture if transmitted.—DAP



6,665,644

43.72.Ne CONVERSATIONAL DATA MINING

Dimitri Kanevsky *et al.*, assignors to International Business Machines Corporation
16 December 2003 (Class 704/275); filed 10 August 1999

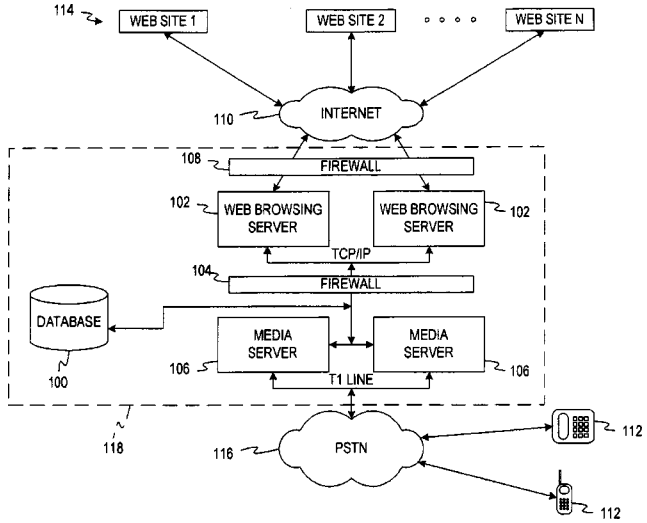
This voice-interactive conversation system collects a wide variety of information about the user, in addition to many direct and obviously voice-related parameters. The patent strongly asserts that all of the wide-ranging personal characteristics, including gender, age, accent, native language, dialect, socioeconomic status, educational level, and emotional state, can all be collected by analysis of the speech signal. All of this information is collected in a database and used to improve the conversational interaction with the user.—DLR

6,721,705

43.72.Ne ROBUST VOICE BROWSER SYSTEM AND VOICE ACTIVATED DEVICE CONTROLLER

Alexander Kurganov and Valery Zhukoff, assignors to Webley Systems, Incorporated
13 April 2004 (Class 704/270.1); filed 5 February 2001

Conversational voice commands over a telephone are interpreted via speaker-independent speech recognition and sent as text messages to data



bases accessed by a web browsing server. Responses from the information sources are converted with speech synthesis to audio messages.—DAP

6,723,904

43.75.Gh AUTOMATED PLAYER FOR STRINGED INSTRUMENTS

Thomas Allen Dolan and Don A. Gilmore, assignors to QRS Music Technologies, Incorporated
20 April 2004 (Class 84/267); filed 19 December 2002

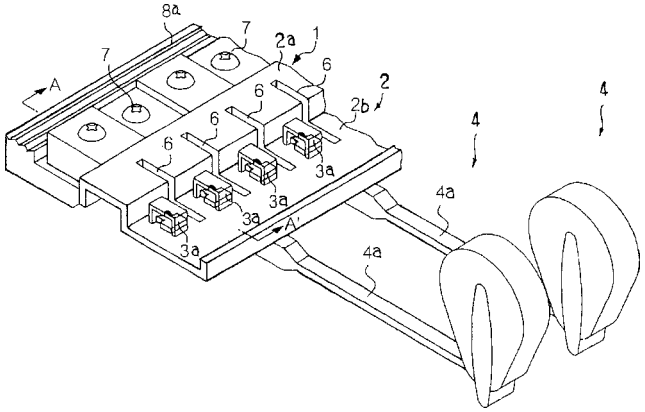
The inventors protest that guitars have lagged behind other instruments in being automated. So they propose a set of microprocessor controlled stepper motors to position the stops over the frets and a rotary plectrum to pluck the strings. What would Nancarrow have done with this?—MK

6,713,751

43.75.Mn EASILY ASSEMBLED OPTICAL FIBER SENSOR AND MUSICAL INSTRUMENT USING THE SAME

Shigeru Muramatsu and Tsutomu Sasaki, assignors to Yamaha Corporation
30 March 2004 (Class 250/221); filed in Japan 30 November 2000

Continuing the long line of Disklavier patents, Yamaha introduces optical sensing by detecting hammer position instead of key position. As shown, the hammers can rotate through slot **6**, blocking the optical path to



sensor head **3a**. The light source and optical fibers are not shown. Once detected, the key signal can engage either the sounding solenoid, synthesizer, or both.—MK

6,720,489

43.75.Mn PIANO MUTER

Daniell Revenaugh, Berkeley, California
13 April 2004 (Class 84/177); filed 10 June 2002

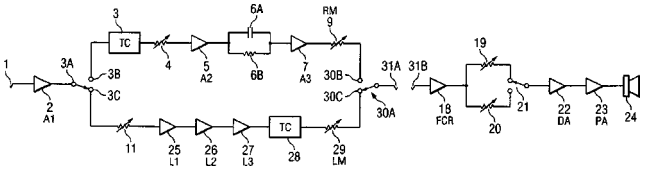
Apartment dwellers now have recourse for their ''loud'' instruments—just install shaped acoustical absorbing material underneath the soundboard. Why not encase the whole instrument?—MK

6,724,897

43.75.Tv AMPLIFIER WITH ALTERNATE OUTPUT LEVEL

Randall C. Smith, Petaluma, California
20 April 2004 (Class 381/61); filed 2 August 2000

The inventor's previous patent, United States Patent 4,211,893, introduced the notion of high-frequency emphasis of an electric guitar to present



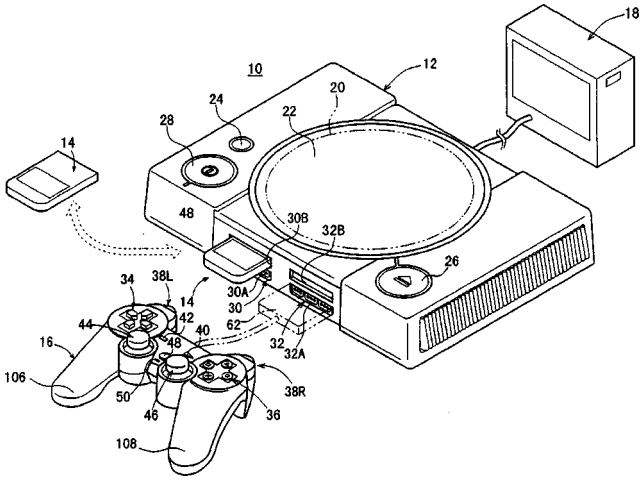
a ''lead guitar'' timbre. Now, he wishes to preserve levels between lead and rhythm signal paths.—MK

6,702,677

43.75.Wx ENTERTAINMENT SYSTEM, ENTERTAINMENT APPARATUS, RECORDING MEDIUM, AND PROGRAM

Takafumi Fujisawa *et al.*, assignors to Sony Computer Entertainment Incorporated
9 March 2004 (Class 463/43); filed in Japan 14 October 1999

Gamers and others interested in the Sony Gameboy will enjoy this



patent. The system architecture is detailed as well as the sound management in the context of (what else?) games.—MK

6,703,552

43.75.Wx CONTINUOUS MUSIC KEYBOARD

Lippold Haken, Champaign, Illinois
9 March 2004 (Class 84/658); filed 19 July 2001

Keyboards without quantized pitches have been known for many years, going back to ribbon controllers and the keyboards of Moog and Buchla. This newer version has sensing in three dimensions and a parabolic fit algorithm to determine position. The mechanical design uses rod magnets and Hall effect sensors.—MK

6,721,701

43.75.Wx METHOD AND APPARATUS FOR SOUND DISCRIMINATION

Stephen Clifford Goss *et al.*, assignors to Lucent Technologies Incorporated
13 April 2004 (Class 704/231); filed 20 September 1999

Automatic sound source identification is an increasingly important problem. The inventors present a very simplistic solution that depends on a close match between input sounds and stored templates. But it should be able to identify the sound of coins in a pay phone.—MK

6,723,910

43.75.Wx REVERBERATION GENERATION PROCESSOR

Chih-Hsien Tsou and Wei-Peng Tsai, assignors to Silicon Integrated Systems Corporation
20 April 2004 (Class 84/630); filed 18 November 2002

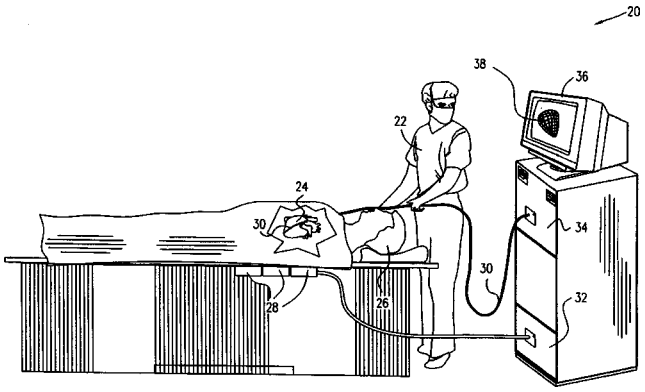
The authors assert that starting from a Fibonacci sequence as noise, they can filter and shape a reverberant response. They missed the critical issue of how to design the filters. They also missed 40 years of research into good sounding artifical reverb.—MK

6,716,166

43.80.Qf THREE-DIMENSIONAL RECONSTRUCTION USING ULTRASOUND

Assaf Govari, assignor to Biosense, Incorporated
6 April 2004 (Class 600/437); filed 10 July 2001

This is a device for mapping and reconstruction of the interior of body organs, such as the heart. It consists of an elongated probe featuring a distal portion adapted for insertion into the body cavity. Multiple acoustic transducers are deployed along the longitudinal axis over the distal portion of the probe. These transducers are adapted to be actuated individually to emit



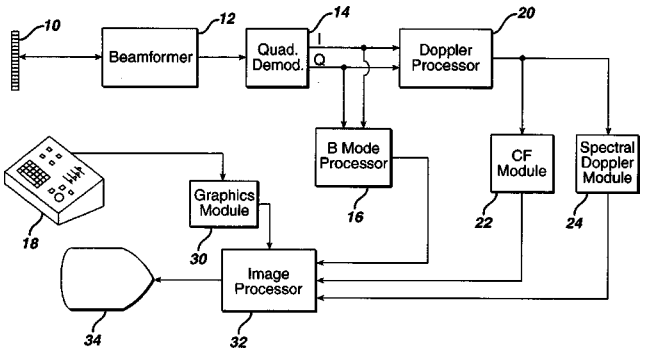
acoustic waves and are further adapted to receive the reflected acoustic waves from the surface of the cavity and to generate, in response to the received waves, electrical signals indicative of times of flight of the waves.—DRR

6,719,697

43.80.Qf ULTRASONIC QUANTIFICATION OF VALVULAR REGURGITANT BLOOD FLOW

Xiang-Ning Li, assignor to Koninklijke Philips Electronics N.V.
13 April 2004 (Class 600/454); filed 27 February 2001

This is an ultrasonic diagnostic imager capable of detecting and quantifying valvular regurgitant blood flow in the heart. This potentially life-threatening condition occurs when a valve in the heart does not fully close during a particular phase of the heart cycle. If this happens, a leak will occur and a jet of blood will escape as the heart contracts, thereby causing the heart to expend more effort than it should and leading to a reduced flow of blood through the body. The device assesses valvular regurgitation by identifying a characteristic regurgitation jet in a colorflow image. An M-line is



placed over the region of regurgitation and Doppler N-mode information is acquired from the flow convergence region adjacent to the regurgitant orifice. The Doppler M-mode information, which is acquired at a higher acquisition rate than the colorflow frame rate, is used to provide a measure of the

regurgitant flow rate and the volume through the orifice. The flow rate can be combined with the velocity data acquired during the regurgitating event to yield a dynamic estimate of the size of the regurgitant valve orifice.—DRR

6,719,698

43.80.Qf ULTRASONIC TRANSDUCER PROBE AND A MEASUREMENT DEVICE UTILIZING THE SAME

Dan Manor *et al.*, assignors to Cardiosonix Limited
13 April 2004 (Class 600/454); filed 18 July 2001

This probe design assists in orienting the probe with respect to a blood vessel being measured, in operating the device during the measurement procedure, and facilitates access of the ultrasound radiation to the blood vessel being measured. The probe housing contains a transducer array. An outer surface of the housing, brought to a measurement location in the patient's body, is contoured so as to have at least one projecting portion. This allows manipulation of the probe with respect to the patient's body so as to partially displace a muscle above the blood vessel, thereby minimizing ultrasound signal attenuation during measurements and also to bring the transducer array close to the blood vessel.—DRR

6,716,168

43.80.Sh ULTRASOUND DRUG DELIVERY ENHANCEMENT AND IMAGING SYSTEMS AND METHODS

Levin Nock *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
6 April 2004 (Class 600/439); filed 30 April 2002

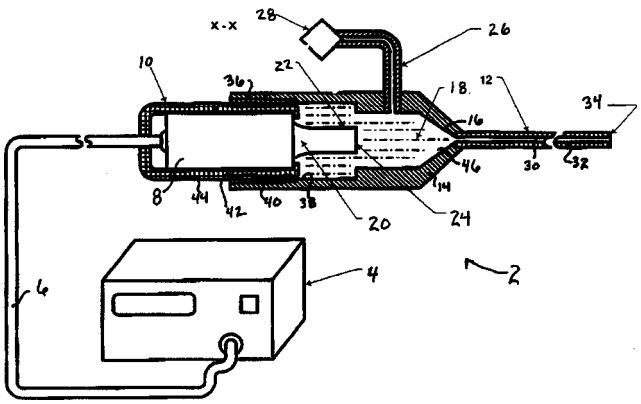
This is a device designed to enhance drug uptake by the application of therapeutic transmissions of acoustic energy while imaging a body region to help guide the therapy with the same transducer and ultrasound system. In one embodiment, a single linear array of transducer elements is applied for both imaging and therapeutic ultrasound. In another embodiment, a transmitter is provided that sends therapeutic transmissions having substantially equal negative and positive peaks at the face of the transducer and also transmits imaging acoustic energy.—DRR

6,723,064

43.80.Sh ULTRASONIC CATHETER DRUG DELIVERY METHOD AND DEVICE

Eilaz Babaev, assignor to Advanced Medical Applications, Incorporated
20 April 2004 (Class 604/22); filed 4 June 2003

This ultrasonic catheter drug delivery system consists of a body or chamber to which is attached a transducer to generate ultrasonic waves. The transducer features a distal tip incorporating a distal radiating surface. When a therapeutic agent from a fluid source is directed to the catheter body or chamber, the radiating surface produces ultrasonic pressure and delivers



liquid and ultrasonic energy simultaneously to a patient's vascularity or to a selected body lumen. The method applies therapeutic agent and ultrasonic waves without the need for direct contact between the ultrasonic transducer and the body. It also dissolves blood clots and stimulates tissue cells.—DRR

6,712,805

43.80.Sh METHOD AND APPARATUS FOR INTRADERMAL INCORPORATION OF MICROPARTICLES CONTAINING ENCAPSULATED DRUGS USING LOW FREQUENCY ULTRASOUND

Ludwig J. Weimann, Burlington, Vermont
30 March 2004 (Class 604/500); filed 28 January 2002

The purpose of this device is to provide *in vivo* intradermal incorporation of microparticles containing encapsulated drugs, therapeutic agents, or cosmetics for topical or subcutaneous application through the use of low-frequency ultrasound. A container having one end covered with a porous membrane contains a microparticle suspension. An ultrasonic horn has its tip submerged in the microparticle suspension containing the treatment drugs. The horn supplies ultrasound radiation to the microparticle suspension at a frequency, an intensity, for a period of time, at some distance from the skin, to effectively generate cavitation bubbles. The cavitation bubbles collapse and transfer their energy into the skin area, thus causing the formation of pores in the skin area. The ultrasonic radiation intensity and distance from the skin help to generate ultrasonic jets, which drive the microparticle suspension through the porous membrane and the formed pores within the skin area. The optimal ultrasound frequency is said to be in the range of 1 kHz to 5 MHz.—DRR

6,723,303

43.80.Sh ULTRASOUND CONTRAST AGENTS INCLUDING PROTEIN STABILIZED MICROSPHERES OF PERFLUOROPROPANE, PERFLUOROBUTANE OR PERFLUOROPENTANE

Steven C. Quay, assignor to Amersham Health, AS
20 April 2004 (Class 424/9.52); filed 6 June 1995

The patent relates to agents supposedly intended to enhance the contrast in an ultrasound image generated for use in medical diagnosis. The contrast-enhancing media consist of extremely small gas bubbles in a solution that is infused into the body during or before the time an ultrasound image is generated. The patent also covers a method for enhancing such images by selecting gases from which a collection of free gas microbubbles can be prepared. These microbubbles may be extremely small in size and yet must survive in the bloodstream long enough to allow contrast-enhanced imaging of parts of the cardiovascular system, peripheral vascular system, and vital organs previously believed to be inaccessible to free gas microbubbles. The selection of particular gases is based on calculations entailing the use of the intrinsic physical properties of a gas and a liquid. In particular, the density of a gas, the solubility of a gas in solution, and the

diffusivity of a gas in solution, which in turn is dependent on the molar volume of the gas and the viscosity of the solution, all come under consideration.—DRR

6,716,174

43.80.Vj FLEXIBLE GEOMETRY FOR REAL-TIME ULTRASOUND VOLUME IMAGING

Xiang-Ning Li, assignor to Koninklijke Philips Electronics N.V.
6 April 2004 (Class 600/447); filed 27 September 2002

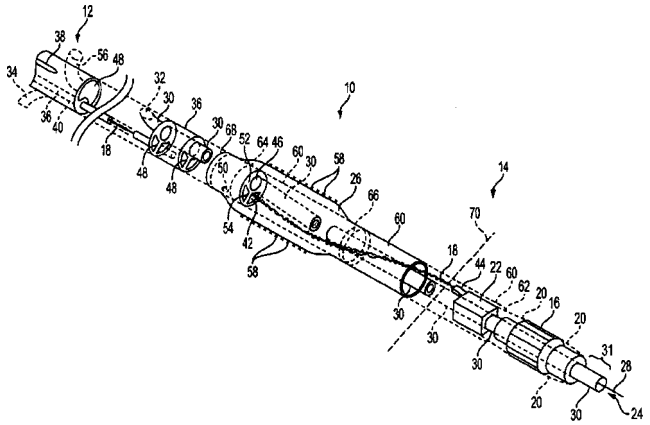
Energy is transmitted into a volume by an array. Echo signals are received. Beams are formed to sample the volume in various geometries such as a circular or elliptical pattern.—RCW

6,712,767

43.80.Vj ULTRASONIC IMAGING DEVICES AND METHODS OF FABRICATION

Norman Hugh Hossack *et al.*, assignors to Volcano Therapeutics, Incorporated
30 March 2004 (Class 600/467); filed 29 August 2002

These devices consist of a catheter with an ultrasound transducer array and an inflatable balloon at the distal end. Inflation fluid, electrical wires, and a guide wire are in lumens of the catheter. A sheath that runs through the



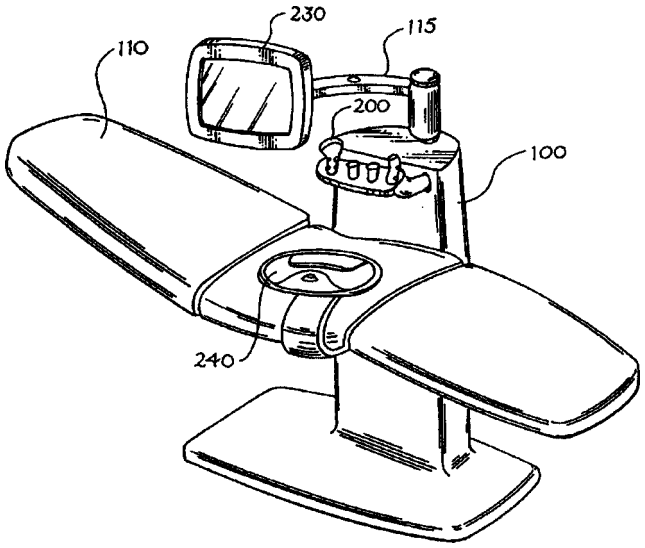
balloon isolates the electrical wire and guide wire from inflation fluid in the balloon.—RCW

6,716,167

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND IMAGING SYSTEM WITH A PATIENT SUPPORT SURFACE

Richard W. Henderson *et al.*, assignors to Siemens Medical Soluions USA, Incorporated
6 April 2004 (Class 600/437); filed 25 September 2001

Combinations of an ultrasound imaging system and a patient support surface are described. A second diagnostic ultrasound imaging system may



be included. Various embodiments may be used alone or in combination.—RCW

6,716,172

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND IMAGING SYSTEM AND METHOD FOR DISPLAYING A PORTION OF AN ULTRASOUND IMAGE

Cynthia L. Kerby and Christian H. Marney, assignors to Siemens Medical Solutions USA, Incorporated
6 April 2004 (Class 600/443); filed 23 December 2002

A region of interest that is selected in an ultrasound image and a portion of another image are saved in a post-scan-converted format. The stored images are displayed and the region of interest is positioned in the portion of the other image. The selected portion is then displayed.—RCW

6,716,184

43.80.Vj ULTRASOUND THERAPY HEAD CONFIGURED TO COUPLE TO AN ULTRASOUND IMAGING PROBE TO FACILITATE CONTEMPORANEOUS IMAGING USING LOW INTENSITY ULTRASOUND AND TREATMENT USING HIGH INTENSITY FOCUSED ULTRASOUND

Shahram Vaezy *et al.*, assignors to University of Washington
6 April 2004 (Class 601/3); filed 7 June 2002

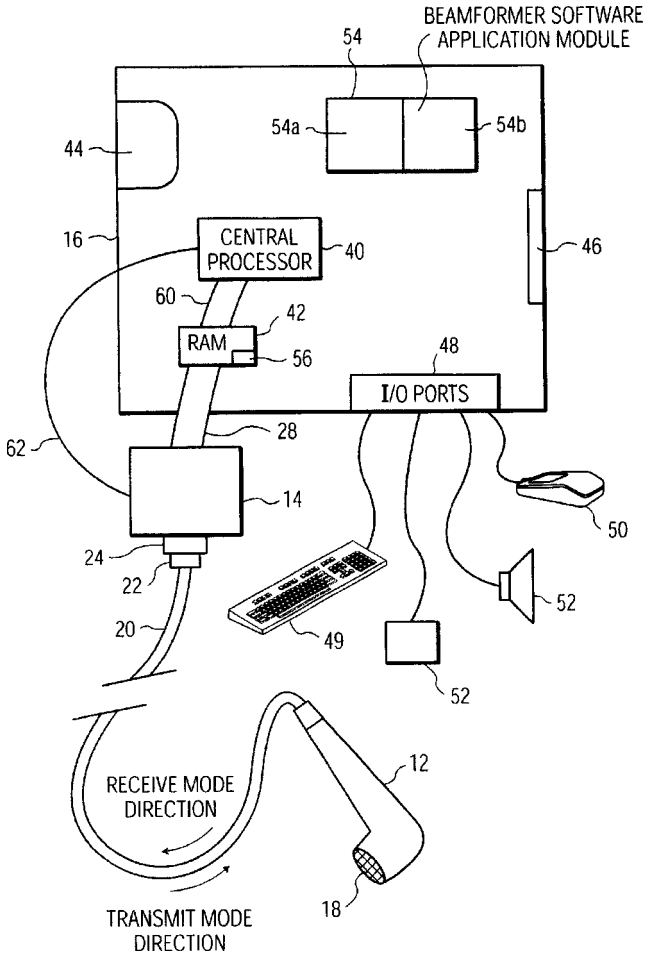
A probe used for ultrasonic imaging and therapy is designed to limit the effects of interference noise in a display by synchronizing transmissions of high-intensity focused ultrasound with transmissions for imaging so that the noise is displayed in an area of the image that does not overlap the treatment site.—RCW

6,716,171

43.80.Vj SYSTEM AND METHOD FOR INTERFACING AN ULTRASOUND TRANSDUCER WITH A COMPUTING DEVICE PERFORMING BEAMFORMING PROCESSING

George Brock-Fisher, assignor to Koninklijke Philips Electronics N.V.
6 April 2004 (Class 600/443); filed 30 September 2002

Signals are exchanged between a transducer array and a computing device to perform partial beamformation on digital data transmitted between the transducer array and the computing device. The computing device con-



tains a random access memory for storage of signals. Processing for both transmit and receive beamformation is implemented in the computing device.—RCW

6,719,693

43.80.Vj APPARATUS AND SYSTEM FOR REAL-TIME SYNTHETIC FOCUS ULTRASONIC IMAGING

William D. Richard, assignor to BECS Technology, Incorporated
13 April 2004 (Class 600/437); filed 28 March 2003

A scalable integrated circuit architecture is employed to construct a synthetic focus in an ultrasonic imaging system that has a large number of transducer elements. Field programmable gate arrays with on-chip memories are used to store digitized ultrasonic echo signals. Logic inside the gate arrays calculates a time-of-flight surface and forms subimages during imaging.—RCW

6,719,694

43.80.Vj ULTRASOUND TRANSDUCERS FOR IMAGING AND THERAPY

Lee Weng *et al.*, assignors to Therus Corporation
13 April 2004 (Class 600/439); filed 22 December 2000

A focused array of transducer elements is employed. Focusing is accomplished using electronic phasing to steer the beam and position the focus. The bandwidth of the array is broadened for imaging by use of an electronic switch to place a resistance in parallel with each element of the array.—RCW

6,719,696

43.80.Vj HIGH RESOLUTION 3D ULTRASOUND IMAGING SYSTEM DEPLOYING A MULTI-DIMENSIONAL ARRAY OF SENSORS AND METHOD FOR MULTI-DIMENSIONAL BEAMFORMING SENSOR SIGNALS

Stergios Stergiopoulos and Amar C. Dhanantwari, assignors to Her Majesty the Queen in right of Canada as represented by the Minister of National Defence
13 April 2004 (Class 600/443); filed in Canada 24 November 1999

Beamformation is accomplished using subapertures that are decomposed into coherent subsets of circular or line array beamformers in different coordinate directions of a multidimensional array. The implementation permits volumetric imaging using a compact multidimensional sensor array and a processing unit capable of producing three-dimensional images in real time and also capable of capturing images as a function of time to store four-dimensional data.—RCW

6,719,699

43.80.Vj ADHESIVE HYDROPHILIC MEMBRANES AS COUPLANTS IN ULTRASOUND IMAGING APPLICATIONS

Larry L. Smith, assignor to Sonotech, Incorporated
13 April 2004 (Class 600/459); filed 7 February 2002

These membranes contain a hydrogel film that is a couplant. The hydrogel adheres to an ultrasound probe and to a protective sleeve that covers the probe. This eliminates the requirement for coupling gels and fluids within the probe sleeve.—RCW

6,719,700

43.80.Vj ULTRASOUND RANGING FOR LOCALIZATION OF IMAGING TRANSDUCER

N. Parker Willis, assignor to SciMed Life Systems, Incorporated
13 April 2004 (Class 600/462); filed 13 December 2002

An ultrasound imaging transducer is operated in one resonant mode to transmit and receive echoes for imaging. The imaging probe is also operated in another resonant mode to transmit or receive an ultrasound position signal. The position of the imaging transducer can then be determined from the transmitted or received position signal.—RCW

Estimation of seabed reflection loss properties from direct blast pulse shape (L)

Mark K. Prior[a] and Chris H. Harrison[b]
NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 23 March 2004; revised 18 June 2004; accepted 18 June 2004)

The decay rate of a smeared multipath arrival received many water depths from a broadband source has been shown to be closely related to bottom reflection loss properties [Smith, J. Acoust. Soc. Am. **50**, 332–336 (1971)]; [Harrison, J. Acoust. Soc. Am. **114**, 2744–2756 (2003)]. This suggests a simple way of measuring bottom properties with active sonars. The method is demonstrated with experimental data measured in a shallow water region of the Mediterranean Sea. Seabed properties are deduced directly by the method in the form of the rate of change of reflection loss with respect to angle close to grazing. Agreement is shown between the results of the method and the results of matched field inversions carried out in the same area. Good agreement is also shown between measurements and predictions of propagation loss made using seabed properties deduced via the method. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1780572]

PACS numbers: 43.20.El, 43.30.Ma [AIT] Pages: 1341–1344

I. INTRODUCTION

A potential method for calculating seabed reflection loss properties from the received pulse shape (direct blast) was suggested by Smith (1971) and Harrison (2003). This could be very useful operationally because it involves no extra measurements or equipment than an existing bistatic active sonar and it is self-calibrating. In essence, there is a simple relationship between the fall-off rate (dB per second) of the tail of the received pulse and the seabed's low-angle reflection loss, independent of any scattering coefficients or other environmental parameters, and independent of the bistatic range. In this paper we provide an experimental demonstration that these predictions are self-consistent, that they agree with local measurements using different techniques in the Mediterranean and that therefore propagation can be well-predicted from the behavior of remote pulsed sources or point targets.

The approach is based on the well-known ''mode-stripping'' phenomenon. Mode-stripping results from two effects: one is that reflection loss (between zero grazing and the critical angle) is usually more or less proportional to angle; the other is that the number of boundary interactions increases as the inverse of ray cycle distance and therefore ultimately also increases with angle. For paths that reflect from both boundaries this leads to intensity having a Gaussian angular dependence and, because one can calculate the travel time of these relatively steep paths without serious refraction problems, one can translate the Gaussian distribution of intensity versus angle into intensity versus time. The method is therefore self-calibrating.

Other methods exist (Pouliquen, 1994) that infer seabed properties from pulse shapes but these are associated with the monostatic detection of sound scattered from the seabed after near-vertical transmission. The approach reported here calculates seabed reflection loss properties from the bistatic detection of multipath arrivals after near-horizontal propagation and is consequently completely unrelated to these other methods.

The approach is complementary to standard geoacoustic inversion techniques and reverberation inversion techniques.

II. SUMMARY OF THEORY

The tail of the multi-path pulse necessarily corresponds to steep rays that ultimately hit both boundaries and are affected only slightly by refraction. Therefore without loss of generality we can assume the water to be isovelocity. The theory can accommodate range dependence (Harrison, 2003) but here we assume a stratified medium. Following Harrison (2003), Smith (1971), and Weston (1971), we assume that the most effective rays are inside the seabed's critical angle and have a reflection loss (measured in dB) proportional to angle $R_{\mathrm{dB}}=10\log(R)=\alpha_{\mathrm{dB}}\theta$. In other words, the intensity reflection coefficient R can be written as

$$R=\exp(-\alpha\theta), \tag{1}$$

with $\alpha=\alpha_{\mathrm{dB}}/(10\log(e))$. In this paper, α is referred to as the ''reflection loss gradient'' and it is measured in nepers per radian. This results in an exponential decay in range r but a Gaussian distribution of received intensity, I, in angle θ,

$$I(\theta)d\theta=B(\theta)\exp\left(-\frac{\alpha\theta^2}{2H}r\right)d\theta, \tag{2}$$

where H is water depth and $B(\theta)$ represents angle-dependent factors such as source, receiver, or target vertical directionality. In all the following we assume that B is either known or uniform, so we replace it by unity. Since travel time t is given by

$$ct=r\sec\theta\cong r+r\theta^2/2, \tag{3}$$

we can map the angle distribution to time to determine the intensity pulse shape, $p(\tau)$, for a one-way path (the direct blast from a distant source),

[a]Electronic mail: prior@saclantc.nato.int
[b]Electronic mail: harrison@saclantc.nato.int

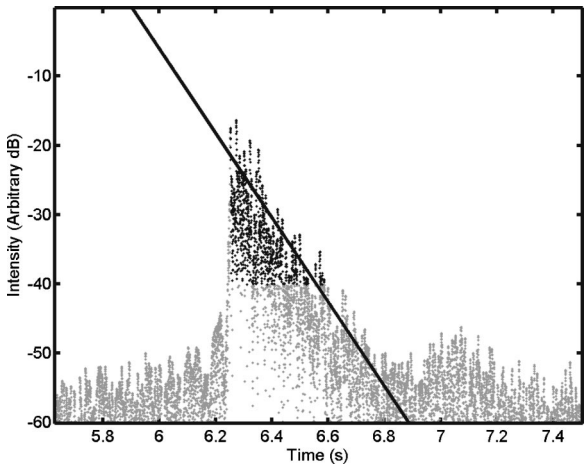


FIG. 1. Intensity versus time for one measured pulse. The straight line shows a least-squares fit to the data used to calculate the rate of change of logarithmic intensity with time. Dark points show the data used to fit the line.

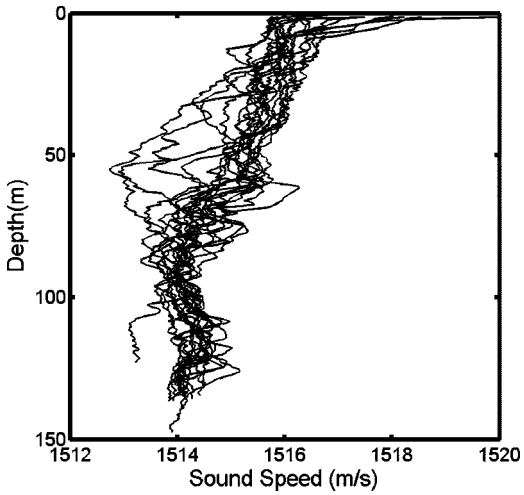


FIG. 2. Sound speed versus depth data deduced from expendable bathythermograph casts taken during the experiment.

$$p(\tau)d\tau=(2t_0)^{-1/2}\frac{\exp\{-\alpha\tau/t_H\}}{\sqrt{\tau}}d\tau, \tag{4}$$

where τ is time after the first arrival $\tau=t-r/c=r\theta^2/2c$, $t_0=r/c$, and $t_H=H/c$. Also for a two-way path (distant point target) the pulse shape is

$$p(\tau)d\tau=\pi/2t_0\exp\{-\alpha\tau/t_H\}d\tau. \tag{5}$$

Thus in both cases the pulse's decay time is $t_H/\alpha=Hc^{-1}\alpha^{-1}$ which, knowing H and c, determines the reflection loss gradient α directly.

One of the tests applied later requires a propagation model, and although any model would suffice, we will use the mode-stripping model derived by integrating Eq. (2) over angle (Harrison, 2003),

$$I=\sqrt{\frac{2\pi}{H\alpha r^3}}\,\mathrm{erf}(\sqrt{\alpha r/2H}\,\theta_c), \tag{6}$$

where θ_c is the critical angle.

III. EXPERIMENTAL ARRANGEMENT

To investigate the predictions made by the theory, measurements were made during the SACLANTCEN PREBASE experiment in the Malta Plateau region, south of Sicily. Transmissions were made from a towed source and received on a single element of a distant vertical line array at position 14° 51′E, 36 20′N. Transmissions were linear frequency-modulated sweeps with frequency changing from 800 Hz to 1400 Hz over the one-second duration of the pulse. The transmitter was a single, free-flooding ring with an effectively omnidirectional beam pattern in the vertical. Upon reception, signals were correlated with the transmitted signal (i.e., replica correlated) giving a temporal resolution of approximately 1.7 milliseconds. Figure 1 shows a typical pulse received. Transmissions were made at a series of locations covering source-to-array ranges from a minimum of 0.9 km to a maximum of 10 km.

An automatic algorithm was applied to each pulse to determine the time of peak arrival. This time was then subtracted from all other times to yield the reduced time variable, τ. The data were then scaled by the ''square root of reduced time'' factor included in the denominator of Eq. (4). The resulting data were then least-squares-fitted with a straight line and the gradient of this line recorded. The fit to the pulse is also shown in Fig. 1.

Seawater sound speed values deduced from expendable bathythermograph casts made in the area are shown in Fig. 2. The data in the figure show that, except for some small-scale surface heating, the water column showed a slightly downward-refracting but almost isovelocity sound speed profile with mean sound speed 1514.8 ms^{-1}. Water depth values were taken from echo-sounder measurements made during the experiment and showed that the water depth varied only by ± 2 m from a mean value of 131 m during all the measurements. These environmental measurements suggested that the environment could be described as being range-independent with an isovelocity water column.

The acoustic data gathered during PREBASE allowed the theoretical predictions described in the previous section to be tested in four ways. First, the theory's prediction of a range-invariant fall-off of logarithmic intensity with time was examined by inspecting gradients of the lines least-squares-fitted to pulses transmitted over a spread of source-receiver ranges. Second, the values of the gradients were combined with seawater sound speed and water depth estimates to yield values for reflection loss gradient. The physical credibility of these values was then investigated by a comparison with values taken from the literature of sediment acoustic properties. The third test involved the estimation of seabed sound speed, density and attenuation, using the deduced values of the reflection loss gradient. This geoacoustic description of the seabed was compared with a description derived from a full matched-field inversion experiment carried out nearby on the Malta Plateau. Finally, the seabed description deduced from the pulse shape method was used to predict propagation loss as a function of range and this prediction was compared with propagation measurements taken during the PREBASE experiment.

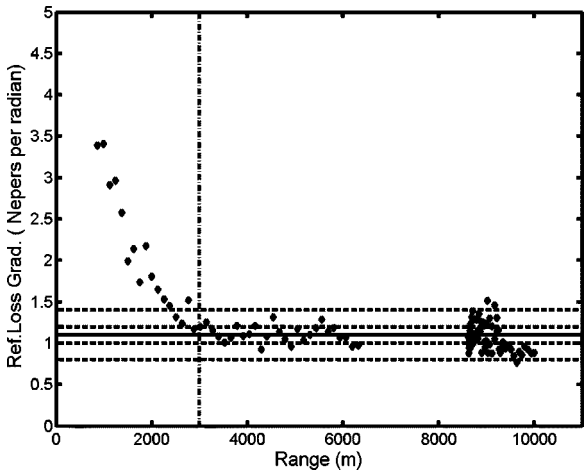


FIG. 3. Values of reflection loss gradient (α) deduced from least-squares-fitted dB per second fall-off, plotted as a function of the source–receiver range. The solid horizontal line shows the mean of values at ranges greater than 3 km, dashed lines show ± 1 and ± 3 standard deviations from the mean.

The results of the four tests of the theoretical predictions are now described.

IV. COMPARISONS OF THEORY AND OBSERVATIONS

A. Range-invariance of pulse shape

One of the predictions of the mode-stripping expressions outlined in Sec. II is the range-invariance [notwithstanding the square root term in the denominator of Eq. (4)] of the rate of change of logarithmic intensity with time. This was investigated by calculating the gradients of least-squares-fitted lines of the type shown in Fig. 1 for pulses transmitted over the entire spread of source–receiver ranges used in the PREBASE experiment. Figure 3 shows values of reflection loss gradient, α, deduced from Eq. (4) using the known water depth and sound speed. The figure shows that, for ranges greater than 3 km, reflection loss gradient, and hence pulse shape, is indeed invariant with range, as predicted by the mode-stripping expression Eq. (4). For ranges less than 3 km, the apparent variation of reflection loss gradient with range is a curve-fitting artifact. In reality, the angle spread is larger than the critical angle so that the pulse is truncated. Nevertheless, the initial fall-off rate is unchanged. The truncation causes the curve fit to produce too steep a fall-off rate.

B. The inferred value of reflection loss gradient

Figure 3 shows values of reflection loss gradient for ranges greater than 3 km with a mean of 1.1 nepers rad^{-1} and a standard deviation of ± 0.1 nepers rad^{-1} about this mean. The following question now arises: ''Does this value of reflection loss gradient correspond to a physically realistic sediment?'' Table I, taken from (Harrison, 2003), shows sound speeds, densities and attenuations for sediments in the range of ''coarse sand'' to ''silty clay.'' A relationship between α and the seabed geoacoustic parameters [Weston, 1971, Eq. (18)] was used to derive the corresponding value of α. The mean value of 1.1 nepers rad^{-1} lies in the range between silt and silty clay and is consequently within reasonable bounds.

TABLE I. Geoacoustic parameters for typical sediment types, taken from Hamilton (1980).

Type	Speed (m/s)	Specific gravity	Absorption (dB/λ)	α (nepers/rad)	θ_c (deg)	$r_0/H = 2/(\alpha\theta_c^2)$
Coarse sand	1836	2.03	0.6	0.31	35.22	17.02
v. fine sand	1709	1.88	0.5	0.43	28.63	16.59
Silt	1615	1.74	0.3	0.65	21.75	21.4
Silty clay	1517	1.48	0.1	3.18	8.59	27.9

C. Comparison with matched-field inversion results

Further confidence in the values of reflection loss gradient deduced from pulse shape would be obtained if it were possible to compare them with values determined via independent methods of estimating seabed properties. A full matched-field inversion experiment (Siderius, 2002) was carried out at a site just to the north of the PREBASE area. This experiment used numerical geoacoustic inversion methods to produce a two-layer fluid description of the seabed in terms of sound speed, density and attenuation of a sediment and basement. The geoacoustic parameters for the sediment layer can be converted into reflection loss gradient, yielding a value of 1.11 nepers rad^{-1}, as shown in Table II. This is within one standard deviation of the mean value deduced using the pulse shape method and such close agreement represents further confirmation of the validity of the pulse shape method for estimating seabed properties. Table I emphasizes the closeness of this agreement and shows the reflection loss gradient varying between 3.18 and 0.31 for sediments in the range of coarse sand and silty clay.

A comparison with the matched-field results can be taken further if the reflection loss gradient estimates are converted into geoacoustic parameters. This process requires the adoption of a model of the seabed and the model adopted here was of a fluid half-space completely described by a sound speed, a density and a compressional wave attenuation. Shear wave properties of the sediment were ignored, for consistency with the seabed description used in the matched field inversion study (Siderius, 2002).

Seabed sound speed was estimated by consideration of the transition range, r_0, beyond which unmodified mode-stripping conditions exist. The Gaussian distribution of intensity with angle shown in Eq. (2) suggests that r_0 can be defined as the range at which the argument of the exponential

TABLE II. A comparison of derived and published geoacoustic parameters.

Type	Speed (m/s)	Specific gravity	Absorption (dB/λ)	α (nepers/rad)	θ_c (deg)
Pulse shape[a]	1577	1.600	0.248	1.1	16.1
Pulse shape[b]	1590	1.615	0.315	1.1	17.7
Matched field inversion[c]	1554.3	1.8	0.1	1.11	12.9

[a]Lower limit.
[b]Upper limit.
[c]Siderius (2002).

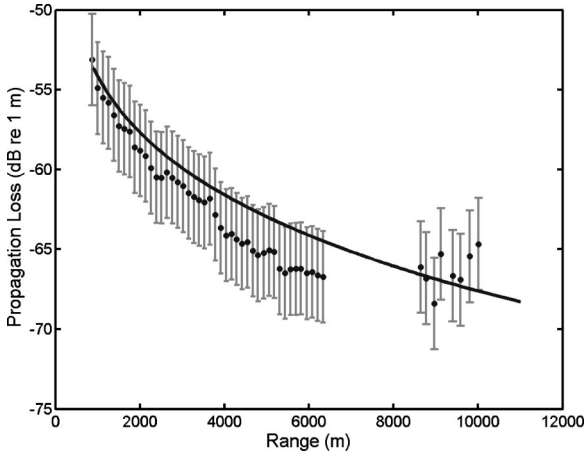


FIG. 4. Propagation loss measured (line with error bars) and predicted using values of the reflection loss gradient (α) and critical angle inferred from the pulse shape, inserted in Eq. (6).

function is equal to unity. Equation (2) shows that r_0 is thus twice the water depth divided by the product of reflection loss gradient and the square of the critical angle (see Table I). The data shown in Fig. 3 allow r_0 to be estimated as being the range from the source below which the deduced values deviated significantly from the mean value. The inspection of Fig. 3 shows that r_0 lies somewhere between 2.5 km and 3 km. Using these values for r_0 allows bounding values of 16.1° and 17.7° to be derived for the critical angle. These yield estimates of seabed sound speed between 1577 m/s and 1590 m/s. These values of sound speed were used to estimate density via Hamilton's regression curves relating sediment sound speed to density (Hamilton, 1978). The corresponding values of specific gravities are 1.600 and 1.615 and the sediment is classified as ''soft, silty.'' These estimated values of sound speed and density were then combined with the measured mean value of α to give estimates of seabed attenuation through Weston's relation [Weston, 1971, Eq. (18)]. This yielded values of 0.24 dB/λ and 0.315 dB/λ. Table II compares these values with those produced for the upper layer using the matched field inversion method. While agreement is not exact, both sets of data lie within the range of ''silt-silty clay,'' values that are shown in Table I.

D. Comparison of measured and predicted propagation loss

In sonar applications, seabed reflection loss properties are required for predicting long-range propagation and reverberation. The most important test of the seabed description is therefore how closely it allows propagation loss to be predicted. Although the same transmissions were used to measure pulse shape and propagation loss, it is emphasized that the two quantities are independent, since reflection loss is derived from the *shape* of the pulse while the propagation loss is determined by the *amplitude* of the pulse. Indeed, the fact that the seabed description can be obtained using uncalibrated data, where the absolute amplitude is completely unknown, is one of the method's attractive features. Furthermore, the expressions for the shape of the pulse, Eq. (4), and the total intensity, Eq. (6), while both derived using ideas based on mode-stripping, operate in different domains.

Figure 4 shows propagation loss measured and predicted using Eq. (6). This expression used only the inferred reflection loss gradient and critical angle to describe the seabed and values for these were as derived from Fig. 3. Good agreement is shown between measurements and predictions with the latter lying within the error bars imposed on the measurements due to calibration uncertainties.

V. CONCLUSIONS

It was shown that seabed reflection loss properties could be estimated from the received pulse shape, many water depths from the source. The rate of change of reflection loss with angle is closely related to the fall-off rate of the pulse and can be determined if the water depth and sound speed are known. A slight extension to the method allowed seabed sound speed to be estimated and, when combined with velocity–density regression curves, this was used to produce estimates of density and attenuation.

The seabed description derived using this method was shown to agree reasonably well with the data derived in a nearby location using a full matched-field inversion technique.

The seabed description, in terms of reflection loss gradient and the critical angle, was inserted into a propagation loss prediction that subsequently agreed (i.e., within experimental error bars) with independent measurements of propagation loss.

The method uses a single source and single-hydrophone receiver with no need for numerical inversion of the acoustic data. Uncalibrated values for received intensity can be used because of the method's sensitivity to the shape of the pulse, not its absolute amplitude. The method is complementary to standard geophysical inversion techniques and could be realized in practical sonar systems without the need for specialist arrays or intensive processing.

ACKNOWLEDGMENTS

We thank the Captain and crew of the RV Alliance and the technicians of SACLANTCEN's engineering department for data collection. The efforts of Georgios Haralabus, the scientist-in-charge of the PREBASE cruise, are also gratefully acknowledged.

Hamilton, E. L. (**1978**). ''Sound velocity–density relations in seafloor sediments and rocks,'' J. Acoust. Soc. Am. **63**, 366–377.

Hamilton, E. L. (**1980**). ''Geoacoustic modeling of the sea floor,'' J. Acoust. Soc. Am. **68**, 1313–1340.

Harrison, C. H. (**2003**). ''Closed-form expressions for ocean reverberation and signal excess with mode stripping and Lambert's law,'' J. Acoust. Soc. Am. **114**, 2744–2756.

Pouliquen, E., and Lurton, X. (**1994**). ''Identification de la nature du fond de la mer à l'aide de signaux d'écho-sondeurs:I. Modélisation d'échos rèverbérés par le fond'' (''Identification of the nature of the seabed using echo-sounder signals:I Modeling seabed reverberation echoes.''), Acta Acust. (Beijing) **2**, 113–126.

Siderius, M., Nielsen, P. L., and Gerstoft, P. (**2002**). ''Range-dependent seabed characterization by inversion of acoustic data from a towed receiver array,'' J. Acoust. Soc. Am. **112**, 1523–1535.

Smith, P. J. (**1971**). ''The averaged impulse response of a shallow-water channel,'' J. Acoust. Soc. Am. **50**, 332–336.

Weston, D. E. (**1971**). ''Intensity-range relations in oceanographic acoustics,'' J. Sound Vib. **18**, 271–287.

Pseudo-surface waves above thin porous layers (L)

Jean F. Allard, Michel Henry,[a] and Vincent Gareton
Laboratoire d'Acoustique de l'Université du Maine, UMR CNRS 6613, Avenue Olivier Messiaen, F-72085 Le Mans Cedex 9, France

(Received 19 December 2003; revised 14 June 2004; accepted 15 June 2004)

The total monopole field above thin porous layers for small source-receiver distances is a complex superposition of different contributions. It is shown that the dependence of the amplitude on the distance from the layer is exponential, close to the layer, and the same as for the surface wave related to the pole located at an angle of incidence θ_p close to $\pi/2$. The localization of this pole obtained for a thin layer from this property is compared to measurements performed with a recent transfer function method based on pressure measurements at different source-receiver distances. In this paper, we evaluate $\cos\theta_p$ from pressure measurements on an axis perpendicular to the surface of the layer at small source-receiver offsets, thus this method can be used with small sample sizes. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1778904]

PACS numbers: 43.20.El, 43.35.Pt, 43.20.Bi [MO] Pages: 1345–1347

I. INTRODUCTION

For nonlocally reacting porous layers with a motionless frame, the surface impedance Z_s depends on the angle incidence, and is given, with the time dependence $\exp(-i\omega t)$, by

$$Z_s(\theta)=i\frac{Z_1}{\phi\cos\theta_1}\cot(k_1 l\cos\theta_1). \tag{1}$$

In this equation, ϕ is the porosity, Z_1 the characteristic impedance, l the thickness, k_1 the wave number in the air saturating the medium, θ_1 the complex angle of refraction given by $\sin\theta_1=(\sin\theta)/n$, n is the refraction index, $n=k_1/k$ where k is the wave number in free air. The reflection coefficient can be written

$$V(\cos\theta)=\frac{\cos\theta-Z/Z_s(\theta)}{\cos\theta+Z/Z_s(\theta)}. \tag{2}$$

The poles are given by

$$\cos\theta_p=-\frac{Z}{Z_s(\theta_p)}. \tag{3}$$

For thin layers such that $|k_1 l|\ll 1$, one of these poles is close to $\pi/2$ as for an impedance plane with a large $|Z_s|$ and the other poles can be ignored.[1,2] The total monopole field p_t over a porous layer is the sum of the direct field $p_d=\exp(ikR_2)/R_2$ where R_2 is the source receiver distance (see Fig. 1), and the reflected field p_r. The exact expression for p_r in the Sommerfield representation[3] is an integral over ξ on the real $\sin\theta$ axis of the complex $\sin\theta$ plane

$$p_r=ik/2\int_{-\infty}^{+\infty}\frac{\xi d\xi}{\mu}V(\mu)H_0^1(\xi kr)\exp[i\mu k(z_1+z_2)],$$
$$\mu=\sqrt{1-\xi^2},\quad \mathrm{Im}\,\xi>0, \tag{4}$$

where H_0^1 is the Hankel function of the first kind, z_1 and z_2 are the the distances from the surface of the layer to the source and the receiver, respectively, and r the horizontal distance from the source to the receiver (see Fig. 1). An asymptotic expression for p_r valid for large kr and for θ_p and the angle of specular reflection θ_0 sufficently close to $\pi/2$, then for thin layers close to grazing incidence, is

$$p_r=\frac{\exp(ikR_1)}{R_1}\left[1-\sqrt{2\pi kR_1}\exp\left(\frac{3\pi i}{4}\right)\times\cos\theta_p\exp(-u^2)\mathrm{erfc}(-iu)\right], \tag{5}$$

where R_1 is distance from the image of the source to the receiver and u is the numerical distance, given by[3]

$$u=\sqrt{2kR_1}\exp\frac{\pi i}{4}\sin\frac{\theta_p-\theta_0}{2}. \tag{6}$$

For sufficiently thin layers, $\mathrm{Im}(u)<0$. At large r and $|u|$, the surface wave related to the pole at θ_p exists and can be a dominant component of the total field. The contribution p_s of the surface wave to the reflected field is given, with the same approximations as for p_r in Eq. (5), by

$$p_s=-\sqrt{\frac{8\pi k}{R_1}}\cos\theta_p\exp i(kr\sin\theta_p+k(z_1+z_2)\cos\theta_p). \tag{7}$$

Close to grazing incidence, R_1 is stationary for small variations of z_2 and the surface wave presents a characteristic exponential dependence in z_2, with an exponential decreasing $\exp(-kz_2\,\mathrm{Im}\cos\theta_p)$ when z_2 increases. This exponential decreasing at large source-receiver distances is considered as a manifestation of a dominant surface wave in the total field.[4] It is shown in the next Sec. II that at small $|u|$ the same exponential dependence of the total field can be observed close to the layer. It is also shown that it is not a manifestation of a dominant surface wave in the total field. We suggest for this case the denomination "pseudo-surface wave." Measurements of $\cos\theta_p$ are performed in Sec. III by using this property of the total field. They are compared with measurements performed with a method recently developed

[a]Electronic mail: mihenry@univ-lemans.fr

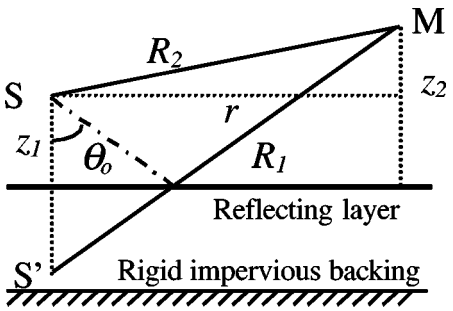


FIG. 1. The source S, its image S', and the receiver at M over the porous surface, $SM=R_2$, $S'M=R_1$, θ_0 is the specular angle of incidence.

by Allard *et al.*[1,2] All the modeling in this paper can be generalized to the case of a thin fluid layer by setting $\phi=1$.

II. VERTICAL DEPENDENCE OF THE TOTAL FIELD

The main result of this section is based on a simple expression of the derivative $\partial p_r/\partial z_2$, p_r being given by Eq. (5). The derivative of $w(u)=\exp(-u^2)\mathrm{erfc}(-iu)$ is $w'(u)=-2uw(u)+2i/\sqrt{\pi}$ (see Ref. 5, Chap. 7). At $\theta_0=\pi/2$, $\partial R_1/\partial z_2=\partial R_2/\partial z_2=0$, and $\partial\theta_0/\partial z_2=-1/r$. The derivative $\partial p_t/\partial z_2$ of the total field at $\theta_0=\pi/2$ with p_r given by Eq. (5) is

$$\frac{\partial p_t}{\partial z_2}=-\frac{\exp(ikr)}{r}\exp\frac{3\pi i}{4}\sqrt{2\pi kr}\cos\theta_p\left(-2uw(u)+\frac{2i}{\sqrt{\pi}}\right)$$

$$\times\sqrt{2kr}\exp\frac{\pi i}{4}\cos\frac{\theta_p-\pi/2}{2}\frac{1}{2r}, \tag{8}$$

where $\cos(\theta_p-\pi/2)/2\simeq 1$ because θ_p is close to $\pi/2$, and $\partial p_t/\partial z_2$ is given by

$$\frac{\partial p_t}{\partial z_2}=\frac{\exp(ikr)}{r}k\cos\theta_p\sqrt{\pi}\left(-2uw(u)+\frac{2i}{\sqrt{\pi}}\right). \tag{9}$$

By using in u the approximation $\sin(\theta_p/2-\pi/4)\simeq-0.5\cos\theta_p$, which is valid if θ_p is sufficiently close to $\pi/2$, this equation can be rewritten

$$\frac{\partial p_t}{\partial z_2}=p_t ik\cos\theta_p. \tag{10}$$

The asymptotic expression for the total pressure verifies for $z_1\ll r$ and $r_2\ll r$ the same equation as p_s. The dependence on z_2 of the total field close to $\theta_0=\pi/2$ must be given by the same factor $\exp(ikz_2\cos\theta_p)$ even at small numerical distances. At small numerical distances, the total field is a superposition of different components where the surface wave is not dominant. This can be shown by evaluating the derivative $\partial(\sqrt{r}p_t)/\partial r$ for $z_1=z_2=0$. Using similar approximations as for $\partial p_t/\partial z_2$, the derivative can be written

$$\frac{\partial(\sqrt{r}p_t)}{\partial r}=\left(ik\sin\theta_p-\frac{\exp(ikr)}{r^2p_t}\right)\sqrt{r}p_t, \tag{11}$$

instead of

$$\frac{\partial(\sqrt{r}p_s)}{\partial r}=ik\sin\theta_p\sqrt{r}p_s, \tag{12}$$

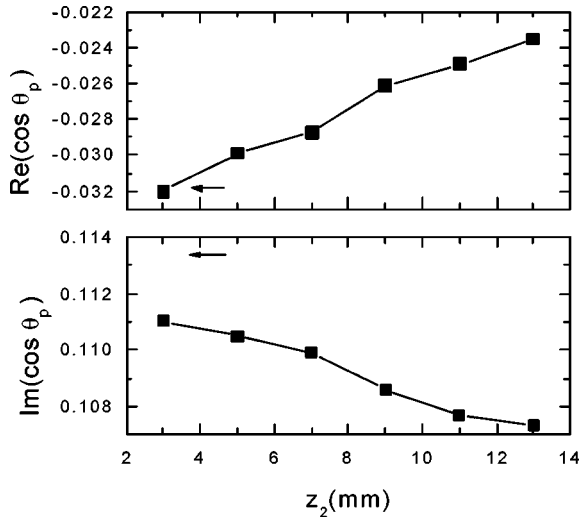


FIG. 2. Simulation of an experiment to measure $\cos\theta_p$ at 1 kHz. Arrows indicate the exact values.

for the surface wave. The term $\exp(ikr)/r^2p_t$ is negligible at large r compared to $ik\sin\theta_p$ when the surface wave gives the main contribution. This is not true at small numerical distances where the derivatives in Eqs. (11) and (12) are noticeably different but Eq. (10) is still valid. A stronger argument concerns the case where the porous layer is semi-infinite with a large flow resistivity. The pole θ_p is close to $\pi/2$ like for thin layers, and Eqs. (5) and (10) are always valid. The difference is that $\mathrm{Re}\sin\theta_p<1$. The surface wave does not exist at large distances and cannot be the main contribution at small $|u|$ because it corresponds to a signal with a horizontal velocity component larger than the sound speed.

If the dependence on z_2 remains the same as for the surface wave for z_2 sufficiently large, it is possible to evaluate $\cos\theta_p$ from pressure measurements on an axis perpendicular to the surface of the layer for small source-receiver distances on porous samples of reasonable lateral dimensions. Equation (10) does not give information concerning the maximum distance z_2 where the exponential dependence of p_t is a good approximation. Simulations have been performed for a medium of fairly high flow resistivity $\sigma=45\,000\ \mathrm{N\,m^{-4}\,s}$. Simulated evaluations of $\cos\theta_p$ from pressure measurements [the total pressure is calculated with the exact expression Eq. (4) for p_r] are performed for a source-receiver horizontal distance $r=1$ m, with a source height $z_1=1$ mm. The exact localization of the pole closest to $\pi/2$ is obtained with an iterative method described in Ref. 2. The height z_2 of the receiver varies from 1 to 13 mm by steps Δz of 2 mm. The evaluation of $\cos\theta_p$ is obtained from

$$\cos\theta_p=\frac{1}{ik\Delta z}\ln\frac{p_t(z_2+\Delta z)}{p_t(z_2)}. \tag{13}$$

Simulations are represented as a function of z_2 varying from 3 to 13 mm in Figs. 2 and 3 for a thickness $l=0.5$ cm and frequencies 1 and 3 kHz, respectively. The difference between the exact $\cos\theta_p$ and the evaluation increases with z_2, and also with l and with frequency. The exact $\mathrm{Re}\cos\theta_p$ and $\mathrm{Im}\cos\theta_p$ for $l=0.5$ cm increase from 1 to 3

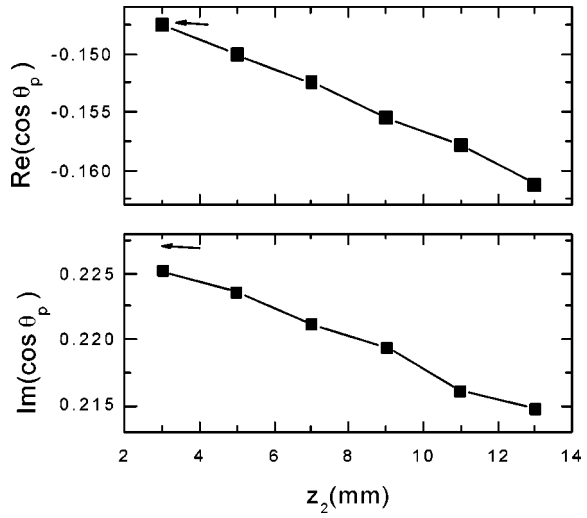


FIG. 3. Simulation of an experiment to measure $\cos\theta_p$ at 3 kHz. Arrows indicate the exact values.

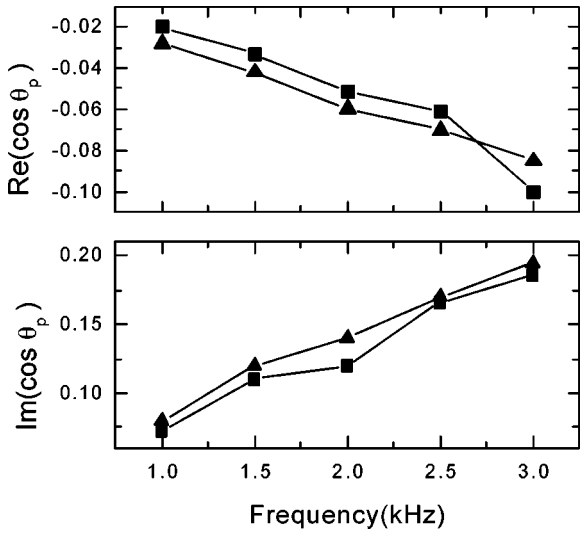


FIG. 4. Measured $\cos\theta_p$ on a foam layer of thickness $l=5$ mm (experiment). Evaluation from vertical pressure variation $z_1=3$ mm, $3\text{ mm}\leqslant z_2\leqslant 12$ mm, $r=0.75$ m (■ ■ ■), evaluation with the previous method $z_1=z_2=3$ mm, $0.2\text{ m}\leqslant r\leqslant 0.5$ m (▲ ▲ ▲).

kHz faster than this difference. The maximum relative differences at 3 kHz are close to 10/100 for $\mathrm{Re}\cos\theta_p$ and 5.5/100 for $\mathrm{Im}\cos\theta_p$ at $z_2=13$ mm.

III. MEASUREMENTS

For a foam of thickness $l=5$ mm, evaluations of $\cos\theta_p$ with Eq. (13) from pressure measurements at $r=0.75$ m have been performed in an anechoic room. The source is a driver unit feeding a pipe, the source-layer distance $z_1=3$ mm. The microphone-layer distance is always smaller than 1.2 cm, and the microphone is moved by steps of 2 mm. A reference signal is given by a fixed microphone, located close to the movable microphone to minimize the effect of the fluctuations of temperature. The mean evaluation of $\mathrm{Re}\cos\theta_p$ and $\mathrm{Im}\cos\theta_p$ obtained for different z_2 is represented in Fig. 4 as a function of frequency. At 2.5 kHz the average deviations for $\mathrm{Re}\cos\theta_p$ and $\mathrm{Im}\cos\theta_p$ at different z_2 are close to the mean evaluations multiplied by 1/10. The dispersion of the measurements at different z_2 increases at lower frequencies and it has not been possible to take into account the variation of the systematic deviation with z_2 which appears in Figs. 2 and 3. A previous method[1,2] of measuring $\cos\theta_p$ for thin layers using Eq. (5) has also been used, in which pressure measurements performed at different r over the layer and over an hard, impervious surface, are compared. Different validations of this previous method have been performed. Comparisons between predicted and measured $\cos\theta_p$ have been performed. For thin layers, θ_p is close to $\pi/2$, and comparisons have been performed between $Z_s(\pi/2)$ measured at grazing incidence with the Tamura method[6] and evaluated with Eq. (3) from measured $\cos\theta_p$. The previous method has been used with the source and the receiver at 3 mm from the layer, and the horizontal distance varying from 20 to 50 cm by steps of 10 cm. The results obtained by both methods are compared in Fig. 4. Both methods give similar results. Therefore, the vertical surface wave behavior of the total field is set in evidence experimentally. In addition to a new property of the vertical dependence of the total field close to the layer, the measurements at constant r give new direct evidence of the validity of the previous method. It may be noticed that the evaluation of $\cos\theta_p$ from pressure measurements at constant r is much more difficult than at constant z_1 and z_2 with the previous method. Pressure variations can be very small at low frequency for a variation of z_2 of some mm. Moreover in the first case the source receiver distance is larger than in the second case, and larger samples are needed.

IV. CONCLUSION

It has been shown that for thin porous layers close to grazing incidence, the total monopole field has the same dependence on the distance to the layer as the surface wave related to the pole located close to grazing incidence, even if this surface wave is not a dominant component of the total field. The localization of this pole can be obtained by a direct use of this property. Similar results have been obtained with this method and a previous method using the dependence of the reflected field on the source-receiver distance.

[1] J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurements 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 measurement near grazing incidence for non locally reacting porous layers," Acust. Acta Acust. **89**, 595–603 (2003).
[3] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media II, Point Source and Bounded Beams*, Springer Series on Wave Phenomena (Springer, New York, 1992).
[4] D. G. Albert, "Observation of acoustic surface waves in outdoor sound propagation," J. Acoust. Soc. Am. **113**, 2495–2500 (2003).
[5] M. Abramovitz and I. A. Stegun, Handbook of Mathematical Functions (Dover, New York, 1970).
[6] M. Tamura, "Spatial Fourier transform method of measuring reflection coefficient at oblique incidence," J. Acoust. Soc. Am. **88**, 2258–2264 (1990).

Time reversal and the spatio-temporal matched filter (L)

D. H. Chambers,[a] J. V. Candy, S. K. Lehman, J. S. Kallman, A. J. Poggio, and A. W. Meyer
Lawrence Livermore National Laboratory, Livermore, California 94551-9900

(Received 8 March 2004; revised 9 June 2004; accepted 12 June 2004)

It is known that focusing of an acoustic field by a time-reversal mirror (TRM) is equivalent to a spatio-temporal matched filter under conditions where the Green's function of the field satisfies reciprocity and is time invariant, i.e., the Green's function is independent of the choice of time origin. In this letter, it is shown that both reciprocity and time invariance can be replaced by a more general constraint on the Green's function that allows a TRM to implement the spatio-temporal matched filter even when conditions are time varying. [DOI: 10.1121/1.1778839]

PACS numbers: 43.60.Gk [EJS] Pages: 1348–1350

I. INTRODUCTION

For over a decade, time reversal focusing has been described as an implementation of a spatio-temporal matched filter. This was first derived by Fink[1] who showed that the fields produced by each element in a time-reversal array added coherently at the focus. Dorme[2] showed how this can be realized on reception of a signal by an array. Recently, Tanter *et al.*[3] proved that time reversal produced the optimal spatial matched filter. This work established that time invariance and reciprocity are *sufficient* for time reversal focusing to be the optimal spatio-temporal matched filter. [If $G(\mathbf{x}_2,t_2;\mathbf{x}_1,t_1)$ is the Green's function for the acoustic system, then time invariance implies $G=G(\mathbf{x}_2,\mathbf{x}_1,t_2-t_1)$, and reciprocity implies $G(\mathbf{x}_2,\mathbf{x}_1,t_2-t_1)=G(\mathbf{x}_1,\mathbf{x}_2,t_2-t_1)$].[1] A related but still unanswered question is whether both time invariance and reciprocity are *necessary* for a time reversal array to produce the optimal spatio-temporal matched filter. Though this would seem physically intuitive, we will show mathematically that time invariance and reciprocity are not strictly necessary for time reversal focusing to be the spatio-temporal matched filter. We first derive the spatio-temporal matched filter for a general (time-variant) wave system using a method similar to that of Tanter[3] and Cox.[4] This is compared with the response of a time-reversal array system to a point source in the medium, which leads to a necessary condition on the Green's function for time reversal focusing to be equivalent to the spatio-temporal matched filter. This condition is naturally satisfied for Green's functions of time-invariant wave systems that obey reciprocity. However, it is also possible to formulate Green's functions that satisfy the condition but do not satisfy reciprocity or time invariance.

II. THE SPATIO-TEMPORAL MATCHED FILTER

In signal processing texts, the matched filter is typically derived for single channel time series (see Refs. 5 and 6). A signal $u(t)$ is input into a filter with impulse response $h(t)$, resulting in an output $y(t)=h(t)*u(t)$ ($*$ indicates convolution). The input is assumed to be a combination of signal $s(t)$ and additive white noise $w(t)$ ($u=s+w$). The filter $h(t)$ is chosen to maximize the output signal-to-noise ratio at a specified time T ($SNR(T)$). The SNR is given by

$$SNR(T)=\frac{|\int_0^T h(t')s(T-t')dt'|^2}{\sigma^2\int_0^T|h(t')|^2\,dt'}, \tag{1}$$

where $E\{w(t)w^*(t+\tau)\}=\sigma^2\delta(\tau)$. The filter that maximizes $SNR(T)$ is $h(t)=s(T-t)$, which is the *matched* filter. Since the denominator of the SNR can be interpreted as the total energy in the filter response in the interval $0<t<T$,[6] the matched filter is also the $h(t)$ that maximizes the numerator given the constraint of constant energy.

For multi-channel signal processing, such as beamforming, the output $y(t)$ is given by

$$y(t)=\sum_{n=1}^{N}\int_0^t h_n(t-t')x_n(t')dt', \tag{2}$$

where $\{x_n(t):n=1,2,...,N\}$ is the set of input time series, and $\{h_n(t):n=1,2,...,N\}$ is the set of filters. In this case the SNR at time T is

$$SNR(T)=\frac{|\Sigma_{n=1}^{N}\int_0^T h_n(t')s(T-t')dt'|^2}{\sigma^2\Sigma_{n=1}^{N}\int_0^T|h_n(t')|^2\,dt'}, \tag{3}$$

where $E\{w_m(t)w_n^*(t+\tau)\}=\sigma^2\delta_{mn}\delta(\tau)$. Maximizing the SNR leads to the MMSE beamformer described by Van Trees.[7] Again, this can be interpreted as maximizing the numerator given the constraint of constant total energy for the filters.

Consider now an array of N acoustic elements (point sources) at positions $\mathbf{a}_n$ ($n=1,2,...,N$) radiating into a medium. Given the set of excitations $\{E_n(t):n=1,2,...,N\}$, the resulting acoustic field is

$$\psi(\mathbf{x},t)=\sum_{n=1}^{N}\int_0^t G(\mathbf{x},t;\mathbf{a}_n,t')E_n(t')dt', \tag{4}$$

where the (complex) Green's function $G(\mathbf{x},t;\mathbf{a}_n,t')$ specifies the response at spatial location $\mathbf{x}$ and time t to an impulse at $\mathbf{a}_n$ and time t' ($G=0$ for $t<t'$ from causality). This expres-

[a]Electronic mail: chambers2@11n1.gov

sion is analogous to Eq. (2) for the beamformer if we associate the field $\psi(\mathbf{x},t)$ with the output $y(t)$, the excitations $E_n(t)$ with the filters $h_n(t)$, and the Green's functions $G(\mathbf{x},t;\mathbf{a}_n,t')$ with the input data $x_n(t)$. Following this analogy with the beamformer, we can define a signal-to-noise ratio of the field at a specific point $\mathbf{x}=\mathbf{x}_0$ and time $t=T$:

$$SNR(\mathbf{x}_0,T)=\frac{|\psi(\mathbf{x}_0,T)|^2}{\Sigma_{n=1}^{N}\int_0^T|E_n(t)|^2\,dt}=\frac{|\Sigma_{n=1}^{N}\int_0^T G(\mathbf{x}_0,T;\mathbf{a}_n,t)E_n(t)dt|^2}{\Sigma_{n=1}^{N}\int_0^T|E_n(t)|^2\,dt}. \quad (5)$$

Physically this is the ratio of the squared magnitude of the field at $\mathbf{x}=\mathbf{x}_0$ and time $t=T$ to the total energy in the excitations in the interval $0<t<T$. (It is also the square of the functional used by Tanter *et al.*[3]).

The spatio-temporal matched filter for the field ψ at $\mathbf{x}=\mathbf{x}_0$ and $t=T$ is the set of excitations $E_n(t)$ that maximize the signal-to-noise ratio. These can be determined by applying the vector form of the Schwartz inequality (as found in Ref. 4) to $SNR(\mathbf{x}_0,T)$:

$$\begin{aligned}SNR(\mathbf{x}_0,T)&=\frac{|\Sigma_{n=1}^{N}\int_0^T G(\mathbf{x}_0,T;\mathbf{a}_n,t)E_n(t)dt|^2}{\Sigma_{n=1}^{N}\int_0^T|E_n(t)|^2\,dt}\\&\leqslant\frac{(\Sigma_{n=1}^{N}\int_0^T|G(\mathbf{x}_0,T;\mathbf{a}_n,t)|^2\,dt)(\Sigma_{n=1}^{N}\int_0^T|E_n(t)|^2\,dt)}{\Sigma_{n=1}^{N}\int_0^T|E_n(t)|^2\,dt}\\&\leqslant\Sigma_{n=1}^{N}\int_0^T|G(\mathbf{x}_0,T;\mathbf{a}_n,t)|^2\,dt. \quad (6)\end{aligned}$$

The signal-to-noise ratio is maximum when the bound becomes an equality, which occurs when $E_n(t)=G^*(\mathbf{x}_0,T;\mathbf{a}_n,t)$. This determines the excitations for the spatio-temporal matched filter. The field produced by these excitations is

$$\psi_{MF}(\mathbf{x},t)=\sum_{n=1}^{N}\int_0^t G(\mathbf{x},t;\mathbf{a}_n,t')G^*(\mathbf{x}_0,T;\mathbf{a}_n,t')dt'. \quad (7)$$

To implement the matched filter for a time-variant system, the Green's function must be known *a priori*. It cannot be obtained from direct experimental measurement because the conditions for which a measured Green's function would be valid would have changed by the time the measurement is completed. An exception might be a time-varying system that is periodic, i.e., $G(\mathbf{x},t;\mathbf{x}',t'+T')=G(\mathbf{x},t;\mathbf{x}',t')$ for some T'. If the period is known, a measured Green's function could be stored for use when conditions are repeated.

III. COMPARISON WITH TIME REVERSAL

We compare the matched filter result to that obtained by assuming the array acts as a time reversal mirror. Suppose a source at position $\mathbf{x}_0$ emits an impulse at time $t=0$. The field sampled at each array element would be $\psi(\mathbf{a}_n,t)=G(\mathbf{a}_n,t;\mathbf{x}_0,0)$. If these are recorded over the period $0<t<T$, time reversed, and emitted by the array, the resulting field would be

$$\psi_{TR1}(\mathbf{x},t+T)=\sum_{n=1}^{N}\int_0^t G(\mathbf{x},t+T;\mathbf{a}_n,t'+T)\times G(\mathbf{a}_n,T-t';\mathbf{x}_0,0)dt',\quad t>0. \quad (8)$$

This could be realized, in principle, for a time-variant wave system because it does not require prior knowledge of the Green's function. However, if one has a model of the Green's function or the conditions are repeatable, the time-reversed field could be emitted simultaneously with the source pulse. The resulting field would be

$$\psi_{TR2}(\mathbf{x},t)=\sum_{n=1}^{N}\int_0^t G(\mathbf{x},t;\mathbf{a}_n,t')\times G(\mathbf{a}_n,T-t';\mathbf{x}_0,0)dt',\quad t>0. \quad (9)$$

This second result is equal to the field produced by the matched filter [Eq. (7)] if the Green's function satisfies the condition

$$G(\mathbf{a}_n,T-t;\mathbf{x}_0,0)=G^*(\mathbf{x}_0,T;\mathbf{a}_n,t),\quad 0<t<T. \quad (10)$$

This states that the time-reversed response of the system at $\mathbf{a}_n$ to an impulse emitted at position $\mathbf{x}_0$ and time 0 is equal to the complex conjugate of the response at $\mathbf{x}_0$ and time T to an impulse emitted at position $\mathbf{a}_n$ and time t. For a time-invariant system the condition reduces to

$$G(\mathbf{a}_n;\mathbf{x}_0;t-t')=G^*(\mathbf{x}_0;\mathbf{a}_n,t-t'),\quad t>t', \quad (11)$$

which is a generalization of reciprocity to a wave system with a complex Green's function. For a real Green's function, the condition agrees with the statements of reciprocity used by Fink[1] and others.[2,3]

IV. SUMMARY

In summary, we have derived a mathematical condition [Eq. (10)] on the Green's function such that time-reversal and matched filtering are equivalent for a general time-variant acoustic system. It is not difficult to construct functions that obey this condition, for example

$$G(\mathbf{x}_1,t_1;\mathbf{x}_2,t_2)=f(\mathbf{x}_1)f^*(\mathbf{x}_2)e^{i\omega(t_1+t_2-T)}. \quad (12)$$

Thus a TRM can act as an optimal spatio-temporal matched filter for an acoustic system whose Green's function satisfies Eq. (10). Time invariance and reciprocity are not strictly necessary. Actual implementation for a time-varying system would require prior knowledge of the Green's function over the time period of interest. This would be possible for laboratory systems and natural systems where conditions repeat periodically. Analysis of specific time-varying systems that possess the required symmetry will be the subject of a future study.

ACKNOWLEDGMENTS

This work was 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.

[1] M. Fink, "Time Reversal of Ultrasonic Fields—Part I: Basic Principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[2] C. Dorme and M. Fink, "Focusing in transmit-receive mode through homogeneous media: The time reversal matched filter approach," J. Acoust. Soc. Am. **98**, 1155–1162 (1995).
[3] M. Tanter, J.-L. Thomas, and M. Fink, "Time reversal and the inverse filter," J. Acoust. Soc. Am. **108**, 223–234 (2000).
[4] H. Cox, "Optimum arrays and the Schwartz inequality," J. Acoust. Soc. Am. **45**, 228–232 (1969).
[5] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).
[6] A. Papoulis, *Signal Analysis* (McGraw–Hill, New York, 1984).
[7] H. L. V. Trees, *Optimum Array Processing* (Wiley, New York, 2002).

On the dichotomy in auditory perception between temporal envelope and fine structure cues[a] (L)

Fan-Gang Zeng,[b] Kaibao Nie, Sheng Liu, Ginger Stickney, Elsa Del Rio, Ying-Yee Kong, and Hongbin Chen

Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology-Head and Neck Surgery, University of California, Irvine, California 92697-1275

(Received 12 December 2003; revised 1 June 2004; accepted 10 June 2004)

It is important to know what cues the sensory system extracts from natural stimuli and how the brain uses them to form perception. To explore this issue, Smith, Delgutte, and Oxenham [Nature (London) **416**, 87–90 (2002)] mixed one sound's temporal envelope with another sound's fine temporal structure to produce auditory chimaeras and found that ''the perceptual importance of the envelope increases with the number of frequency bands, while that of the fine structure diminishes.'' This study addressed two technical issues related to natural cochlear filtering and artificial filter ringing in the chimaerizing algorithm. In addition, this study found that the dichotomy in auditory perception revealed by auditory chimaeras is an epiphenomenon of the classic dichotomy between low- and high-frequency processing. Finally, this study found that the temporal envelope determines sound location as long as the interaural level difference cue is present. The present result reinforces the original hypothesis that the temporal envelope is critical for speech perception whereas temporal fine structure is critical for pitch perception, but does not support the assertion regarding the temporal envelope and fine structure as the acoustic basis for the ''what'' and ''where'' mechanisms.
© *2004 Acoustical Society of America.* [DOI: 10.1121/1.1777938]

PACS numbers: 43.64.Sj, 43.66.Ba, 43.66.Hg [KWG] Pages: 1351–1354

I. INTRODUCTION

Classical dichotomies in auditory perception include pitch encoding with temporal and spectral cues (Licklider, 1951), sound localization with interaural time and level differences (Rayleigh, 1907), and listening with either the ''auditory'' or ''speech'' mode (Liberman and Mattingly, 1989). Recently, Smith *et al.* (2002) showed a new dichotomy in auditory perception between temporal envelope and fine structure cues. They first digitally filtered a wideband signal (80–8 820 Hz) into 1–64 frequency bands and then used the Hilbert transform to decompose the band-limited signal into a slowly varying envelope component and a fast varying fine-structure component. To assess relative contributions of temporal envelope and fine structure to auditory perception, they produced ''auditory chimaeras'' by mixing one sound's envelope with another sound's fine structure. By conducting listening tests with the chimaeric sounds, they found that the envelope is most important for speech reception, and the fine structure is most important for pitch perception and sound localization.

Smith *et al.*'s (2002) report has generated much interest in the field and already stimulated several new studies in speech perception (Xu and Pfingst, 2003) and in cochlear implants (Hong *et al.*, 2003; Litvak *et al.*, 2003). Here we address two technical issues in Smith *et al.*'s chimaerizing algorithm. One issue is related to an over-interpretation of the role of the fine structure cue with a few frequency bands (e.g., 1 and 2). The other issue is related to a filter ringing artifact that may have contributed to an over-interpretation of the role of the envelope cue with a large number of frequency bands (e.g., 32, 48, and 64). In addition, we take issue with the generality of Smith *et al.*'s assertion that the temporal envelope and fine structure are the acoustic basis of ''what'' and ''where'' cortical mechanisms.

II. ENVELOPES RECOVERED BY COCHLEAR FILTERING

Smith *et al.* (2002) found 70%–90% correct performance for chimaeric sounds with one- or two-band speech fine structure and noise envelope. This result was reminiscent of the classical work by Licklider and Pollack (1948) showing that infinite amplitude clipping had minimal effects on speech recognition. Because both the noise envelope in the speech-noise chimaeric sound and the envelope in the infinitely clipped speech were relatively flat, the good speech performance appeared to suggest a significant contribution of the fine structure to speech intelligibility. This suggestion is inconsistent with previous work showing that the fine structure with flattened $\frac{1}{4}$-oct band temporal envelopes contributed only about 17% speech intelligibility (Drullman, 1995). This inconsistency lies in the auditory filter's ability to recover the narrow-band speech envelope from the broad-band speech fine structure (Ghitza, 2001).

To demonstrate the auditory filter's ability to recover the narrow-band speech envelope, Fig. 1 shows the output of the gammachirp auditory filters (Irino and Patterson, 2001) in response to the original speech (top panel) as well as to the

[a]Portions of this work were presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL, 2003.

[b]Address correspondence to University of California, 364 Med Surge II, Irvine, CA 92697-1275. Electronic mail: fzeng@uci.edu

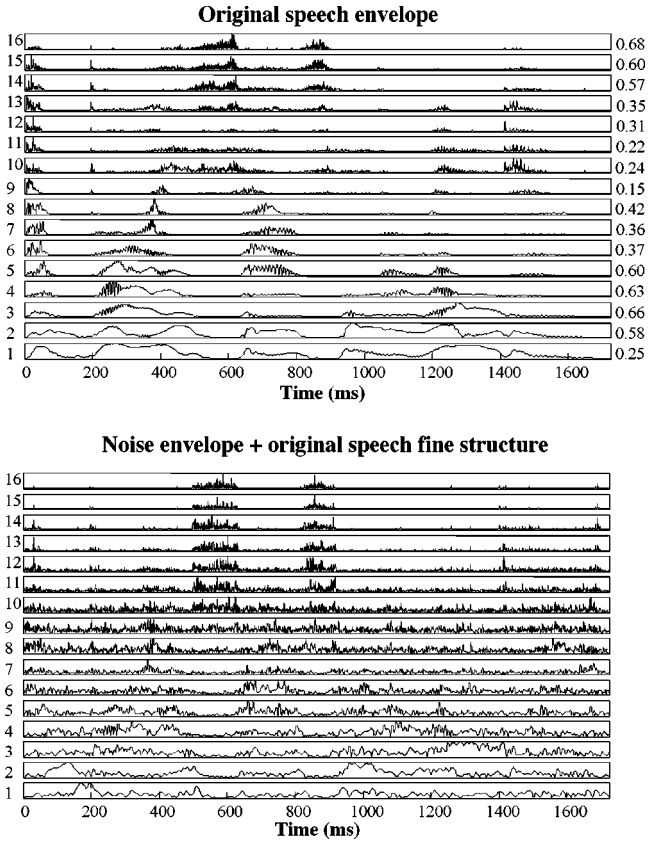


FIG. 1. Narrow-band envelopes derived from the output of 16 gammachirp filters in response to the original speech ''A boy fell from the window'' (top panel) and the one-band chimaeric sound with the speech fine structure and the noise envelope (bottom panel). Numbers displayed to the left of the panels represent different frequency bands. Numbers displayed to the right of the top panel are correlation coefficients between the original speech and the chimaeric sound's narrow-band envelopes.

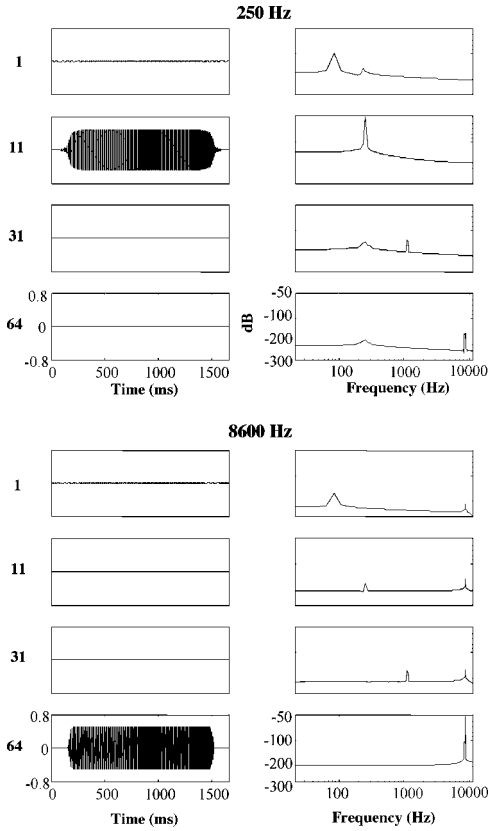


FIG. 2. Waveforms (left panels) and their Fourier amplitude spectra (right panels) at the output of selective bands (labeled by the number displayed to the left of the waveform panels) in a 64-band processor in response to two sinusoids at 250 (top panels) and 8600 Hz (bottom panels). The spectral peak at 250 Hz in the top-right four panels represents the 250-Hz stimulus component while additional peaks shifted systematically from low to high frequencies representing the ringing at each band's center frequency. Similarly, the spectral peak at 8600 Hz in the bottom-right four panels represents the stimulus while the additional peaks represent ringing artifacts.

one-band chimaerized stimulus with the noise envelope but the speech fine structure (bottom panel). The chimaeric sound's narrow-band envelopes were significantly correlated with the original speech envelopes with the mean coefficient of 0.44 from all 16 bands and a range from 0.68 at the high frequency band to 0.15 at the intermediate frequency band, suggesting that the auditory filters could at least partially recover the original narrow-band speech envelope from the one-band speech fine structure. To quantify the contribution of the recovered narrow-band speech envelopes, we additionally amplitude modulated them to noise having the same bandwidth as the original gammachirp filters (Shannon *et al.*, 1995) and measured their performance using the HINT sentences (Nilsson *et al.*, 1994) in four normal-hearing subjects. The mean percent correct score was 79% (SD=9%) for the original one-band chimaerized sound, similar to the 70% score obtained in the Smith *et al.* study. However, we found a mean score of 40% (SD=12%) for the additionally synthesized sound, suggesting that the recovered envelope could account for at least half of the performance from the one-band chimaerized sound that contained the speech fine structure and the noise envelope. Therefore, taking cochlear filtering into account, the present result has effectively removed the number of bands as a significant factor and reinforced the original idea that the temporal envelope is critical for speech recognition (Van Tasell *et al.*, 1987; Drullman, 1995; Shannon *et al.*, 1995).

III. ARTIFICAL FINE STRUCTURE INTRODUCED BY FILTER RINGING

Smith *et al.* (2002) observed a contribution of the envelope to pitch recognition only with a large number of frequency bands (e.g., 48 and 64). We found this result counter intuitive. Consider the following simplistic case in pitch perception, namely chimaerizing two sinusoids. In this case, the Hilbert envelope of a sinusoid is the amplitude and its Hilbert fine structure is the cosine of the phase of the analytic signal $[\cos(2\pi ft)]$, with frequency being the only difference between the two sinusoids. To the extent that pitch is related to frequency, amplitude (i.e., temporal envelope) should not determine the pitch of a sinusoid.[1] Why then could Smith *et al.* come to the conclusion that pitch is determined by the envelope with the large number of bands?

Figure 2 shows the waveform and its Fourier spectrum at the output of a few selected bands (the total number of bands=64) in response to a sinusoid at either 250 Hz (top panels) or 8600 Hz (bottom panels). In the top panels, the 11th band had a center frequency of 258 Hz and expectedly produced the largest response to the 250-Hz sinusoidal

stimulus. The other bands' responses were much smaller (hardly seen in the waveform) but contained both the leaked stimulus at 250 Hz (the peak aligned to that in the 11th band) as well as the ringing located at the filter's center frequency (the additional peak whose position systematically increased with the band number). The same pattern can be observed for the filter responses to the 8600-Hz sinusoidal stimulus.

When chimaerizing between the two sinusoids, the large envelope in the band that contained the stimulus (e.g., the 64th band containing the 8600 Hz sinusoid) would be combined with the fine structure in the same band in response to the 250-Hz sinusoid, which had a dominant component at the ringing frequency of 8565 Hz. This ringing frequency was close enough to 8600 Hz, resulting in an impression that the envelope determines pitch when the number of frequency bands is large as reported by Smith *et al.* (2002). With a large number of bands, the difference is small between the ringing frequency and the original signal frequency, making the differentiation between Smith *et al.*'s envelope argument and the present ringing argument subtle in a practical sense. However, the theoretical differentiation is important because the ringing is primarily determined by the filter properties, such as the center frequency and the bandwidth, having nothing to do with the original signal; we suggest that the data showing dependence of pitch perception on the temporal envelope obtained by Smith *et al.* at these large frequency bands (32–64) are an artifact of filter ringing.

IV. "WHERE" VERSUS "WHAT": A NEW OR OLD DICHOTOMY?

Smith *et al.* (2002) showed a dichotomy in speech recognition and sound localization by presenting chimaeric speech sounds with either a 700-μs delayed envelope or fine structure. With the 16- and 32-band conditions (their Fig. 4), they clearly demonstrated that the envelope cue is important for speech recognition and the fine-structure cue is important for sound localization, revealing a possible acoustic basis for the hypothesized "what" and "where" pathways in the auditory cortex (Tian *et al.*, 2001). This result is provocative but lacks the necessary generality to support its assertion.

First, to examine what acoustic cues were responsible for the observed dichotomy in the Smith *et al.* results, we used an identical algorithm and procedure to measure the intelligibility and localization of an exemplary condition in the Smith *et al.* (2002) study, in which one sentence's envelope was mixed via a 16-band chimaerizer with another sentence's fine structure that had a 700-μs ITD (leading in the right ear). In addition, we processed those stimuli by the classic low- and high-pass filtering (sixth-order Butterworth filters). Figure 3 shows averaged lateralization (top panel) and intelligibility (bottom panel) data as a function of the cutoff frequency for low-pass (filled circles) and high-pass (open triangles) filtering of the chimaeric sound in three normal-hearing listeners. Note first in the low-pass condition that the subjects used mostly the ITD cue in the fine structure to lateralize to the right ear. Note second in the high-pass condition that the subjects did not use the ITD cue in the fine structure but rather the ILD cue in the envelope to lateralize the sound (i.e., a center position in the head). This was true

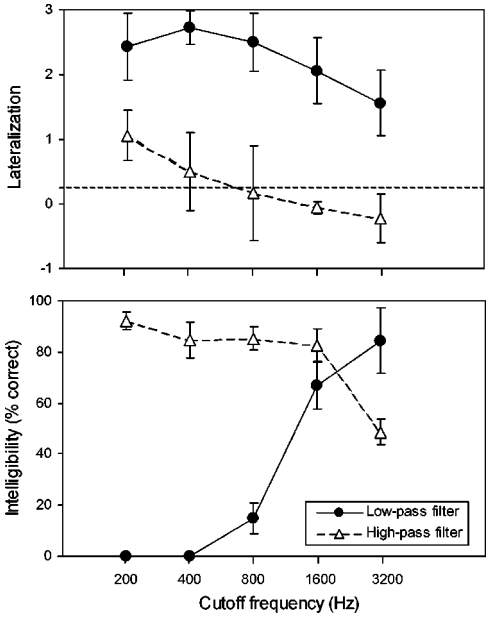


FIG. 3. Lateralization (top panel) and intelligibility (bottom panel) results from 16-band chimaerized sounds as a function of cutoff frequency for low-pass (filled circles) and high-pass (open triangles) filters. The averaged data from three normal-hearing subjects are represented by symbols while the error bars represent plus or minus one standard deviation. In the top panel, the "3" on the *y* axis represents the subject's complete lateralization to the right ear while the "0" represents the center position. The intelligibility scores are a percentage of the key words correctly identified.

even with the 200-Hz high-pass condition, in which the subjects only slightly lateralized (scale=1) to the right ear. A repeated-measure ANOVA confirmed that the difference between the low- and high-pass conditions was significant [$F(1,2)=135.32; p<0.01$]. This result suggests that lateralization using the ITD cue in the present chimaeric sound is essentially a low-frequency phenomenon.

Note a totally different pattern of results for the intelligibility data, in which those low-frequency components (<800 Hz) that dominated lateralization contributed essentially nothing to speech intelligibility (see the three leftmost filled circles representing 200-, 400-, and 800-Hz low-pass cutoff frequencies). On the contrary, it was the high-frequency components that contributed to speech intelligibility. Thus the present low-pass and high-pass filtering data clearly demonstrate that the perceptual dichotomy between the temporal envelope and fine structure cues observed in the Smith *et al.* (2002) study was due to the well-known difference between the use of the low-frequency ITD cue for lateralization and the use of the high-frequency cue for intelligibility (Rayleigh, 1907; French and Steinberg, 1947; Oppenheim and Lim, 1981).

Second, to directly test the generality of Smith *et al.*'s assertion relating the temporal envelope and fine structure to the "what" and "where" mechanisms, we constructed two chimaeric stimuli that contained either a 15-dB ILD cue or conflicting ITD and ILD cues extracted from head-related transfer functions.[2] The ILD cue reflects an overall level difference that is inherently embedded in the temporal envelope, whereas the ITD cue in naturally embedded in the temporal fine structure. Figure 4 shows the effect of the ILD cue on lateralization with the ITD cue towards either left (top

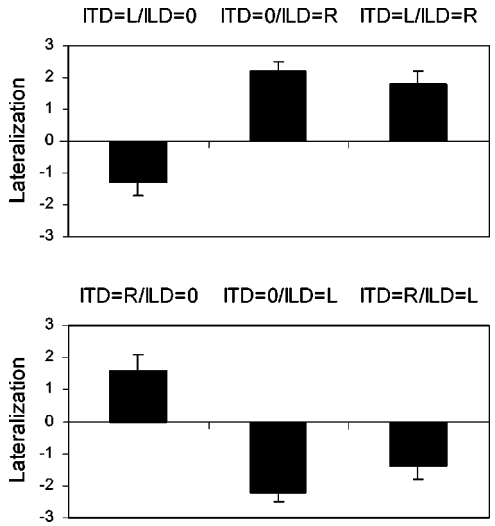


FIG. 4. Lateralization result for the ITD-only cue (left bars), the ILD-only cue (middle bars), and the conflicting ITD and ILD cue (left bars). The top panel shows results from a condition where the ITD cue is either left or neutral whereas the bottom panel shows the result from a condition where the ITD cue is either right or neutral. The data are averaged from five normal-hearing subjects with the error bars representing plus or minus one standard deviation.

panel) or right (bottom panel). Three experimental conditions included the original ITD-only control in Smith *et al.*'s study (left bars), the ILD-only condition (middle bars), and the conflicting ITD and ILD condition (right bars). Although the conflicting ITD cue slightly reduced the ILD-alone effect (0.5 to 1 on the −3 to 3 scale used), the result clearly showed that whenever the ILD cue is present, the temporal envelope, rather than the fine structure, largely determines sound location.

V. FINAL REMARKS

Smith *et al.* (2002) presented an innovative algorithm (''auditory chimaeras'') to assess the relative contribution of temporal envelope and fine structure to auditory perception. Here we addressed two technical issues that one should be aware of when using their chimaerizing algorithm and in interpreting the results derived from auditory chimaeras. First, one should be aware of the ear's natural ability to recover the narrow-band envelope with broad-band processing for a small number of frequency bands (e.g., 1 and 2). Second, one should be concerned about filter ringing artifacts with narrow-band processing for a large number of bands (e.g., 32, 48, and 64). In addition, we performed a classic filtering manipulation on the chimaerized sounds and found clear evidence suggesting that the dichotomy revealed by the auditory chimaeras is an epiphenomenon of classic duplex perception between low- and high-frequency pathways. Finally, we provided two counter examples showing that the temporal envelope via an ILD cue determines sound location for both conditions where the ITD cue was either neutral or in conflict with the ILD cue. The present result reinforces the previously proposed hypothesis that the temporal envelope is critical for speech recognition whereas the temporal fine structure is critical for pitch perception. However, the present result does not support the assertion regarding the temporal envelope and fine structure as the acoustic basis for ''what'' and ''where'' mechanisms.

ACKNOWLEDGMENTS

We thank Bertrand Delgutte, Andrew Oxenham, Zachary Smith, Ken Grant, Robert Shannon, Van Summers, and an anonymous reviewer for helpful comments on earlier versions of this manuscript. We also thank Smith, Delgutte, and Oxenham for providing the chimaerizing algorithm, Ruth Litovsky for providing the head related transfer function, and G. Bruce Henning for referencing to Oppenheim and Lim's study. Preparation of this manuscript was supported by a grant from the National Institutes of Health (RO1-DC-02267).

[1]It is noted that actual pitch perception of a pure tone is dependent on frequency as well as amplitude (Stevens, 1935). However, this amplitude-dependent change in pitch is generally small (~10%) and requires a large change in amplitude (~40 dB).

[2]The head-related transfer function (HRTF) was provided by Ruth Litovsky and recorded in a human manikin. One sentence was convolved with a HRTF recorded from the right ear whereas another sentence was convolved with a HRTF recorded from the left ear. Chimearizing these two HRTF-convolved sentences produced the stimuli used in Fig. 4, in which they contained conflicting ITD and ILD cues.

Drullman, R. (**1995**). ''Temporal envelope and fine structure cues for speech intelligibility,'' J. Acoust. Soc. Am. **97**, 585–592.

French, N. R., and Steinberg, J. C. (**1947**). ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119.

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

Hong, R. S., Rubinstein, J. T., Wehner, D., and Horn, D. (**2003**). ''Dynamic range enhancement for cochlear implants,'' Otol. Neurotol. **24**, 590–595.

Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.

Liberman, A. M., and Mattingly, I. G. (**1989**). ''A specialization for speech perception,'' Science **243**, 489–494.

Licklider, J. C. R. (**1951**). ''A Duplex Theory of Pitch Perception,'' Experientia **7**, 128–134.

Licklider, J. C. R., and Pollack, I. (**1948**). ''Effects of differentiation, integration, and infinite peak clipping on the intelligibility of speech,'' J. Acoust. Soc. Am. **20**, 42–51.

Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003**). ''Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains,'' J. Acoust. Soc. Am. **114**, 2079–2098.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). ''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.

Oppenheim, A. V., and Lim, J. S. (**1981**). ''The importance of phase in signals,'' Proc. IEEE **69**, 529–541.

Rayleigh, L. (**1907**). ''On our perception of sound direction,'' Philos. Mag. **13**, 214–232.

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

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

Stevens, S. (**1935**). ''The relation of pitch to intensity,'' J. Acoust. Soc. Am. **6**, 150–154.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (**2001**). ''Functional specialization in rhesus monkey auditory cortex,'' Science **292**, 290–293.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.

Xu, L., and Pfingst, B. E. (**2003**). ''Relative importance of temporal envelope and fine structure in lexical-tone perception,'' J. Acoust. Soc. Am. **114**, 3024–3027.

Plane wave reflection coefficient from near field measurements

Chris H. Harrison[a] and Peter L. Nielsen
NATO Undersea Research Centre, Viale S. Bartolomeo 400, 19138 La Spezia, Italy

(Received 19 February 2004; revised 10 May 2004; accepted 18 June 2004)

The plane wave reflection coefficient for all angles and frequencies is a unique descriptor of the acoustic properties of a seabed, but it is difficult to measure directly in shallow water. However, one can calculate the point receiver result by a plane wave expansion given the plane wave reflection coefficient. In this paper it is shown that it is also possible to perform the more useful inverse operation of calculating the plane wave reflection coefficient given point source and receiver measurements. This means that it is possible to make propagation predictions in the presence of complex bottom structure using short range reflection measurements but without ever determining the detailed geoacoustic parameters. Simulations show that the method works for multiple isospeed layers and also continuously varying sediment sound speed profiles. The feasibility of experimental determination of the plane wave reflection coefficient is addressed. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781328]

PACS numbers: 43.20.El, 43.30.Ma [AIT] Pages: 1355–1361

I. INTRODUCTION

A common problem in geoacoustic inversion is that of ambiguity, namely that more than one set of geoacoustic parameters may fit the propagation measurements. In contrast, the plane wave reflection coefficient for all angles and frequencies is a unique descriptor of the acoustic properties of a seabed, but it is difficult to measure directly with point source measurements or narrow beam measurements (Schmidt and Jensen, 1986). Brekhovskikh (1980) showed the relationship between the spherical wave reflected from a half space and the plane wave reflection coefficient by using the Sommerfeld integral (see, e.g., Jensen *et al.*, 1994, Eqs. 2.104, 2.51) to express the reflection as a spectrum of plane waves. The resulting integrand is highly oscillatory although there is a stationary phase point corresponding to the angle of the Fermat ray path.[1] Making use of this behavior one can apply the method of steepest descents to find approximate solutions to the integral under various conditions, and Brekhovskikh (1980) produced several useful formulas. Westwood (1989) took this approach further by allowing the behavior of the complex plane wave reflection coefficient to shift the stationary phase point and thereby provide a more faithful solution in the presence of phase changes (as are common with interfering reflections from two or more boundaries).

In this paper, rather than searching for better approximations, we demonstrate that the plane wave and spherical wave reflection coefficients (as defined below) for a layered medium are related by a Hankel (or Fourier–Bessel) transform. Furthermore, one can calculate the plane wave reflection coefficient directly, given the spherical reflected wave at all angles, or at a set of ranges for fixed depth. In principle this transformation could be applied to the ''move-out'' type of measurement as used, for instance, by Holland and Osler (2000). The resulting (complex) plane wave reflection coefficient could be used to predict propagation regardless of the complexity of the geoacoustics. By simulating such a shallow water experimental arrangement we investigate this transformation numerically with three types of layered seabed.

II. REFLECTION FROM A HALF-SPACE

Brekhovskikh's derivation for a sediment half-space below an infinitely deep water column is summarized as follows. The field from a point source in free space can be expressed as a plane wave expansion through the Sommerfeld integral

$$\frac{\exp(ikR)}{R} = i\int_0^\infty \frac{\exp(\pm i\gamma|z|)}{\gamma} J_0(Kr)KdK, \tag{1}$$

where the horizontal and vertical wavenumbers K and γ are related to the wavenumber in the medium k by $k^2=K^2+\gamma^2$, and slant range R is related to horizontal range r and depth z by $R^2=r^2+z^2$. Therefore the part of the field in the water column due to the source image in the seabed alone is

$$\psi(r,z,z_s) = V_S(r,z,z_s)\frac{\exp(ikR)}{R}$$

$$= i\int_0^\infty V_P(K)\frac{\exp(i\gamma(z+z_s))}{\gamma} J_0(Kr)KdK, \tag{2}$$

where we have adopted Brekhovskikh's convention of z pointing upwards with origin at the seabed, and $|z_s|$ is the source height $[R^2=r^2+(z+z_s)^2]$. The plane wave reflection coefficient (PWRC) is V_P, and we define the spherical wave reflection coefficient (SWRC) as V_S so that it excludes the main spherical spreading and travel time. Note that the integration is along the real K axis from zero to infinity. By making the substitution $K=k\cos\theta$ the integral can also be written in terms of grazing angle [or incident angle as in Brekhovskikh (1980)].

The integral in K [Eq. (2)] can be recognized as a Hankel transform. If we rearrange the terms and take Hankel transforms of both sides, we obtain the transform pair

[a]Electronic mail: harrison@saclantc.nato.int

$$V_S(r,z,z_s)=i\int_0^\infty V_P(K)\,\frac{\exp(i\gamma(z+z_s))}{\gamma}J_0(Kr)KdK \times\left\{\frac{\exp(ikR)}{R}\right\}^{-1}, \tag{3}$$

$$V_P(K)=-i\int_0^\infty V_S(r,z,z_s)\,\frac{\exp(ikR)}{R}J_0(Kr)rdr \times\left\{\frac{\exp(i\gamma(z+z_s))}{\gamma}\right\}^{-1}, \tag{4}$$

which displays remarkable symmetry. With the second equation we obtain the PWRC $V_P(K)$ in terms of a given (complex) SWRC $V_S(r,z,z_s)$. This is more directly applicable to the "move-out" type of reflection measurements provided there is some (relative) phase information, for instance, a measured phase or travel time relative to the direct path. Notice that in the first integral r, z, z_s, R are constant while K (and therefore γ) varies, whereas in the second integral K, γ, z, z_s are constant while r (and therefore R) varies.

Given a set of experimental reflection measurements with horizontally moving source and fixed receiver, a procedure would be to apply Eq. (4) to find the plane wave reflection coefficient. One could then either apply the PWRC to ray trace propagation programs without change, or convert it to geoacoustic parameters using standard inversion techniques.

III. REFLECTION FROM MULTIPLE LAYERS/ ARBITRARY LAYERING

One might object that the above finding is only valid for a bottom half-space since it appears from Brekhovsikh's approach to rely on the image representation and the Sommerfeld integral. In reality, for a horizontally separated source and receiver, reflections from deep layers will be steeper than reflections from shallow layers. One could counter the objection by arguing that in a multi-layer environment each boundary has its own image, however, this view point is made complicated by partial reflection and refraction in the intermediate layers.

The more general approach in Jensen *et al.* (1994) assumes a water half-space containing the source and receiver above a multi-layered seabed. A harmonic source results in a three-dimensional Helmholtz equation which reduces to one vertical dimension after taking a horizontal Hankel transform (Jensen *et al.*, 1994, pp. 206–208; Morse and Ingard, 1968). Assuming each layer to have a uniform sound speed and density, the vertical Green's function can then be written in terms of up- and downgoing waves in each layer. The layer containing the source has an additional term of the form (Jensen *et al.*, 1994, Eq. 4.13)

$$\hat{\phi}(K)=\frac{i\exp(-i\gamma(z-z_s))}{\gamma}. \tag{5}$$

Note that in Jensen *et al.* (1994) a unit source S_ω is given by $S_\omega=-4\pi$, (see his Eqs. 4.12, 4.3).

In the usual derivation we would arrive at the scalar potential and hence the acoustic pressure by taking the Bessel transform of this quantity. This leads to the theory adopted in the model OASES (Schmidt, 1999) which we use later. Instead, here we use the source term to write the continuity conditions at the bottom of the water layer in terms of the unknown SWRC V_S and the up- and down going coefficients in the first bottom layer A_1, B_1. (The continuity conditions at the other boundaries are irrelevant for this argument.)

Continuity of pressure:

$$i\rho_0\exp(-i\gamma_0 z_s)/\gamma_0+\rho_0 V_S=\rho_1 A_1+\rho_1 B_1. \tag{6}$$

Continuity of particle velocity:

$$i\gamma_0\exp(-i\gamma_0 z_s)/\gamma_0-\gamma_0 V_S=\gamma_1 A_1-\gamma_1 B_1. \tag{7}$$

Comparing this with the equivalent continuity conditions for a plane wave with unit incident amplitude and reflection coefficient V_P and coefficients in the first layer A_1', B_1' we have the following

Continuity of pressure:

$$\rho_0+\rho_0 V_P=\rho_1 A_1'+\rho_1 B_1'. \tag{8}$$

Continuity of particle velocity:

$$\gamma_0-\gamma_0 V_P=\gamma_1 A_1'-\gamma_1 B_1'. \tag{9}$$

For each horizontal wavenumber K we see that these two sets of equations are identical if we put

$$V_P=-iV_S\gamma_0\exp(-i\gamma_0 z_s) \tag{10}$$

and similarly

$$A_1'=-iA_1\gamma_0\exp(-i\gamma_0 z_s) \quad \text{and}$$

$$B_1'=-iB_1\gamma_0\exp(-i\gamma_0 z_s).$$

The changes to A_1', B_1' make no difference at all since they are dummy variables. However, when we Hankel transform back to the radial coordinate r, we obtain the spherical wave

$$\psi(r,z,z_s)=\int_0^\infty V_S(K)\exp(i\gamma_0 z)J_0(Kr)KdK, \tag{11}$$

where $k_0^2=\gamma_0^2+K^2$, and, substituting with Eq. (10),

$$\psi(r,z,z_s)=i\int_0^\infty V_P(K)\,\frac{\exp(i\gamma_0(z+z_s))}{\gamma_0}J_0(Kr)KdK. \tag{12}$$

This is identical to Eq. (2) despite the multi-layered environment. Thus we have proved that even though the angles to deeper layers are steeper than to shallow layers, by the time the integral over wavenumber K (or vertical angle) is completed the net difference is zero. The angles, though different for one source position, cover the same angle range in the integral—in effect they are dummy variables. Therefore the inversion of spherical wave reflection measurements to find plane wave reflection coefficient is still valid for multiple layers.

From the above discussion of boundary conditions at the seabed it is apparent that one could also formulate a wavenumber-integration (or fast field) propagation model in

terms of the plane wave reflection coefficient *instead of* geoacoustic parameters. This has the important implication in naval sonar that one could predict sound propagation faithfully using the PWRC (derived from SWRC) without ever measuring or inverting for the geoacoustic properties of the bottom layers.

IV. APPROXIMATIONS

Numerical integration of Eqs. (3) and (4) is nontrivial because the integrand is highly oscillatory. It is well known that the main contribution in each comes from the vicinity of one or more stationary phase points, and in fact both integrals can be written (to a very good approximation away from the end points) in the form of a convolution with a Fresnel-type term for each stationary phase point. This is shown explicitly by writing the Bessel function of Eq. (3) in its asymptotic form, changing variable from K to $\theta = \mathrm{a}\cos(K/k)$, then changing again to the difference δ between this (plane wave) angle and the sonar geometric angle θ_o, i.e., $\theta_o = \mathrm{a}\tan((z+z_s)/r)$ and $\theta = \theta_o + \delta$. Thus

$$V_S(r,z,z_s) = \int \exp(-ikR\delta^2/2) V_P(\theta_o+\delta)\sqrt{\cos(\theta_o+\delta)}\,d\delta \times \mathrm{e}^{i\pi/4}\sqrt{\frac{kR}{2\pi\cos\theta_o}}. \quad (13)$$

Similarly Eq. (4) can be reduced to

$$V_P(\phi) = -i\int \exp(ikZ\chi^2/2) V_S(r)\sqrt{\sinh(\xi_o+\chi)}\,d\chi \times \mathrm{e}^{i\pi/4}\sqrt{\frac{kZ}{2\pi\cot\phi}}, \quad (14)$$

where $\xi = \xi_o + \chi$, $r = (z+z_s)\sinh\xi$, $Z = (z+z_s)\sin\phi$ and $\xi_o = \mathrm{a}\sinh(\cot\phi)$. By inspection, if the Fresnel term is narrow compared with the other terms (large number of wavelengths between source and image receiver), then V_S and V_P are equal. Equation (13) was the subject of several approximate analytical solutions by Brekhovskikh (1980) and Westwood (1989). One can find similar solutions for Eq. (14), however they will not be pursued here. Rather, the equations are presented to highlight the sensitivity of V_P to ''edges'' and phase curvature in V_S, and vice versa. Clearly the stationary phase point can be moved by the existence of a linear phase in either V. Spikes or discontinuities in either V can obviously produce significant effects by disrupting the balance of the positive and negative contributions from the exponential term.

V. EXAMPLES

To demonstrate the integrity of Eqs. (3) and (4) we take three example seabeds (with the sea surface removed to infinity) each viewed two ways; first from afar—in plane wave geometry, then in ''move-out'' geometry, i.e., receiver at fixed height above the seabed, and source moving at fixed height horizontally from zero to infinity. Although we have removed the sea surface we have shallow water geometry in mind where experimentally one would extract the bottom return by time gating. We simulate the first case with OASR and the second using OASES (Schmidt, 1999) as a starting point. We then use a numerical implementation of Eq. (4) to calculate PWRC from the simulated SWRC (move-out) data, and Eq. (3) to calculate the SWRC from the PWRC data. The results of the integral and the simulation are then superimposed in subsequent plots.

In a simulation, as opposed to an experiment, one could simply subtract the direct path contribution $[\exp(ikR)/R]$ from the unmodified OASES output to obtain the reflected path alone. Here, instead, we mimic a move-out experiment by performing a broad band calculation, running OASES many times, then separating the direct and bottom reflected portions in the time domain. Thus the results also demonstrate the feasibility of deriving a PWRC in an experiment. The plots correspond to the bottom reflected part alone at 500 Hz.

OASES produces a complex amplitude at one frequency for each receiver depth and range. The band from 10 to 1000 Hz was covered in 1-Hz steps, and these were zero padded and inverse Fourier transformed to form a time series. As far as possible, geometry was chosen so that the bottom paths did not overlap any other path in time so that separation from the direct path was straightforward. To mimic an experiment the time-separated direct and bottom returns were individually Fourier transformed and normalized according to their respective nominal slant ranges by $\exp(ikR)/R$. The ratio of these two quantities at the center band (500 Hz) is V_S. Surprisingly, such time gating does not spoil the spectrum of V_S or V_P, for the following reason. The bottom reflection is causal, so regardless of its spectrum it cannot exist before a certain arrival time, although it can have a long tail. The direct path which arrives earlier is also causal and, provided the water is not dispersive, it can have a tail as short as the transmit pulse. Therefore it is possible to arrange the geometry so that the complete time response consists of the *sum* of direct and bottom returns but there is definitely no overlap in time between them. Because, in this special case, time gating can be expressed as a *sum*, the process of extraction is linear and so the FT of the bottom pulse extracted by time gating is indeed the FT of the bottom reflection alone.

A. Case 1: Three uniform layers

The first example has three layers (water, sediment, and basement) with sound speeds 1500, 1600, and 1700 m/s. Otherwise the 4-m sediment layer and the basement have the same properties, namely, specific gravity 1.4, and absorption 0.14 dB per wavelength. The source and receiver are at 80 and 129.75 m above the seabed, notionally 0.25 m below the sea surface in water 130 m deep although the reflecting sea surface has been removed in these tests. OASR (PWRC) is shown in Fig. 1 and OASES (SWRC) in Fig. 2, each in the form of a loss $[\mathrm{RL} = -20\log_{10}(\mathrm{V})$, whether plane or spherical wave]. (The missing section for small angles in OASES is caused by the difficulty of time separating the direct path.) Note first that the two curves are very different, with loss peaks of different magnitude and location and occasional negative SW reflection loss. Nevertheless, the transformed

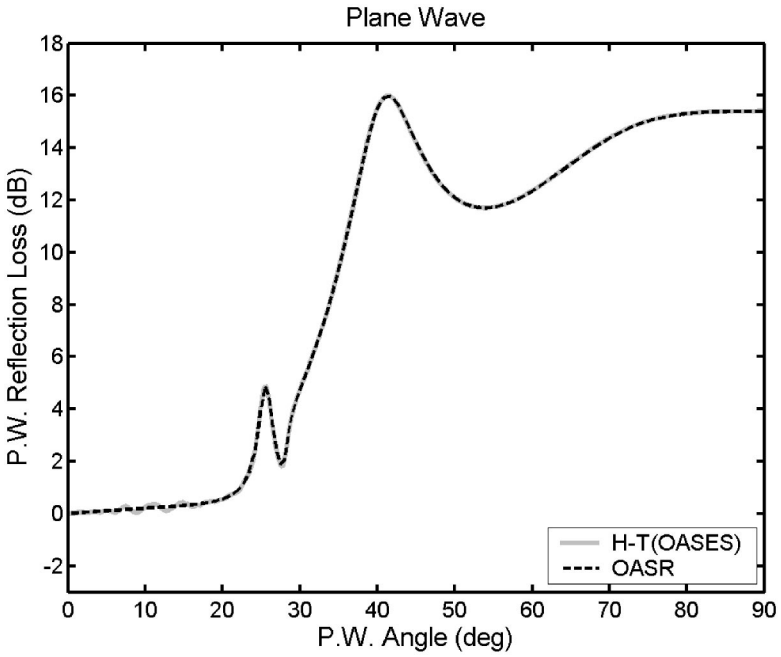


FIG. 1. Case 1: three-layer. Plane wave reflection loss from OASR (black dashed) and Hankel transformed bottom return extracted from OASES by time domain separation of bottom paths (gray solid).

OASR [labeled H-T(OASR)] agrees with OASES in Fig. 2 to within a line thickness. Similarly the transformed OASES (having patched the missing end section with separate linear interpolation of phase and amplitude) agrees with OASR in Fig. 1. The numerical transform of the transform, e.g., of OASR, is not shown since it is indistinguishable from the original in all examples, as it should be.

B. Case 2: Continuously varying sound speed

In the second example we introduce a linear sound speed in the middle layer and extend it to be 100 m thick with the aim of forming a caustic in the water column, i.e., a pronounced difference between SWRC and PWRC. This type of environment with its resulting ''negative bottom losses'' was investigated by Stickler (1977). Sound speeds in the top and bottom layers are 1500 and 1800 m/s, and speeds are continuous across the boundaries. In all layers the specific gravity is 1.0 and the attenuation is zero. Source and receiver are both at 15 m above the seabed. The modulus of the PWRC (OASR) in Fig. 3 looks rather uninteresting except for the pronounced critical angle at 34°, but hidden is the fact that the phase increases more or less as the cube of angle (Ainslie *et al.*, 2004), thus exhibiting curvature and interacting strongly with the Fresnel term shown explicitly in Eq. (13). The transformation and the OASES plot, with which it agrees, show the caustic in the form of the negative loss of almost 10 dB at 5.3° in Fig. 4. Again it is possible to transform the OASES solution back to a PWRC as shown in Fig. 3, and we note the extreme disimilarity of the PWRC and SWRC if taken at face value.

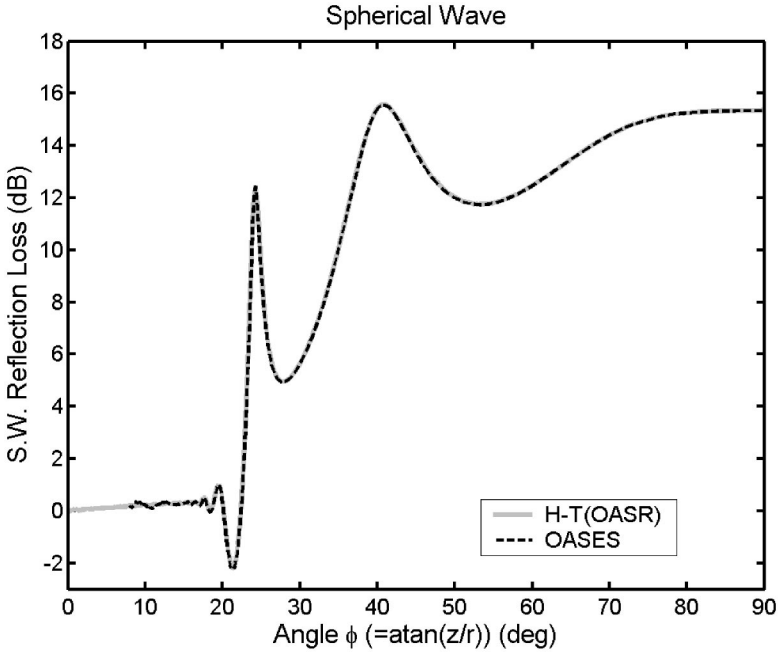


FIG. 2. Case 1: three-layer. Spherical wave reflection loss extracted from OASES by time domain separation of bottom paths (black dashed) and Hankel transformed OASR (gray solid).

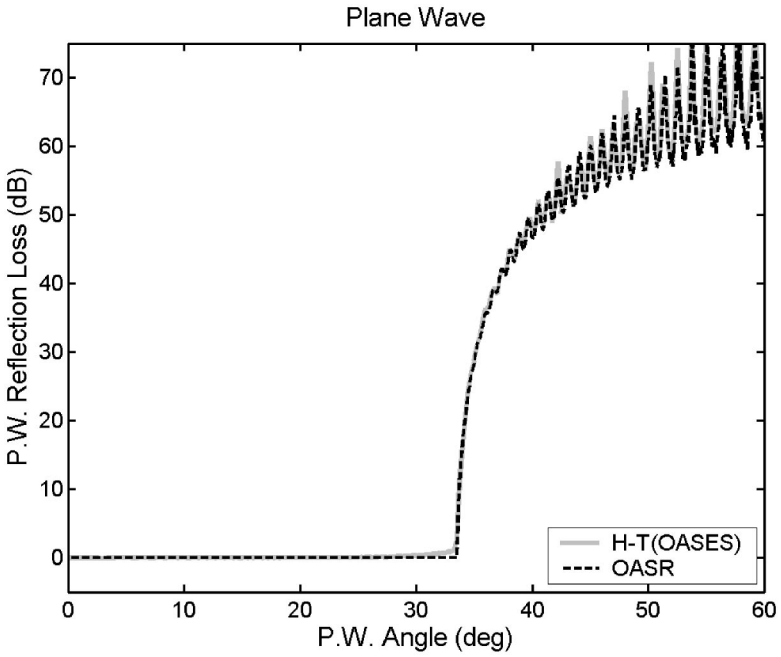


FIG. 3. Case 2: uniform gradient. Plane wave reflection loss from OASR (black dashed) and Hankel transformed bottom return extracted from OASES by time domain separation of bottom paths (gray solid).

C. Case 3: 36 layers

The third example is the test case ''Na'' from the Geoacoustic Inversion Workshop'97 (Tolstoy *et al.*, 1998). The 36 layers have various thicknesses spanning 55 m. Sound speeds increase from about 1500 to 1700 m/s and specific gravities from 1.4 to 1.8. Attenuation is constant at 0.23 dB per wavelength. Sonar geometry is the same as case 1. Despite the extra detail the agreement between PWRC (OASR) and the transformed OASES in Fig. 5 is still extremely good except, understandably, near the patched low angles. In Fig. 6 agreement of the transformed OASR with OASES is excellent.

The examples in these figures were chosen to demonstrate some features that one might guess to be difficult to measure or invert experimentally. Certainly it would be hard to predict correctly the four or so properties for each of the 36 layers in case 3 by standard inversion techniques. Therefore (notwithstanding arguments concerning parameter correlation and ambiguity) one might reasonably deduce that it would be almost impossible to get correct and reliable propagation predictions by reusing these 144 geoacoustic parameters in a propagation model. On the contrary, these plots emphasize that one definitely can get reliable propagation predictions through the PWRC, having first measured the SWRC, regardless of the complexity of the geoacoustics, or even with no knowledge of the geoacoustics at all.

Another important consideration is the effect of changing the sound speed in the water. If we keep it uniform but alter it by a few percent in the Hankel transform we find that differences are imperceptible, even in the 36-layer case. If

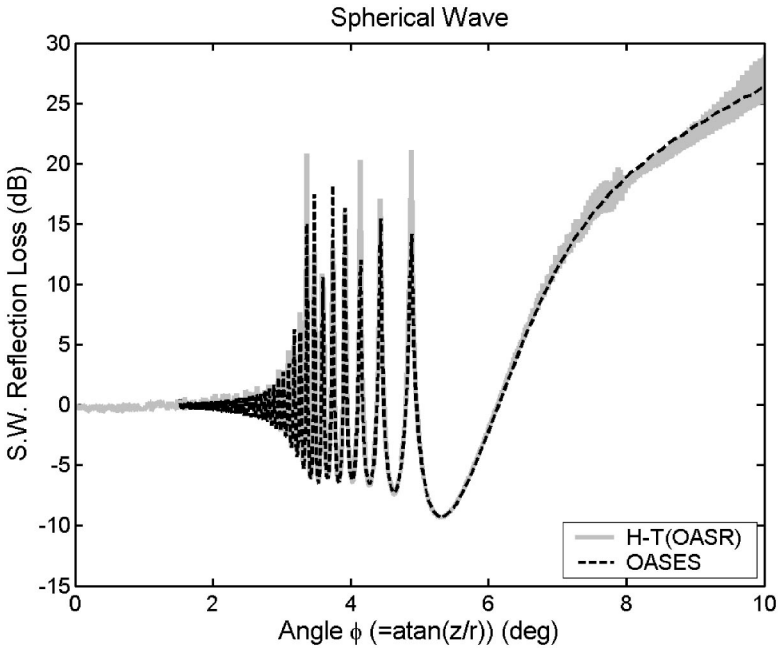


FIG. 4. Case 2: uniform gradient. Spherical wave reflection loss extracted from OASES by time domain separation of bottom paths (black dashed) and Hankel transformed OASR (gray solid).

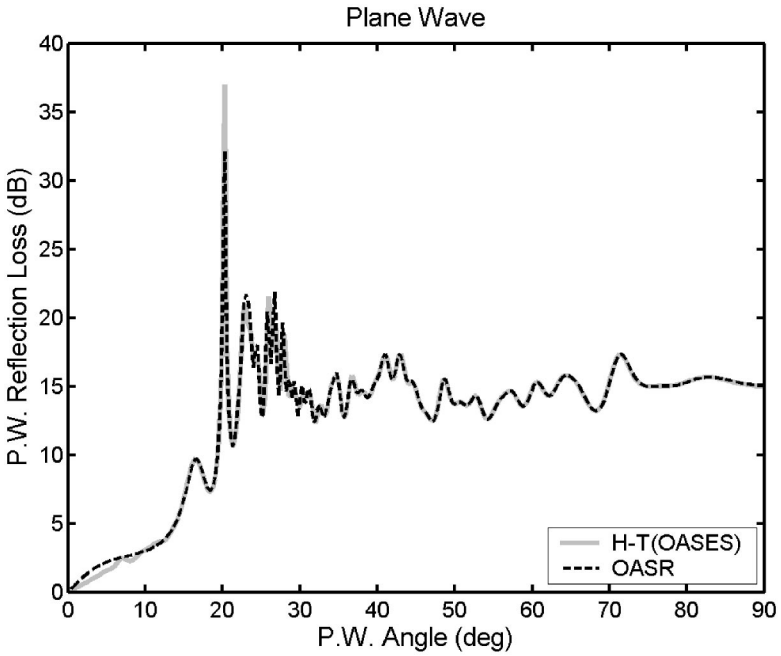


FIG. 5. Case 3: 36-layer. Plane wave reflection loss from OASR (black dashed) and Hankel transformed bottom return extracted from OASES by time domain separation of bottom paths (gray solid).

we introduce a gradient into the water column, the theory in Sec. 4.2.3 of Jensen *et al.* (1994) shows that this can be accommodated in the Hankel transform by altering the z dependence in Eqs. (3) and (4), however a demonstration of this is out of the scope of this paper.

VI. CONCLUSIONS

Equations (3) and (4) state that the plane wave and spherical wave reflection coefficients are related by a Hankel transform. Thus given point measurements with a phase, i.e., the spherical wave reflection coefficient (SWRC), one can reconstruct the plane wave reflection coefficient (PWRC). Demonstrations of this conversion were given for three cases, all involving multiple layers, and one including a pronounced velocity gradient generating a caustic. The 36-layer case is a powerful demonstration that the plane wave reflection coefficient retains all the acoustic information associated with the 36 layers. As a consequence one can manipulate this information and reconstruct acoustic fields from it without ever explicitly converting it into geoacoustic parameters. One therefore avoids the problems of geoacoustic ambiguity.

The examples also show that the plane wave and spherical wave reflection coefficients may be significantly different at frequencies of 500 Hz even with vertical separation of receiver and source image of 210 m (80 m+130 m). Therefore measurements with source and receiver at any depth in shallow water and frequencies of this order or lower are likely to require transformation. Differences can be regarded as caused by the interaction of the Fresnel term [Eqs. (13) and (14)] with the complex reflection coefficient, and the

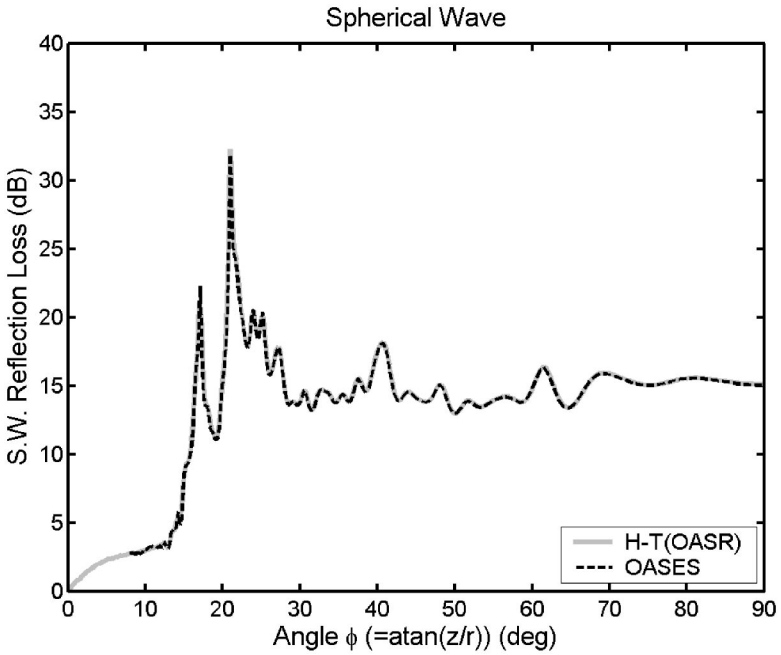


FIG. 6. Case 3: 36-layer. Spherical wave reflection loss extracted from OASES by time domain separation of bottom paths (black dashed) and Hankel transformed OASR (gray solid).

most serious cases involve sudden amplitude changes or phase curvature, which can lead to significant shifts of nulls, ringing, and so on.

[1]There may be more stationary phase points if there is a lateral wave or additional layers.

Ainslie, M. A., Robins, A. J., and Simons, D. G. (**2004**). ''Caustic envelopes and cusp coordinates due to the reflection of a spherical wave from a layered sediment,'' J. Acoust. Soc. Am. **115**, 1449–1459.

Brekhovskikh, L. M. (**1980**). *Waves in Layered Media* (Academic, New York).

Holland, C. W., and Osler, J. (**2000**). ''High resolution geoacoustic inversion in shallow water: a joint time- and frequency-domain technique,'' J. Acoust. Soc. Am. **107**, 1263–1279.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**1994**). *Computational Ocean Acoustics* (AIP, New York).

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw–Hill, New York).

Schmidt, H. (**1999**). ''OASES User's Guide and Reference Manual,'' Dept. of Ocean Engineering, MIT.

Schmidt, H., and Jensen, F. B. (**1986**). ''Evaluation of experimental techniques for determining the plane wave reflection coefficient at the sea floor,'' in *Ocean Seismoacoustics*, edited by T. Akal and J. M. Berkson (Plenum, New York), pp. 721–730.

Stickler, D. C. (**1977**). ''Negative bottom loss, critical-angle shift, and the interpretation of the bottom reflection coefficient,'' J. Acoust. Soc. Am. **61**, 707–710.

Tolstoy, A., Chapman, N. R., and Brooke, G. (**1998**). ''Workshop'97: Benchmarking for geoacoustic inversion in shallow water,'' J. Comput. Acoust. **6**, 1–28.

Westwood, E. K. (**1989**). ''Complex ray methods for acoustic interaction at a fluid-fluid interface,'' J. Acoust. Soc. Am. **85**, 1872–1884.

A fast Fourier transform on multipoles (FFTM) algorithm for solving Helmholtz equation in acoustics analysis

Eng Teo Ong[a)]
Institute of High Performance Computing, 1 Science Park Road #01-01 The Capricorn, Science Park II, Singapore 117528

Heow Pueh Lee[b)]
Institute of High Performance Computing, 1 Science Park Road #01-01 The Capricorn, Science Park II, Singapore 117528 and National University of Singapore, Department of Mechanical Engineering, 10 Kent Ridge Crescent, Singapore 119260

Kian Meng Lim[c)]
National University of Singapore, Department of Mechanical Engineering, 10 Kent Ridge Crescent, Singapore 119260

(Received 1 October 2003; revised 30 May 2004; accepted 2 June 2004)

This article presents a fast algorithm for the efficient solution of the Helmholtz equation. The method is based on the translation theory of the multipole expansions. Here, the speedup comes from the convolution nature of the translation operators, which can be evaluated rapidly using fast Fourier transform algorithms. Also, the computations of the translation operators are accelerated by using the recursive formulas developed recently by Gumerov and Duraiswami [SIAM J. Sci. Comput. **25**, 1344–1381 (2003)]. It is demonstrated that the algorithm can produce good accuracy with a relatively low order of expansion. Efficiency analyses of the algorithm reveal that it has computational complexities of $O(N^a)$, where a ranges from 1.05 to 1.24. However, this method requires substantially more memory to store the translation operators as compared to the fast multipole method. Hence, despite its simplicity in implementation, this memory requirement issue may limit the application of this algorithm to solving very large-scale problems. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1775273]

PACS numbers: 43.20.Fn [LLT] Pages: 1362–1371

I. INTRODUCTION

Boundary element method (BEM) is an effective numerical method for solving acoustics problems governed by the Helmholtz equation. It is particularly suitable for solving infinite domain problems, since it does not require artificial truncation and specifications of the boundary conditions, which are essential in domain solution methods such as the finite difference and finite element methods. However, the conventional BEM generates dense linear system, which requires $O(N^3)$ and $O(N^2)$ operations when solved using direct methods, such as the Gaussian elimination, and Krylov subspace iterative methods,[2] such as the generalized minimal residual (GMRES), respectively. The computation obviously becomes prohibitively large if the problem size N exceeds several thousands.

To improve the computation, numerous fast algorithms have been developed, and the fast multipole method (FMM) is definitely one of the most widely implemented algorithm. FMM was developed by Greengard and Rokhlin[3] for solving potential fields in N-body simulations. It was then shortly employed by Nabors *et al.*[4] to solve integral equation using the method of moments, where the FMM algorithm was used to accelerate the matrix-vector multiplication process required in the iterative solvers.

Although it is possible to apply the original FMM[3] algorithm to wave phenomenon problems governed by the Helmholtz or Maxwell equations, such implementation is not efficient as the method scales like $O(p^5)$ [see Eqs. (11) and (12)], where p is the truncation order of the multipole expansions. Unlike in the Laplace case, the use of the hierarchical procedure is considerably more complicated for the Helmholtz problem. This is mainly due to the oscillatory nature of the solution, which demands more stringent criteria for the convergence of the multipole approximations, namely, the wave-number-radius criterion (to be elaborated in Sec. IV). The important consequence is that the order of expansion p required by the FMM algorithms can be quite large. The actual FMM implementations for the wave problems did not surface until the work by Rokhlin,[5] who derived the diagonal form of the translation operators, which reduced the operation cost to $O(p^3)$ [see Eqs. (13) and (14)]. It laid down the firm foundation for further investigations by numerous other authors, such as Coifman *et al.*,[6] Dembart and Yip,[7] Gyure and Stalzer,[8] Darve,[9] and Song and Chew[10] and Song *et al.*,[11,12] and Koc and Chew[13] who contributed to the incorporation of the hierarchical structure into FMM, which is usually known as the multilevel fast multipole algorithm.

Another group of fast methods utilizes the fast Fourier transform (FFT) to accelerate the matrix-vector-

[a)]Electronic mail: onget@ihpc.a-star.edu.sg
[b)]Electronic mail: hplee@ihpc.a-star.edu.sg
[c)]Electronic mail: limkm@nus.edu.sg

 © 2004 Acoustical Society of America

multiplication process, such as the precorrected FFT method,[14] the adaptive integral method (AIM),[15] and the FMM-FFT method.[13] These methods exploit the convolution nature of the integral equation, in which the evaluation of the integral can be accelerated using FFT algorithms. Generally, the various methods differ in the ways in which the problems are mapped onto regularly spaced grids. For example in the precorrected FFT method,[14] the sources at the grids are obtained by matching potentials of the actual and grid sources at selected points, whereas in AIM,[15] they are derived by matching the truncated multipole moments. For the FMM-FFT[13] method, the grid point sources correspond to the multipole moments components that are associated with the given far-field direction.

Other fast solvers for the Helmholtz and Maxwell equations include the Matrix Decomposition methods,[16] and matrix sparsification techniques using the singular value decomposition[17] and wavelet transform.[18]

In this paper, we present an alternative fast algorithm, which we refer to it as the fast Fourier transform on multipoles (FFTM) method. This method arises from an important observation that potential evaluation using a multipole to local translation operator can be expressed as a series of discrete convolutions of the multipole moments with their associated spherical harmonics functions, and FFT algorithms can be employed to evaluate the discrete convolutions rapidly. To a certain extent, this algorithm can be seen as a variant of the FFT approach, where the multipole moments are the grid point sources in this case. On the other hand, it also resembles the FMM in the usage of the multipole and local expansions and their translation operators. It is remarked that this approach is similar to the FMM-FFT method,[13] which also utilizes of the discrete convolutions nature of the multipole expansions to speed up the matrix-vector calculation. The primary difference is that FMM-FFT uses the diagonal integral form of the translation operators [see Eqs. (13) and (14)], whereas in FFTM we use the discrete summation form [see Eqs. (11) and (12)], in which the computations of the translation operators are accelerated by using the fast and exact recursive formulas developed recently by Gumerov and Duraiswami.[1] In this case, the numerical errors and the instability problems,[7,9,19] that are associated with the evaluations of the integral translation operators are circumvented in the FFTM approach. FFTM was implemented for the evaluation of Coulombic potential in charge particle simulations.[20] As part of our continuous effort to extend the FFTM algorithm into new areas of application, we have implemented it for solving the Helmholtz equation commonly found in acoustics analyses.

This paper is organized as follows. In Sec. II, we briefly review the boundary integral formulation for the Helmholtz equation. This is followed by the mathematical formulas of the various translation operators required in the FFTM implementation. In Sec. IV, the FFTM algorithm is first briefly described, and some aspects of its accuracy and efficiency are then discussed. In Sec. V, we present two numerical examples to demonstrate the accuracy and efficiency of the method. Finally, the conclusions of this study are summarized in Sec. VI.

II. BOUNDARY ELEMENT METHOD FOR HELMHOLTZ EQUATION

The propagation of acoustic waves, which are harmonic in time through a fluid medium Ω, is governed by the Helmholtz equation[21]

$$\nabla^2\psi + k^2\psi = 0, \tag{1}$$

where ψ is the complex-valued velocity potential, and k is the wave number defined by $k=\omega/c$, with ω being the angular frequency and c the speed of sound in the given medium. The velocity potential is related to the acoustic pressure by $P=i\rho_f\omega\psi$, where ρ_f is the density of the fluid medium, and $i=\sqrt{-1}$. In the case of acoustic radiation in exterior problems, ψ must also satisfy the Sommerfeld radiation condition at infinity, that is,

$$\lim_{r\to\infty} r\left(\frac{\partial\psi}{\partial r} - ik\psi\right) = 0. \tag{2}$$

The boundary integral formulation of the Helmholtz equation can be derived by using the weighted residual technique together with the Green's identities, which gives[21]

$$\alpha(x)\psi(x) = \int_\Gamma \left[\frac{\partial\psi(y)}{\partial n} G(x,y) - \psi(y)\frac{\partial G(x,y)}{\partial n}\right] d\Gamma(y), \tag{3}$$

where x and y denote the field and source points, respectively, n is the unit normal at the boundary $\Gamma(y)$ that is directed away from the domain Ω, and $\alpha(x)$ is known as the jump term, which arises when x is moved to the boundary, and is dependent on the geometry of the boundary at x. The free-space Green's function $G(x,y)$, commonly known as the fundamental solution in BEM, and for the three-dimensional case is given by

$$G(x,y) = \frac{e^{ik\|x-y\|}}{4\pi\|x-y\|}. \tag{4}$$

The standard BEM implementation for solving Eq. (3) comprises the following three steps, namely: (i) discretizing the boundary into elements; (ii) forming a linear system of equations; and finally (iii) solving the dense linear system of equations. As mentioned before, solving the dense linear system via direct or iterative methods becomes prohibitively costly in terms of computational time and storage when N exceeds several thousands. Fortunately, due to the matrix-free nature of the iterative solvers, the matrix-vector multiplication task for generating the basis vectors can be significantly reduced with appropriate approximations. In this study, we present a fast algorithm that can speed up the matrix-vector calculation process.

III. TRANSLATION OPERATORS FOR HELMHOLTZ EQUATION

This section presents the translation operators used in the FFTM algorithm. Two different forms of translation operators exist, namely, the discrete summation and the diagonal integral forms. They were discussed in great details by Epton and Dembart.[22] FFTM works with the discrete summation form, and hence we present this form of the transla-

tion operators in details. The integral form of the translation operators will also be briefly discussed to illustrate the implementation differences between the two approaches. For simplicity, the formulas were written in the context of point sources.

A. Multipole expansion

Suppose there are n_q point sources of strength q_i that are located at positions $x_i=(\rho_i,\alpha_i,\beta_i)$, for $i=1,...,n_q$, and are bounded within a sphere S_a of radius a centered at the origin, that is $|\rho_i|<a$. Then for any point $y=(r,\theta,\phi)\in R^3$ with $r>a$, the potential generated by these sources is given by

$$\psi(y)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} M_n^m h_n(kr)Y_n^m(\theta,\phi), \tag{5}$$

where the multipole moment M_n^m is defined as

$$M_n^m=\sum_{i=1}^{n_q} q_i j_n(k\rho_i)Y_n^{-m}(\alpha_i,\beta_i). \tag{6}$$

Here, $h_n(kr)$ and $j_n(kr)$ are the spherical Hankel and Bessel functions of the first kind, which satisfy the Sommerfeld condition in Eq. (2). $Y_n^m(\theta,\phi)$ is the spherical harmonics of degree n and order m, and is defined as

$$Y_n^m(\theta,\phi)=\sqrt{\frac{(n-|m|)!}{(n+|m|)!}}P_n^{|m|}(\cos\theta)e^{im\phi}, \tag{7}$$

where $P_n^{|m|}(\cos\theta)$ is the associated Legendre function of the first kind.

In the case of distributed surface sources, as in boundary elements, the multipole moment has the following form:

$$M_n^m=\sum_{i=1}^{n_q}\int_{\Gamma_e} f(\rho_i,\alpha_i,\beta_i)j_n(k\rho_i)Y_n^{-m}(\alpha_i,\beta_i)d\Gamma, \tag{8}$$

where $f(\rho,\alpha,\beta)$ defines the variation of surface sources in three-dimensional space, and n_q here denotes the number of elements that fall within the bounding sphere S_a.

Equations (6) and (8) are the translation operators that convert a system of points or surface sources that is arbitrarily distributed within a sphere centered at the origin into multipole moments M_n^m defined at the origin. They are denoted by $Q2M$ hereafter.

B. Local expansion

Suppose there are n_q point sources of strength q_i that are located at positions $x_i=(\rho_i,\alpha_i,\beta_i)$, for $i=1,...,n_q$, and fall outside a sphere S_a of radius a centered at the origin, that is $|\rho_i|>a$. Then for any point $y=(r,\theta,\phi)\in S_a$, the potential generated by these sources is given by

$$\psi(y)=\sum_{s=0}^{\infty}\sum_{t=-s}^{s} L_s^t j_s(kr)Y_s^t(\theta,\phi), \tag{9}$$

where L_s^t is the local expansion coefficient, which is defined as

$$L_s^t=\sum_{i=1}^{n_q} q_i h_s(k\rho_i)Y_s^{-t}(\alpha_i,\beta_i). \tag{10}$$

The physical interpretation of local expansion coefficients is that they correspond to the potential and its gradients evaluated at the origin, which are generated by the n_q sources outside S_a. Equation (9) is the translation operator that converts local expansion coefficients defined at the origin into potential at some other arbitrary point $y=(r,\theta,\phi)$ within S_a. It is denoted by $L2P$ hereafter.

C. Multipole to local translation operators

Suppose there are n_q point sources of strength q_i that fall within a sphere S_X of radius a centered at the position $X=(\rho,\alpha,\beta)$, and that $\rho>(c+1)a$ with $c>1$. Then the corresponding multipole expansion converges inside a sphere S_O of radius a centered at the origin. For any point $y=(r,\theta,\phi)\in S_O$, the potential generated by these sources, which are converted to multipole moments M_n^m defined at X, is given by the local expansions in Eq. (9), and the local expansion coefficients are obtained by

$$L_s^t=\sum_{n=0}^{\infty}\sum_{m=-n}^{n} T_{s,n}^{t,m}M_n^m, \tag{11}$$

where $T_{s,n}^{t,m}$ is referred to as the translation operator for the associated multipole and local coefficients. Equation (11) is the discrete summation form of the multipole to local expansions translation operator, with the translation operator given by the formula

$$T_{s,n}^{t,m}=\sum_{a=0}^{\infty}\sum_{b=-a}^{a} \upsilon_{ans}^{bmt}h_a(kr)Y_a^b(\theta,\phi), \tag{12}$$

where the coefficient υ_{ans}^{bmt} is related to the Wigner 3-j symbols.[1,7,22]

It is noted from Eq. (12) that the computation of the translation operator require summations over additional indexes (a,b), due to the computation of the Wigner 3-j symbols. Although many of the υ_{ans}^{bmt} coefficients are zeros, evaluating the local expansion coefficients via Eqs. (11) and (12) still requires $O(p^5)$ operations, where p is the truncated order of expansion. Hence, most implementations of FMM do not use the above discrete summation form. Instead, they are based on the diagonal integral form derived by Rokhlin,[5] which can be briefly described as follows.

Consider two clusters of points C_a and C_b, which are enveloped within two non-intersecting spheres of radii R_a and R_b, and centered at O_a and O_b, respectively. Then for any pair of points (x_i,y_i), where $x_i\in C_a$ and $y_i\in C_b$, their corresponding Green's function as given in Eq. (4) can be approximated as

$$\frac{e^{ik\|x_i-y_i\|}}{\|x_i-y_i\|}\approx\int_S [e^{ik\langle\hat{s},x_i-O_a\rangle}T_{p,O_a-O_b}(\hat{s})]e^{ik\langle\hat{s},O_a-y_i\rangle}d\hat{s}, \tag{13}$$

where $\hat{S}$ is a sphere of unit radius, $\langle\ \rangle$ denotes the scalar product, and

$$T_{p,O_a-O_b}(\hat{s})=\sum_{m=0}^{p}\frac{(2m+1)i^m}{4\pi}h_m^{(1)}(k\|O_a-O_b\|)\times P_m(\cos(\hat{s},O_a-O_b)). \tag{14}$$

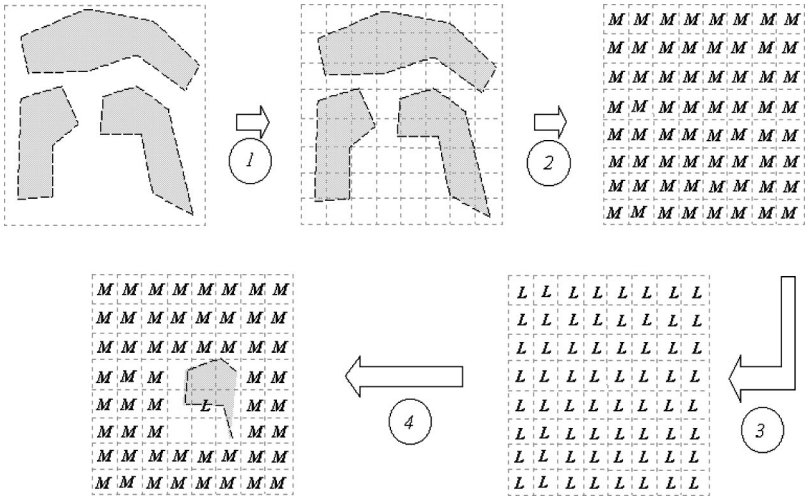


FIG. 1. 2D pictorial representation of FFTM algorithm. Step (1) Division of problem domain into many smaller cells and allocate panels; Step (2) Computation of multipole moments M for all cells. Step (3) Evaluation of local expansion coefficients L at cell centers by discrete convolutions via FFT. Step (4) For a given cell, compute the potentials at nodal locations using L, which accounts for effects of the ''distant'' charges, and also adding the ''near'' sources contributions (panels within the neighboring cells of the cell concerned) directly onto nodal locations.

An important feature in Eq. (14) is that the translation function is diagonal, that is, the summation is taken over the index m only. In general, it would require $O(p^2)$ Gauss quadrature points to integrate Eq. (13) accurately. Hence, the total operation cost in this case is only $O(p^3)$, which is significantly more efficient than the discrete summation form as given in Eqs. (11) and (12). It is mentioned here that for a desired order of accuracy, the p value depends greatly on the value of kD, where D is the larger of R_a and R_b. An empirical formula suggested by Coifman *et al.*[6] is

$$p = kD + \beta \ln(kD + \pi), \tag{15}$$

where β is 5 and 10 for single and double precision, respectively.

Note that both forms of the multipole to local expansions translation operators, i.e., Eq. (11) and the term within the square bracket of Eq. (13), are in a convolution form. The latter is essentially the fundamental principle upon which the FMM-FFT method[13] is based on, whereas FFTM works with discrete summation form in Eq. (11). But instead of dealing with the complicated Wigner 3-j symbols, the translation operators $T_{s,n}^{t,m}$ are computed using the recursive formulas developed by Gumerov and Duraiswami.[1]

D. Recurrence relations for translation operators

Recently, Gumerov and Duraiswami[1] derived recurrence relations for evaluating $T_{s,n}^{t,m}$ in Eq. (11), which only require $O(p^4)$ operations. These recursive relations are essentially derived by applying the differentiation theorems for spherical functions, which is similar to the work by Chew.[23] Although this approach is still one order less efficient than the integral form, it is easy to implement, and numerically stable. Furthermore, we would like to remark that the impact of the order of expansion p on the efficiency of FFTM is usually much less significant than that for the FMM, since low order p is usually used. This point will be clarified in Sec. IV. In the remainder of this section, the essential recurrence formulas will be summarized.

The recurrence formulas work in the following manner. First, the ''initial'' functions, which initiate the recurrence process, are computed as

$$\begin{aligned} T_{0,n}^{0,m} &= \sqrt{4\pi} h_n(kr) Y_n^m(\theta,\phi), \\ T_{n,0}^{m,0} &= \sqrt{4\pi}(-1)^n h_n(kr) Y_n^{-m}(\theta,\phi), \end{aligned} \tag{16}$$

for $n = 0,...,2p$, and $m = -n,\ldots,n$.

Then the ''sectorial'' functions, which are of the forms $T_{s,|m|}^{t,m}$ and $T_{|s|,n}^{s,m}$, are computed using the following relations:

$$\begin{aligned} b_{m+1}^{-(m+1)} T_{s,m+1}^{t,m+1} &= b_s^{-t} T_{s-1,m}^{t-1,m} - b_{s+1}^{t-1} T_{s+1,m}^{t-1,m} \\ b_{m+1}^{-(m+1)} T_{s,m+1}^{t,-(m+1)} &= b_s^{t} T_{s-1,m}^{t+1,-m} - b_{s+1}^{-(t+1)} T_{s+1,m}^{t+1,-m} \end{aligned} \tag{17}$$

for $m = 0,...,p-1$, $s = 1,...,2p-m$, and $t = -s,\ldots,s$, and

$$b_n^m = \begin{cases} \sqrt{\dfrac{(n-m-1)(n-m)}{(2n-1)(2n+1)}}, & 0 \leqslant m \leqslant n \\ -\sqrt{\dfrac{(n-m-1)(n-m)}{(2n-1)(2n+1)}}, & -n \leqslant m < 0 \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Finally, the remaining translation operators can be computed with the following recurrent relation:

$$a_n^m T_{s,n+1}^{t,m} = a_{n-1}^m T_{s,n-1}^{t,m} - a_s^t T_{s+1,n}^{t,m} + a_{s-1}^t T_{s-1,n}^{t,m} \tag{19}$$

for $n = 1,...,p$, $m = -n+1,...,n-1$, $s = 1,...,p-1$, and $t = -s,\ldots,s$, and

$$a_n^m = \begin{cases} \sqrt{\dfrac{(n+1+|m|)(n+1-|m|)}{(2n+1)(2n+3)}}, & n \geqslant |m| \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

IV. FAST FOURIER TRANSFORM ON MULTIPOLE

A. Fast Fourier transform on multipole algorithm

We would like to emphasis that the FFTM algorithm presented here is directly applicable to both the Laplace and Helmholtz problems, where the major distinction is due to the differences in the translation operators used. Hence, in this paper, we will only briefly describe the core of the algorithm. Readers are referred to our previous work[20] for detailed elaboration on the algorithm. Basically, this algorithm comprises the following four steps, as depicted in Fig. 1.

1. Spatial discretization

This step divides the problem domain into many smaller cells and allocates the elements among them. The primary objective here is to identify closely packed elements that are to be approximated by multipole moments defined at the cell centers. It also helps to differentiate the ''near'' and ''distant'' elements easily. In general, the nearness definition is given as follows. First, two cells are considered neighbors if they share at least one common vertex, and this gives a maximum of 27 nearest neighbors for any given cell, including itself. This is referred to as the first layer stencil of ''near'' neighbors. The second layer stencil of ''near'' neighbors of a given cell includes all its nearest neighbors' nearest neighbors, which gives a total of 125 cells. Following this definition, the D layer stencil of ''near'' neighbors would have at most $(2D+1)^3$ neighboring cells.

2. Converting cluster of panel sources into multipole moments

The operator that performs this task is given in Eq. (8), which is denoted by $Q2M$ translation operator. It converts the cluster of distributed surface sources within the cells to an equivalent set of point sources defined at the center of the cells. Applying $Q2M$ to all the cells transforms the original system of panel discretized problem to one that contains only regularly spaced point sources, as depicted in Fig. 1 after step (2).

3. Evaluating local expansion coefficients due to multipole moments

This step evaluates the local expansion coefficients at the cell centers, due to the multipole moments in all the cells. The regular spacing of the cell centers enable the calculations to be done rapidly through discrete convolutions using FFT algorithms. Mathematically, the multipole to local expansion translation formula, as given in Eq. (11), can be written as a series of three-dimensional discrete convolutions as follows:

$$L_s^t(x,y,z) \approx \sum_{n=0}^{p} \sum_{m=-n}^{n} \left[\sum_{x'} \sum_{y'} \sum_{z'} M_n^m(x',y',z') \times T_{s,n}^{t,m}(x-x',y-y',z-z') \right], \tag{21}$$

where $T_{s,n}^{t,m}$ denotes the translation operator of the multipole moment M_n^m for the local expansion coefficients L_s^t, and the indices (x,y,z), and (x',y',z') denote the discrete locations of the field points and multipole moments, respectively. The discrete convolution in the square bracket, as according to the discrete convolution theorem, can be evaluated rapidly using FFT algorithms.

Now, suppose the total number of cells used in step (1) is $N_x \times N_y \times N_z$. Then in order to account for all the multipole moments effects completely, the translation operators $T_{s,n}^{t,m}$ have to be evaluated for the range $(-N_i, N_i)$, for $i = x$, y, and z, and defined in wrap-around order. As for M_n^m, its size is also increased accordingly by zero-padding. In other words, the actual size of the discrete convolution is $8N_x \times N_y \times N_z$. However, the memory storage for $T_{s,n}^{t,m}$ can be reduced to about $2N_x \times N_y \times N_z$ by exploiting its symmetry. Also to exclude the effects of the ''near'' neighbors in the convolution process, the entries of $T_{s,n}^{t,m}$ for $x-x' \leqslant D$, $y-y' \leqslant D$, and $z-z' \leqslant D$ that correspond to the D-stencil neighboring cells are set to zero.

4. Evaluating of potentials at collocation points

The final step is to evaluate the potential at the collocation points. By using the $L2P$ translation operator in Eq. (9), one can evaluate the potential due to the ''distant'' elements at the desire locations. Finally, the total potential is obtained by adding the ''near'' element contributions directly onto the collocation points, which is denoted by $Q2P$.

B. Accuracy of fast Fourier transform on multipole

It was demonstrated for the calculation of Coulombic potential[20] in N-body simulation that FFTM can be significantly more accurate than FMM for a given order of expansion p. This is primarily due to the fact that the accuracy of the multipole approximation is inversely related to the relative distance c between the interacting cells. By foregoing the hierarchical procedure used in FMM, the parameter c in FFTM can be significantly larger than FMM, and hence providing FFTM the superior accuracy over FMM for a given expansion order p.

Similar error bound for the multipole approximation was also suggested for the Helmholtz equation[5] as

$$|\Phi_{\text{exact}} - \Phi_{\text{multipole}}| = O(1/c)^p \quad \text{for } p \geqslant ka, \tag{22}$$

where ka is a measure of the radius of the enclosing sphere in terms of the wavelength. It is important to note that $\Phi_{\text{multipole}}$ does not begin to converge until p is greater than ka. We refer to this condition as the wave-number-radius criterion. Basically, this criterion demands that the order of expansion must be large enough to capture the oscillating field within the sphere. The significance of this criterion will be discussed in the next subsection. Here, we would like to point out that the error bound in Eq. (22) is only partially true, and this can be illustrated using the following simple example.

Consider two interacting cells of unit length $(a = \sqrt{3}/2)$ that are separated by L units as shown in Fig. 2. In cell A, we place a point source at (0.5,0.5,0.5), and then compute the errors of its absolute potential at (0.5,−0.5,0.5) location in cell B using the various FFTM schemes, which are denoted by (ka,p). The convergence behavior of the FFTM errors with respect to the cell interacting distant L are plotted in Fig. 3. In general, it is observed that the errors would decrease initially as the separation between cells A and B widens. But this decreasing trend ceases at some point, and the accuracy would stabilize at a certain value after some fluctuations. Now, this error convergence behavior can be explained as follows. Recall from Eq. (4) that the Green's function comprises of two parts, namely, $1/r$ and e^{ikr}, in which the former term decreases monotonically to zero as r increases, whereas the latter one is oscillatory in nature. The initial reduction in the errors as r increases is likely due to the dominant monotonic behavior of the first term. However

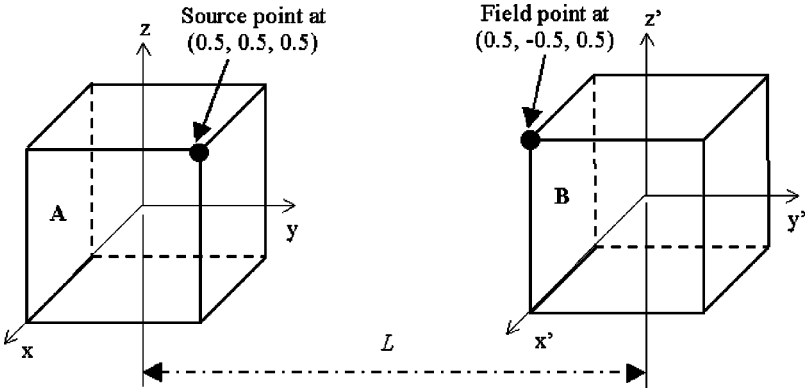


FIG. 2. Multipole expansions error analysis using the simple source and field points ensemble, with interacting cells separated by L distant.

beyond a certain distance, this effect becomes lesser and the variation of the field is dominated by the second term, which remains oscillatory for large values of r. Hence, in order to capture this oscillatory field accurately, we need to increase the expansion order p.

In this case, the above-mentioned argument of FFTM being more accurate than FMM, due to the difference in the relative distance c between the interacting cells, does not hold for the Helmholtz equation. Nevertheless, FMM still requires a significantly higher order of expansion than FFTM to attain a given order of accuracy. This is mainly because of the wave-number-radius criterion, which will be discussed in the following subsection.

C. Impact of wave-number-radius criterion on fast multipole method and fast Fourier transform multipole

FMM works in a hierarchical fashion by translating multipole and local expansions at different levels, regardless of whether it is the discrete summation or the diagonal integral forms. This recursive regrouping process results in an oct-tree structure that has cells of different sizes at different levels of the tree. Suppose that at the finest level L, the enclosing sphere radius is a. Then at any higher level 1, its size would be $(2^{L-1})a$. Now, according to the wave-number-radius criterion, the order of expansion p has to be increased accordingly, as the algorithm moves upward the oct-tree, so that the oscillating fields within a larger sphere can be captured correctly. However, when translating local expansions downward the oct-tree, the reverse is done to avoid unnecessary computations, since a lower order of expansion is needed to represent the field inside a smaller sphere. In terms of the far-field representations, this means that more/less Gauss quadrature points are required to evaluate the integral as the algorithm moves upward/downward the hierarchy. The addition and removal of the Gauss quadrature points are commonly known as the interpolation and anterpolation (filtering) schemes,[8,9,12] respectively.

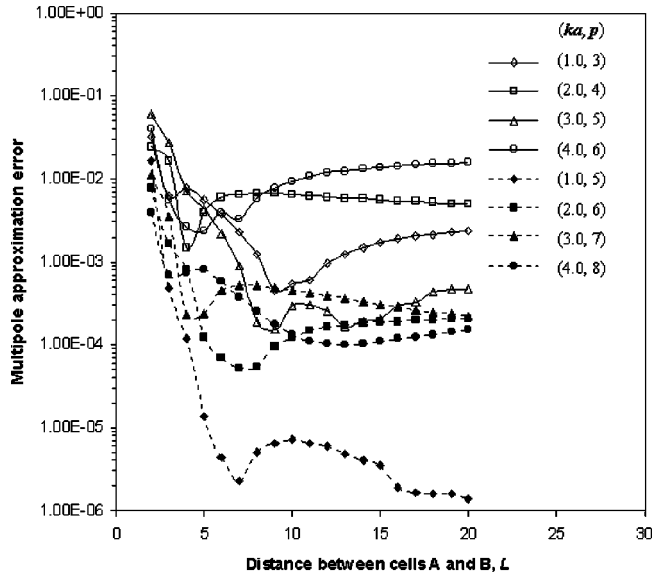


FIG. 3. Multipole expansion error convergence plots with respect to the relative distances between the interacting cells L.

The first consequence of this process is that it complicates the implementation of the FMM for Helmholtz equation as compared to the Laplace equation. More importantly, it affects the efficiency of the FMM quite significantly. For the Laplace equation, where the order of expansion remains the same at all levels, the computational cost decreases considerably as the algorithm moves up the hierarchy, since the number of cells at the higher level decreases rapidly. However, for the Helmholtz equation, the efficiency gained from the decrease in the number of cells is offset by the proportional increase in the order of expansion (more Gauss quadrature points for the far-field representation). In fact, the computational complexity of FMM for the Laplace equation is often claimed to be $O(N)$, whereas for the Helmholtz case it is $O(N \log N)$. Furthermore, using large values of p can lead to numerical instability,[7,9,19] sometimes called the subwavelength breakdown problem, which is associated with the rapid growth of the spherical harmonics and Hankel functions.

On the other hand, the wave-number radius criterion is expected to have less impact on FFTM, simply because it forgoes the hierarchical approach used in FMM. Here, we only need to ensure that this criterion is satisfied by the cells (equivalent to the cells at the finest level of FMM) used to discretize the problem. The FFTM implementations for the Laplace and Helmholtz equations are identical, except for the difference translation operators used. In other words, the FFTM implementation for the Helmholtz equation is rather straightforward as compared with FMM.

TABLE I. Pulsating sphere example for $k=4.0\,\mathrm{m}^{-1}$. Analytical and standard BEM solutions are $(-390.62+97.654i)$ and $(-389.65+95.069i)$, respectively.

Order of expansion p	FFTM schemes $ka=1.0$	$ka=1.5$	$ka=2.0$
2	$-390.92+95.479i$	$-388.17+96.463i$	—
3	$-389.92+95.188i$	$-389.86+94.869i$	$-388.67+96.029i$
4	$-389.57+95.114i$	$-389.61+95.093i$	$-389.50+95.165i$
5	$-389.61+95.085i$	$-389.64+95.071i$	$-389.61+95.065i$
6	$-389.66+95.069i$	$-389.66+95.065i$	$-389.66+95.063i$

V. NUMERICAL EXAMPLES

In this section, we examine the performance of the FFTM algorithm for the Helmholtz equation. We apply it on two examples, namely, (i) a unit pulsating sphere, and (ii) a scattering problem in underwater acoustics. First, it is demonstrated that FFTM is an accurate method for solving the Helmholtz equation. For the efficiency study, we have solved the scattering problem for five frequencies, ranging from 2 to 10 kHz at 2 kHz intervals, which results in problem sizes of 2842 to 71 002. Note that the element meshes for the examples increase with the frequencies being analyzed. This is necessary because more elements are required to capture the rapidly oscillating field at higher frequencies. In this study, we have used approximately 8–12 linear elements per wavelength λ.

A. Accuracy analysis of FFTM

1. Unit pulsating sphere

The first example, often regarded as the standard test problem, is the pulsating sphere of unit radius in an infinite acoustics domain. Its normal velocity is taken to be unity, which gives the normal surface pressure $\partial P/\partial n = i\rho\omega$. The analytical solution for its surface pressure P is then given by

$$P = \frac{\rho\omega}{1+k^2}(k+i), \tag{23}$$

where the variables are defined in Eq. (1). For air medium, $\rho=1.21\ \mathrm{kgm}^{-3}$ and $c=343\,\mathrm{ms}^{-1}$.

Two values of the wave number are considered here, namely, $k=4$ and 8 (m^{-1}). The analyses have been performed using three different schemes, where ka are given by 1.0, 1.5, and 2.0, and the order of expansion p varies from 2 to 6. The ''near'' neighboring cells are defined by the first layer stencil $(M=27)$.

The analytical solutions of P for the two cases $k=4$ and 8 using Eq. (23) are $(-390.62+97.654i)$ and $(-408.65+51.081i)$, respectively. For the numerical solutions, we have calculated the average value of the complex pressure over the sphere. Note that the comparisons between the FFTM solutions and the analytical solutions are not very useful. This is because the FFTM solutions can only converge to the numerical solutions obtained using the standard BEM. Here, we have also calculated the numerical solutions using the standard BEM, which are $(-389.65+95.069i)$ and $(-409.16+49.775i)$ for $k=4$ and 8, respectively. These results are different from the analytical solutions primarily due to the discretization error and approximations in the standard BEM. The FFTM solutions for the two cases are tabulated in Tables I and II, respectively.

It is observed that the FFTM solutions converge quite rapidly to the standard BEM solutions. From this example, the FFTM can produce accurate solutions (less than 0.1% error with respect to standard BEM solutions) if the order of expansion p is chosen to be at least $(ka+2)$. However, by this accuracy analysis alone, which is based on the average pressure field, it is not sufficient to draw a conclusion on the accuracy of the FFTM. In the following example, we study the accuracy of FFTM in terms of the primary solution, that is the pressure distributions.

2. Scattering problem in underwater acoustics

Underwater acoustics is commonly used for surveillance, location, and identification of submerged objects, analogous to radar above ground. In this example, we study the scattering patterns of a submerged cylinder due to a point incident sound wave of unit magnitude and frequency 2 kHz, which is located 10 m above it. The cylinder is 5 m long and 0.5 m in radius. To avoid the edge singularity problem, we have rounded the edges of the cylinder with a fillet of radius 0.1 m. The cylinder is assumed to be sound hard, that is $\partial P/\partial n=0$ on the cylindrical surface. Figure 4 show the schematic diagram of the problem.

TABLE II. Pulsating sphere example for $k=8.0\,\mathrm{m}^{-1}$. Analytical and standard BEM solutions are $(-408.65+51.081i)$ and $(-409.16+49.775i)$, respectively.

Order of expansion p	FFTM schemes $ka=1.0$	$ka=1.5$	$ka=2.0$
2	$-410.54+49.103i$	$-410.46+50.970i$	—
3	$-409.10+49.956i$	$-409.32+49.831i$	$-409.42+49.693i$
4	$-409.14+49.810i$	$-409.11+49.846i$	$-409.07+49.964i$
5	$-409.15+49.765i$	$-409.14+49.772i$	$-409.14+49.781i$
6	$-409.17+49.776i$	$-409.16+49.775i$	$-409.16+49.778i$

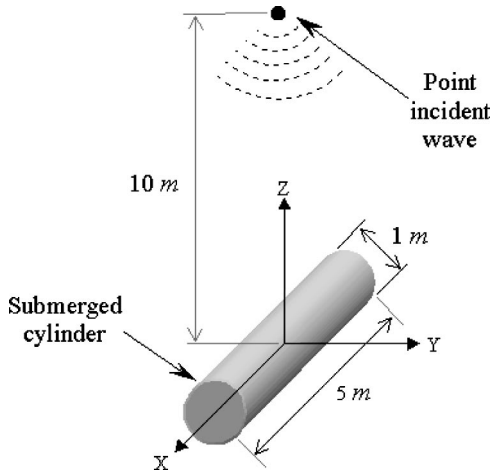


FIG. 4. Schematic diagram for the scattering cylinder example in underwater acoustics analysis.

In this example, we compute the L_2 norm errors of the pressure distribution solutions, defined as

$$\text{Error}=\left(\frac{\Sigma_{i=1}^{N}|P_i^{\text{BEM}}-P_i^{\text{FFTM}}|^2}{\Sigma_{i=1}^{N}|P_i^{\text{BEM}}|^2}\right)^{1/2}, \tag{24}$$

where P_i^{BEM} and P_i^{FFTM} denote ith nodal surface pressure of the standard BEM and the FFTM schemes, respectively. The analyses were done using the first ($M=27$) and second layer ($M=125$) stencils for the ''near'' cells, and the error plots for the various FFTM schemes are depicted in Fig. 5.

It is again demonstrated that the FFTM can produce accurate solution using a relatively low order of expansion. From the plots, it is observed that the FFTM schemes can achieve less than 1.0% error if $p\geqslant(ka+2)$, whereas for error less than 0.1%, it would require $p\geqslant(ka+3)$. In general, the second layer stencil schemes can give more accurate solutions, and have slightly faster convergence rates. However, we would like to point out that the accuracy of the two sets of FFTM schemes is not significantly different (generally less than one order for the same (ka,p) scheme). This observation can be explained as follows. Notice that the only difference between the two sets of schemes is due to the evaluations of the second layer of nearest cell contributions (at most $5^3-3^3=98$ cells). The one-layer stencil schemes approximated them using the multipole expansions, while the two-layer stencil schemes accounted for them directly. Generally, this second layer of cells only contribute to a small fraction of the total potential, which is particularly true for the large-scale problems, and hence they do not change the final results significantly. Unfortunately, we cannot quantify this ratio in general, because it is likely to be problem dependent.

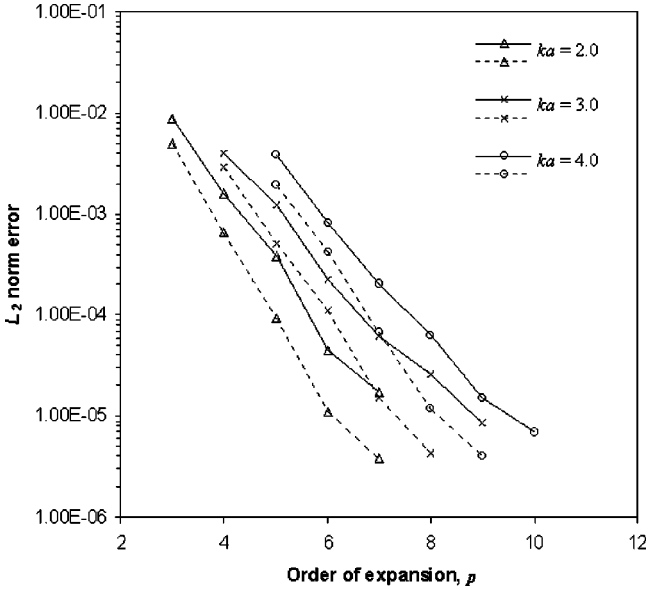


FIG. 5. L_2 norm error convergence plots for the surface pressure solutions using first layer (solid lines) and second layer (dashed lines) stencils for ''near'' cells definition.

B. Efficiency analysis of fast Fourier transform multipole

In this section, we study the efficiency of the FFTM algorithm to demonstrate that it has computational complexity, both in terms of memory storage and CPU time, that is significantly better than $O(N^2)$ in the standard BEM approach.

The underwater scattering problem is used to perform this efficiency analysis. The example is solved for five different incident wave frequencies, ranging from 2 to 10 kHz at intervals of 2 kHz, and the corresponding problem sizes are 2842, 11 362, 25 562, 45 442, and 71 002, respectively. In the light of the accuracy analysis above, we will only use the one-layer stencil here, since it should provide comparable accuracy as compared to the two-layer stencil, at lower computational costs. The analysis is done separately for the two major parts of the algorithm, namely, at the initialization and iteration stages. At the initialization stage, we determine the computational costs, both in terms of the CPU time and memory storage requirement, for forming the various trans-

TABLE III. Computational costs for solving the underwater acoustics examples for $<1.0\%$ accuracy.

Problem size (frequency)	FFTM scheme (ka,p)	Memory storage (MB)		CPU times (seconds)	
		Translation operators	Transformation matrices	Initialization stage	Iteration stage
2842 (2kHz)	1.6, 4	14.7	3.72	11.6	0.21
11362 (4kHz)	2.5, 5	60.5	25.3	60.8	0.85
25562 (6kHz)	2.7, 5	120.4	60.7	166.3	2.31
45442 (8kHz)	2.9, 5	210.3	118.3	358.6	4.46
71002(10kHz)	3.0, 5	431.5	192.0	610.9	8.71

TABLE IV. Computational costs for solving the underwater acoustics examples for <0.1% accuracy.

Problem size (frequency)	FFTM scheme (ka,p)	Memory storage (MB)		CPU times (seconds)	
		Translation operators	Transformation matrices	Initialization stage	Iteration stage
2842 (2kHz)	3.6, 8	31.5	13.2	34.1	0.26
11362 (4kHz)	3.8, 8	123.5	52.9	143.0	1.20
25562 (6kHz)	4.3, 9	235.2	152.0	430.8	3.01
45442 (8kHz)	4.5, 9	518.4	280.6	794.1	6.31
71002(10kHz)	4.5, 9	929.3	428.7	1286.8	12.8

formation matrices (mainly the $Q2P$ matrices) and the FFT of the translation operators. But at the iteration stage, we are only concerned with the CPU time taken for one matrix-vector multiplication, because the memory storage requirements here is considerably less than that required at the initialization stage.

Two different orders of accuracy are considered here, namely, <1.0% and <0.1% in terms of the L_2 norm errors as defined in Eq. (24). Based on the above accuracy analysis, the FFTM schemes with p given by the smallest integer larger than $(ka+2)$ and $(ka+4)$ are used for 1% and 0.1% accuracy, respectively. The memory storage requirements and the CPU times for these two orders of accuracy are tabulated in Tables III and IV, respectively. Note that the results presented in the tables correspond to the FFTM schemes that require minimum memory storage. These tables are organized as follows:

(1) Column 1: Problem size as defined by frequency.

(2) Column 2: FFTM scheme defined by (ka,p).

(3) Columns 3 and 4: Memory storage requirements (in MB) for, (i) translation operators $T_{s,n}^{t,m}$, and (ii) transformation matrices (mainly $Q2P$).

(4) Columns 5 and 6: CPU times (in seconds), (i) at initialization stage, and (ii) for computing one GMRES iteration.

The computational complexity plots for the total memory storage requirement (column 3+column 4) and the CPU times at the initialization stage (columns 5), and the CPU times for computing one GMRES iteration (column 6), are depicted in Fig. 6. From the plots, we observe that the curves can be approximated by power functions of the form AN^B, which corresponds to a straight line in a log-log plot. In Table V, we tabulated the approximated power functions for the various curves using the least-square fit. It is found that the complexity of FFTM is lower than quadratic complexity growth. However, the constants in the power functions are not small as desired, especially for the memory storage requirements. This is a common shortcoming of FFT based methods[13–15,20] as compared to the FMM, because one needs to store the translation operators for all the cells in the problem domain regardless of whether they are empty or not. Also unlike the case of Laplace equation,[20] where there are only $O(p^2)$ distinct translation operators, there are $O(p^4)$ translation operators for the Helmholtz equation. This excessive memory storage requirement by the FFTM has greatly limited its practicability in solving very large problems. To be more explicit, using the approximated power functions in Table V, FFTM can only solve problem of sizes of about 200 000 and 100 000 for 1.0% and 0.1% accuracy, respectively, with 2 GB of memory. But for the FMM, it was reported in Ref. 12 that the method can solve problems with 600 000 unknowns with 2 GB of memory (probably for 1.0% accuracy).

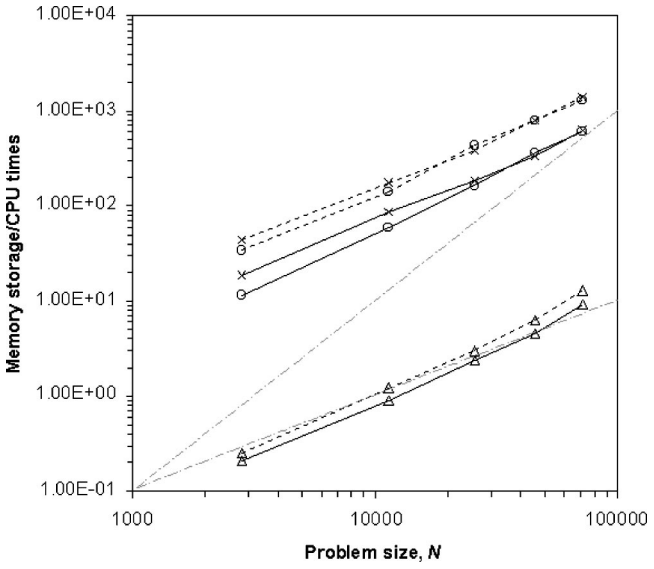


FIG. 6. Computational complexity plots of the memory storage requirements (x) and the CPU times at the initialization (○) and iteration (△) stages, for <1.0% (solid lines) and <0.1% (dashed lines) accuracy.

VI. CONCLUSIONS

In this paper, we presented an alternative fast algorithm called the fast Fourier transform on multipoles (FFTM) method for the efficient solution of the Helmholtz equation. This method employs the multipole and local expansions to approximate the far field potentials. The speed up in the algorithm primarily comes from the realization of the convolution nature of the multipole to local expansion translation operators, which can be evaluated rapidly using FFT algo-

TABLE V. Estimated power function least-squares fit for the computational complexity of FFTM.

Computational variable	Power function least-squares fit	
	1.0% accuracy	0.1% accuracy
Memory storage requirement	3.80E-3 $N^{1.07}$	9.76E-3 $N^{1.05}$
CPU time at initialization stage	6.04E-4 $N^{1.24}$	3.65E-3 $N^{1.14}$
CPU time for one matrix-vector product	1.74E-5 $N^{1.17}$	1.80E-5 $N^{1.19}$

rithms. We have also employed the recursive formulas developed by Gumerov and Duraiswami[1] when computing the translation operators, which avoid dealing with the complicated Wigner 3-j symbols, and more importantly reduce the computational cost from $O(p^5)$ to $O(p^4)$, where p is the order of expansion.

It is demonstrated that the algorithm can achieve reasonably good accuracy with relatively low order of expansion. But, unlike the case of Laplace equation, the multipole approximation does not improve infinitely with increasing separation between the interacting cells. Nevertheless, the order of expansion required by FFTM is often much lower than FMM due to the wave-number-radius criterion. The efficiency studies of the algorithm also show that it has computational complexity of $O(N^a)$, where a ranges from 1.05 to 1.07 for the memory storage requirement, and from 1.14 to 1.24 for the CPU times.

The main advantage of this algorithm is its ease of implementation as compared to the FMM. Basically, FFTM replaces the complicated hierarchical procedure (including the interpolation and anterpolation steps) of the FMM algorithm, with discrete convolutions that can be evaluated rapidly with FFT routines. However, the present algorithm requires relatively excessive memory storage for the $O(p^4)$ translation operators. Further investigation is currently underway to improve on this shortcoming.

[1] N. A. Gumerov and R. Duraiswami, ''Recursions for the computation of multipole translation and rotation coefficients for the 3D Helmholtz equation,'' SIAM J. Sci. Comput. (USA) **25**, 1344–1381 (2003).

[2] C. T. Kelly, *Iterative Methods for Linear and Nonlinear Equations* (SIAM, Frontier in Applied Mathematics, Philadelphia, 1995).

[3] L. Greengard and V. Rokhlin, ''A fast algorithm for particle simulations,'' J. Comput. Phys. **73**, 325–348 (1987).

[4] K. Nabors, F. T. Korsmeyer, F. T. Leighton, and J. White, ''Preconditioned, adaptive, multipole-accelerated iterative methods for three-dimensional first-kind integral equations of potential theory,'' SIAM J. Sci. Comput. (USA) **15**, 713–735 (1994).

[5] V. Rokhlin, ''Diagonal forms of translation operators for Helmholtz equation in three-dimensions,'' Appl. Comput. Harmon. Anal. **1**, 82–93 (1993).

[6] R. Coifman, V. Rokhlin, and S. Wandzura, ''The fast multipole method for the wave equation: A pedestrian prescription,'' IEEE Antennas Propag. Mag. **35**, 7–12 (1993).

[7] B. Dembart and E. Yip, ''The accuracy of fast multipole methods for Maxwell's equations,'' IEEE Comput. Sci. Eng. **5**, 48–56 (1998).

[8] M. F. Gyure and M. A. Stalzer, ''A prescription for the multilevel Helmholtz FMM,'' IEEE Comput. Sci. Eng. **5**, 39–47 (1998).

[9] E. Darve, ''The fast multipole method: numerical implementation,'' J. Comput. Phys. **160**, 195–240 (2000).

[10] J. M. Song and W. C. Chew, ''Multilevel fast multipole algorithm for solving combined field integral equations of electromagnetic scattering,'' Microwave Opt. Technol. Lett. **10**, 14–19 (1995).

[11] J. M. Song, C. C. Lu, and W. C. Chew, ''Multilevel fast multipole algorithm for electromagnetic scattering by large complex objects,'' IEEE Trans. Antennas Propag. **45**, 1488–1493 (1997).

[12] J. M. Song, C. C. Lu, W. C. Chew, and S. W. Lee, ''Fast Illinois Solver Code (FISC),'' IEEE Antennas Propag. Mag. **40**, 27–34 (1998).

[13] S. Koc and W. C. Chew, ''Calculation of acoustical scattering from a cluster of scatterers,'' J. Acoust. Soc. Am. **103**, 721–734 (1998).

[14] X. C. Nie, L. W. Li, and N. Yuan, ''Precorrected-FFT algorithm for solving combined field integral equations in electromagnetic scattering,'' J. Electromagn. Waves Appl. **16**, 1171–1187 (2002).

[15] E. Bleszynski, M. Bleszynski, and T. Jaroszwicz, ''AIM: Adaptive integral method for solving large-scale electromagnetic scattering and radiation problems,'' Radio Sci. **31**, 1225–1251 (1996).

[16] E. Michielssen and A. Boag, ''Multilevel evaluation of electromagnetic fields for the rapid solution of scattering problems,'' Microwave Opt. Technol. Lett. **7**, 790–795 (1994).

[17] S. Kapur and D. E. Long, ''IES^3: Efficient electrostatic and electromagnetic simulation,'' IEEE Comput. Sci. Eng. **5**, 60–67 (1998).

[18] R. L. Wagner and W. C. Chew, ''A study of wavelets for the solution of electromagnetic integral equations,'' IEEE Trans. Antennas Propag. **43**, 802–810 (1995).

[19] S. Koc, J. M. Song, and W. C. Chew, ''Error analysis for the numerical evaluation of the diagonal forms of the scalar spherical addition theorem,'' SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **36**, 906–920 (1999).

[20] E. T. Ong, K. M. Lim, K. H. Lee, and H. P. Lee, ''A fast algorithm for three-dimensional potential fields calculation: Fast Fourier Transform on Multipoles (FFTM),'' J. Comput. Phys. **192**, 244–261 (2003).

[21] L. C. Wrobel, *The Boundary Element Method Volume 1: Applications in Thermo-Fluids and Acoustics* (Wiley, Chichester, West Sussex, 2002).

[22] M. Epton and B. Dembart, ''Multipole translation theory for the three-dimensional Laplace and Helmholtz equations,'' SIAM J. Sci. Comput. (USA) **16**, 865–897 (1995).

[23] W. C. Chew, ''Recurrence relations for three-dimensional scalar and addition theorem,'' J. Electromagn. Waves Appl. **6**, 133–142 (1992).

An integral equation approach to three-dimensional acoustic radiation and scattering problems

S. A. Yang[a)]
Department of Systems and Naval Mechatronic Engineering, National Cheng Kung University, Tainan, Taiwan, Republic of China

(Received 28 December 2003; revised 15 June 2004; accepted 17 June 2004)

This paper provides numerical implementation and tests for a boundary integral equation method proposed recently by Krutitskii [IMA J. Appl. Math. **64**, 259–269 (2000)] for evaluating three-dimensional acoustic radiation and scattering fields in the exterior Neumann problem for several obstacles. Auxiliary boundaries with certain prescribed boundary conditions are introduced inside the obstacles for the purpose of eliminating the fictitious characteristic frequencies. Applying the simple source formulation on all boundaries leads to a set of uniquely solvable integral equations with weakly singular kernels. To facilitate the numerical implementation, the modified integral equations are further transformed to ordinary ones that contain nonsingular kernels only. Test cases consist of the acoustic radiation and scattering by one sphere and a pair of equal spheres. Comparisons are made with analytical solutions and Schenck's CHIEF method. Numerical results demonstrate the efficiency, robustness, and accuracy of the proposed method. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779751]

PACS numbers: 43.20.Fn, 43.20.Tb [JBS] Pages: 1372–1380

I. INTRODUCTION

This paper is concerned with the numerical solutions of the Neumann problem in three-dimensional acoustic radiation and scattering using a modified boundary integral equation (BIE) method. Boundary integral equation methods have been used to solve exterior wave propagation problems including acoustics, electromagnetism, and the surface-wave-body interactions for many years. The most important features of the methods are that they only require discretization of the surface rather than the volume and the outgoing radiation conditions at infinity are automatically satisfied. The methods are nowadays extensively used in other problems with infinite or semi-infinite domains such as those occurring in Stokes flow,[1] porous media flow,[2] aerodynamics,[3] elastostatics,[4] elastodynamics,[5] fracture mechanics,[6] geomechanics,[7] inverse analysis,[8] and many others. Effectively treating singular kernels and boundary elements are known to be two major problems when compared with domain-based methods such as the finite-element and finite-difference methods. However, the intensive investigation into these two subjects has helped to popularize the technique amongst practicing engineers.

At the characteristic frequencies of the associated interior Dirichlet problem, it is known that the simple source formulation for the exterior Neumann problem does not have a solution and that there is no unique solution of the surface Helmholtz integral formulation. This nonexistence or nonuniqueness nuisance does not have any physical significance, and is a purely mathematical problem arising from the breakdown of boundary integral representation at the characteristic frequencies. Various methods have been proposed to deal with this nonexistence or nonuniqueness difficulty in the last several decades.

Burton and Miller[9] developed a composite integral equation to circumvent the nonuniqueness difficulty at the characteristic frequencies. They linearly combined the Helmholtz integral equation with its normal derivative and proved the uniqueness of the solution. The major problem of this approach has been the evaluation of the hypersingular integrals involving a double normal derivative of the free-space Green's function. Burton and Miller[9] also presented a method of double surface integrals that reduced the order of the singularity. Such a formulation, however, is inefficient in computation due to the presence of double integrals. Most of the Burton–Miller-type techniques developed thereafter thus emphasized enhancing the numerical efficiency of evaluating hypersingular integrals. On the basis of the Hadamard[10] finite-part interpretation, Krishnasamy *et al.*[11] applied the Taylor series expansion and Stokes' integral theorem for the regularization. Chien, Rajiyah, and Atluri[12] used some identities arising in the integral equation related to an interior Laplace problem so as to reduce the order of the kernel singularity. Yang[13] refined the process and obtained modified integral equations containing bounded kernels only. Liu and Rizzo[14] presented a weakly singular form of the hypersingular integral equations by employing a two-term Taylor series subtraction from the density function. Certain integral identities for the static Green's function were used to evaluate the added-back terms. Yang[15] expressed the unknown functions as a truncated Fourier–Legendre series. Some weakly singular integrals and the hypersingular integral were analytically evaluated by using some properties of Legendre functions. Yan, Hung, and Zheng[16] developed a double surface integral method so as to transform the double surface integrals into a product of two discretized operator matrices. Harris and Chen[17] proposed a high order Galerkin method by using a

a)yha111@mail.ncku.edu.tw

 © 2004 Acoustical Society of America

singularity subtraction technique to reduce the hypersingular operator to a weakly singular one. The numerical procedures included two particular iterative solvers: the conjugate gradient normal method and the generalized minimal residual method.

Schenck[18] presented a combined Helmholtz integral equation formulation (CHIEF) that added some interior integral relations to the surface Helmholtz integral equation. The resulting overdetermined system of equations may then be solved by a least-squares technique. This approach perhaps is the most commonly used one in engineering applications. The selection of the optimum number and appropriate positions of the interior points, however, may become an issue as the wave frequency increases. One usually needs at least one ''good'' interior point, the so-called CHIEF point, to insure uniqueness of the solution at the characteristic frequencies. A ''good'' CHIEF point means that it must be away from the neighborhood of a nodal point of the corresponding interior eigenmode. Numerous variants of CHIEF have been put forth for effectively treating interior points. Seybert and Rengarajan[19] provided an ''interior point pressure checking'' procedure to detect the accuracy of the CHIEF solution. The concept of adding more constraints for each interior point was applied in Segalman and Lobitz's ''superCHIEF'' method,[20] Seybert and Wu's ''enhanced'' CHIEF method,[21] and Wu and Seybert's ''CHIEF-block'' method.[22] See also a recent work by Chen, Liu, and Hong[23] about using CHIEF for multiply connected problems. Although the above-mentioned techniques may make the proper selection of interior points less critical, they also increase the complexity of computation. Benthien and Schenck[24] suggested checking *a posteriori* the effectiveness of CHIEF points by evaluating the interior field. They indicated that CHIEF continues to be a practical approach for avoiding the characteristic frequency problem.

The off-boundary type of Burton–Miller formulation was developed for simultaneously avoiding the characteristic frequency and singular kernel problems.[25,26] This approach is based on the Helmholtz integral relation, but places all the collocation points interior to the boundary surface. It is usually more costly to implement than, say, CHIEF owing to its complicated integrands. The wave superposition method, another off-boundary approach, was developed by placing point sources interior to the boundary surface.[27,28] This method linearly combined monopoles and dipoles to circumvent the nonexistence difficulty arising from using the monopole distribution alone. The main deficiency of both types of the off-boundary methods lies in the selection of the inner surface, particularly when the boundary surface has discontinuities. Miller *et al.*[29] proposed some rules for selecting the inner surface; however, the optimum inner surface was determined by a process of trial and error. Krutitskii[30] developed a so-called method of interior boundaries by introducing an additional boundary inside the scattering/radiation obstacle. Satisfaction of the prescribed boundary condition on the interior surface leads to a set of weakly singular integral equations in a form of the single layer potential. Yang[31] numerically carried out Krutitskii's formulation and presented a detailed discussion.

The main goal of the paper is to develop a numerical implementation of a three-dimensional BIE method[30] that is more efficient, reliable, and robust than the previously mentioned techniques. Section II introduces the boundary integral formulations based on Krutitskii's method. Section III transforms the related integral equations into a form of containing bounded kernels only. The ordinary integral formulation of CHIEF is derived in Sec. IV. Section V examines the availability and accuracy of the modified formulations. Numerical examples include the acoustic radiation and scattering by one sphere and a pair of spheres. Section VI concludes this paper.

II. FORMULATION OF THE PROBLEM AND KRUTITSKII'S METHOD

The propagation of acoustic waves through an unbounded ideal homogeneous medium is described by the Helmholtz wave equation

$$(\nabla^2 + k^2)\phi = 0, \tag{1}$$

where ∇^2 is the Laplacian operator in three dimensions, k is the wave number, and ϕ is the velocity potential. In Eq. (1) a steady-state excitation with a time factor $\exp(-i\omega t)$ is assumed, where $i=\sqrt{-1}$, ω is the angular frequency, and t is the time. The Neumann boundary condition on a sufficiently smooth body surface ∂B can be written as

$$\frac{\partial \phi}{\partial n_r} = f(r), \tag{2}$$

where $\partial/\partial n_r$ denotes differentiation along the outward normal direction at $r \in \partial B$ and $f(r)$ is a given function. The velocity potential ϕ should also meet the Sommerfeld radiation conditions at infinity, i.e.,

$$\phi = O(r^{-1}), \quad \frac{\partial \phi}{\partial r} - ik\phi = O(r^{-1}), \tag{3}$$

as $r \to \infty$. The solution to the boundary-value problem, Eqs. (1)–(3), exists and is unique.[32]

The previous problem can be transformed to a Fredholm integral equation of the second kind that, however, suffers from either the nonuniqueness or nonexistence problem, as was mentioned in Sec. I. Now we give a brief description of the boundary integral equation method proposed by Krutitskii,[30] where these difficulties were circumvented. We introduce an auxiliary surface ∂b interior to ∂B with the boundary condition[30]

$$\frac{\partial \phi(r)}{\partial n_r} - i\phi(r) = 0, \quad r \in \partial b, \tag{4}$$

where $\partial/\partial n_r$ denotes differentiation along the outward normal direction at $r \in \partial b$. Both normal vectors on the real and interior boundaries are notably pointing to the acoustic medium (infinity). The surface ∂b is chosen in such a way[30] that for any k from the set $[0,k_0]$ the following interior Dirichlet problem bounded by ∂b:

$$(\nabla^2 + k^2)\phi = 0, \tag{1'}$$

$$\phi(r) = 0, \quad r \in \partial b, \tag{5}$$

has only the trivial solution; in particular, the inner surface ∂b can be taken such that its diameter d satisfies the estimation

$$d<\ln\left(1+\frac{1}{2k_0^2}\right), \tag{6}$$

where k_0 is an arbitrary fixed positive number and $k \in [0,k_0]$. A convenient selection of ∂b for computation is taking ∂b as a sphere[30] with radius r^* that satisfies the relationship $r^*<\pi/k_0$, where $\pi\approx 3.1416...$ is the first characteristic value of the problem (1) and (5) for a unit sphere. This selection criterion is adopted for all test examples presented in Sec. V. Any solution to the modified boundary-value problem, Eqs. (1)–(4), is also a solution to the original one, Eqs. (1)–(3).[30]

To satisfy some additional smoothness conditions, we have to look for the unknown function ϕ in a form of the single layer potential on $\partial B\cup\partial b$:[30]

$$\phi(p)=\int_{\partial B\cup\partial b}\sigma(q)G_k(p,q)\,dS(q), \tag{7}$$

where σ is the source density. The free-space Green's function G_k for the Helmholtz equation is given by

$$G_k(p,q)=\frac{e^{ikR(p,q)}}{4\pi R(p,q)}, \tag{8}$$

where $R(p,q)$ denotes the distance between the field point p and the source point q. Further application of boundary conditions (2) and (4) leads to[30]

$$-\frac{\sigma(p)}{2}+\int_{\partial B\cup\partial b}\sigma(q)\frac{\partial G_k}{\partial n_p}dS(q)=f(p),\quad p\in\partial B, \tag{9}$$

$$-\frac{\sigma(p)}{2}+\int_{\partial B\cup\partial b}\sigma(q)\left(\frac{\partial}{\partial n_p}-i\right)G_k dS(q)=0,\quad p\in\partial b. \tag{10}$$

The representation of Eqs. (9) and (10) is also called an indirect approach due to solving the source density σ first, rather than the velocity potential ϕ. The nonexistence problem is known to appear when using Eq. (9) alone; however, a combination of Eqs. (9) and (10) gives rise to unique solutions for all wave frequencies $k\in[0,k_0]$.[30] Note that k_0 is an arbitrary positive fixed number, so it can be taken as large as we need. The additional constraint, Eq. (10), is therefore used to overcome the difficulty of the characteristic frequency. The CHIEF method similarly applies a related interior integral equation as the constraint for the solution. Unlike the CHIEF points, the positions of the interior points for the current approach are explicitly determined.

Consider next the acoustic scattering of a plane wave by sound hard obstacles. The velocity potential ϕ becomes

$$\phi(p)=\int_{\partial B\cup\partial b}\sigma(q)G_k(p,q)\,dS(q)+\phi^{\text{inc}}(p), \tag{11}$$

where ϕ^{inc} denotes the incident velocity potential. The two related boundary integral equations can be written in a similar manner to the radiation case as follows:

$$-\frac{\sigma(p)}{2}+\int_{\partial B\cup\partial b}\sigma(q)\frac{\partial G_k}{\partial n_p}dS(q)+\frac{\partial\phi^{\text{inc}}(p)}{\partial n_p}=0,\quad p\partial B, \tag{12}$$

$$-\frac{\sigma(p)}{2}+\int_{\partial B\cup\partial b}\sigma(q)\left(\frac{\partial}{\partial n_p}-i\right)G_k\,dS(q)+\frac{\partial\phi^{\text{inc}}(p)}{\partial n_p}-i\phi^{\text{inc}}(p)=0,\quad p\in\partial b. \tag{13}$$

Again, the velocity potential ϕ is obtained from Eq. (11) after solving the source density σ from the boundary integral equation (12) and the supplemental equation (13).

III. REDUCTION OF WEAKLY SINGULAR INTEGRANDS TO BOUNDED INTEGRANDS IN INTEGRAL EQUATIONS

The systems of Eqs. (7), (9), and (10) for the radiation problem and Eqs. (11)–(13) for the scattering problem notably contain weakly singular kernels when $p\to q$. The polar coordinate transformation has been a traditional technique to deal with such singular integrals; standard Gaussian quadrature is then applied after the discretization. See, for example, Ref. 12 for the detailed implementation. This section converts the singular integral equations to a form of containing bounded kernels. Directly applying standard quadrature is amenable to solution without other specific treatment. The current regularization is therefore carried out in a global sense, rather than locally along each element as in the conventional boundary element methods.

Applying the subtraction-addition technique, see, e.g., Ref. 33 among others, in Eq. (9) yields

$$-\frac{\sigma_1(p)}{2}+\int_{\partial B}\sigma_1(q)\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)dS(q)+\int_{\partial B}(\sigma_1(q)-\sigma_1(p))\frac{\partial G}{\partial n_q}dS(q)+\sigma_1(p)\int_{\partial B}\frac{\partial G}{\partial n_q}dS(q)+\int_{\partial b}\sigma_2(q)\frac{\partial G_k}{\partial n_p}dS(q)=f(p),\quad p\in\partial B, \tag{14}$$

where σ_1 is the source function on the body surface ∂B and σ_2 is the source function on the auxiliary inner surface ∂b. The free-space Green's function $G(p,q)=(4\pi R)^{-1}$ is the fundamental solution of the Laplace equation. Further substituting the Gauss flux theorem,[4]

$$\int_{\partial B}\frac{\partial G(p,q)}{\partial n_q}dS(q)=-\frac{1}{2},\quad p\in\partial B, \tag{15}$$

into Eq. (14) results in

$$-\sigma_1(p)+\int_{\partial B}\sigma_1(q)\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)dS(q)+\int_{\partial B}(\sigma_1(q)-\sigma_1(p))\frac{\partial G}{\partial n_q}dS(q)+\int_{\partial b}\sigma_2(q)\frac{\partial G_k}{\partial n_p}dS(q)=f(p),\quad p\in\partial B, \tag{16}$$

where

$$\lim_{p\to q}\left[\sigma_1(q)\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)\right]=0,$$

$$\lim_{p\to q}\left[(\sigma_1(q)-\sigma_1(p))\frac{\partial G}{\partial n_q}\right]=0,$$

and the third integrand is regular since $p\in\partial B$ and $q\in\partial b$, i.e., $p\neq q$. Equation (16) is therefore singular free since all integrands are bounded and explicitly determined when $p\to q$.

Equation (10) can be rewritten as, again in terms of the subtraction-addition technique,

$$-\frac{\sigma_2(p)}{2}+\int_{\partial B}\sigma_1(q)\left(\frac{\partial}{\partial n_p}-i\right)G_k\,dS(q)+\int_{\partial b}\sigma_2(q)\times\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)dS(q)+\int_{\partial b}(\sigma_2(q)-\sigma_2(p))\times\frac{\partial G}{\partial n_q}dS(q)+\sigma_2(p)\int_{\partial b}\frac{\partial G}{\partial n_q}dS(q)-i\int_{\partial b}\sigma_2(q)\times(G_k-G)\,dS(q)-i\int_{\partial b}(\sigma_2(q)-\sigma_2(p))G\,dS(q)-i\sigma_2(p)\int_{\partial b}G\,dS(q)=0,\quad p\in\partial b. \tag{17}$$

The analytical integration value of the last singular integral does not exist in general for an arbitrarily shaped body. Section II suggested selecting the interior surface ∂b as a spherical surface with radius r^*. The last integral of Eq. (17) then becomes

$$\int_{\partial b}G\,dS(q)=r^*, \tag{18}$$

The identity is a result of the application of equipotential body.[34] Substitution of Eq. (18) and the Gauss flux theorem (15) into Eq. (17) leads to

$$-\sigma_2(p)+\int_{\partial B}\sigma_1(q)\left(\frac{\partial}{\partial n_p}-i\right)G_k\,dS(q)+\int_{\partial b}\sigma_2(q)\times\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)dS(q)+\int_{\partial b}(\sigma_2(q)-\sigma_2(p))\times\frac{\partial G}{\partial n_q}dS(q)-i\int_{\partial b}\sigma_2(q)(G_k-G)\,dS(q)-i\times\int_{\partial b}(\sigma_2(q)-\sigma_2(p))G\,dS(q)-i\sigma_2(p)r^*=0,\quad p\in\partial b, \tag{19}$$

where the first integral is regular since $p\in\partial b$ and $q\in\partial B$,

$$\lim_{p\to q}\left[\sigma_2(q)\left(\frac{\partial G_k}{\partial n_p}-\frac{\partial G}{\partial n_q}\right)\right]=0,$$

$$\lim_{p\to q}\left[(\sigma_2(q)-\sigma_2(p))\frac{\partial G}{\partial n_q}\right]=0,$$

and $\lim_{p\to q}[(\sigma_2(q)-\sigma_2(p))G]=0$. In the fourth integral, we have

$$\lim_{p\to q}(G_k-G)=\frac{ik}{4\pi}. \tag{20}$$

The identity is obvious from standard calculus. Equation (19) is therefore a boundary integral equation of the second kind containing bounded integration functions only. At the first glance the modified equations (16) and (19) are much more complicated than the original equations (9) and (10). The computation, however, requires only two more terms, i.e., G and $\partial G/n_q$. Moreover, these two terms are of body shape dependence only; that is, they are not functions of the wave frequency or the boundary condition of a given acoustical problem.

The velocity potential ϕ in the radiation field B_e can be calculated from Eq. (7) after solving the source functions σ_1 and σ_2. If one is concerned with the excess acoustic pressure p, the relationship between p and ϕ can be expressed as follows:

$$p=i\omega\rho_0\phi, \tag{21}$$

where ρ_0 denotes the density of the fluid at the equilibrium state. For ϕ on the body surface ∂B, Eq. (7) contains a weakly singular integral and can be transformed to the following regular form:

$$\phi(p)=\int_{\partial B}\sigma_1(q)(G_k-G)\,dS(q)+\int_{\partial B}\left[\sigma_1(q)-\sigma_1(p)\frac{\sigma^*(q)}{\sigma^*(p)}\right]G\,dS(q)-\frac{\sigma_1(p)}{\sigma^*(p)}\Phi_e+\int_{\partial b}\sigma_2(q)G_k\,dS(q),\quad p\in\partial B, \tag{22}$$

where $\sigma^*\neq 0$ is a source distribution on ∂B and Φ_e is an equipotential defined by

$$\Phi_e=-\int_{\partial B}\sigma^*(q)G\,dS(q). \tag{23}$$

Readers are referred to Ref. 34 for the proof that a unique solution of (23) exists. The first integrand in Eq. (22) is bounded when $p\to q$, as was given in Eq. (20); the second integrand is bounded since its value can be set equal to zero when $p\to q$; the last integral is regular since $p\neq q$. Equation (23) also contains a singular integral when $p\to q$. The equipotential Φ_e is known to remain constant in the interior domain B_i; its value can then be computed by locating the field point p in B_i so as to avoid the singularity. The source function σ^* can be obtained in terms of the fact that the velocity inside an equipotential body vanishes. Therefore, taking the normal derivative of Eq. (23) at $p\in\partial B$ yields

$$\sigma^*(p)=-2\int_{\partial B}\sigma^*(q)\frac{\partial G}{\partial n_p}dS(q),\quad p\in\partial B. \tag{24}$$

The nonsingular form of the preceding equation can be written as

$$\sigma^*(p) = -2\int_{\partial B}\left[\sigma^*(q)\frac{\partial G}{\partial n_p} - \sigma^*(p)\frac{\partial G}{\partial n_q}\right] dS(q) - 2\sigma^*(p)\int_{\partial B}\frac{\partial G}{\partial n_q} dS(q), \quad p \in \partial B. \tag{25}$$

Further applying the Gauss flux theorem (15) in the last integral of Eq. (25) leads to

$$0 = \int_{\partial B}\left[\sigma^*(q)\frac{\partial G}{\partial n_p} - \sigma^*(p)\frac{\partial G}{\partial n_q}\right] dS(q), \quad p \in \partial B. \tag{26}$$

The equipotential Φ_e can then be obtained via Eq. (23) after solving σ^* from Eq. (26). Substitution of σ^* and Φ_e into Eq. (22) yields the surface function ϕ by directly evaluating regular integrals.

The main advantage of the above singularity-free integral equations lies in that the numerical solution can be obtained by directly applying standard quadrature rules. Previous experience in executing the proposed formulation is almost not required when compared with, say, CHIEF or the off-boundary approach. The corresponding singularity-free representation of Eqs. (11)–(13) for the scattering problem can be obtained straightforwardly according to the previously presentation.

IV. CHIEF METHOD

The CHIEF method for solving exterior acoustic radiation problem is based on the Helmholtz integral relation

$$\int_{\partial B}\left[\phi(q)\frac{\partial G_k}{\partial n_q} - G_k\frac{\partial \phi}{\partial n_q}\right] dS(q) = \begin{cases} \phi(p), & p \in B_e, \\ \phi(p)/2, & p \in \partial B, \\ 0, & p \in B_i, \end{cases} \tag{27}$$

where B_e is the domain exterior to ∂B and B_i is the domain interior to ∂B. Schenck[18] indicated that there is a unique solution satisfying a combination of equations for $p \in \partial B$ and $p \in B_i$ in (27). This is also called a direct approach due to solving ϕ directly. When p is on ∂B, the singular integral equation (27) can be recast to

$$\int_{\partial B}\left(\phi(q)\frac{\partial G_k}{\partial n_q} - \phi(p)\frac{\partial G}{\partial n_q}\right) dS(q) - \int_{\partial B}(G_k - G) \times \frac{\partial \phi}{\partial n_q} dS(q) - \int_{\partial B} G\left(\frac{\partial \phi}{\partial n_q} - \frac{\partial \phi}{\partial n_p}\right) dS(q) - \frac{\partial \phi}{\partial n_p} \times \left[\int_{\partial B}\left(1 - \frac{\sigma^*(q)}{\sigma^*(p)}\right) G\, dS(q) - \frac{\Phi_e}{\sigma^*(p)}\right] = \phi(p), \tag{28}$$

where the equipotential Φ_e and source function σ^* were defined in Eq. (23). The first, third, and fourth integrands can be set to zero as $p \to q$; the second integrand is bounded as $p \to q$, as was given in Eq. (20). Equation (28) is therefore an integral equation containing regular integrals for solving the radiation problem.

The Helmholtz integral relation for the scattering problem can be written as

$$\phi^{\text{inc}}(p) + \int_{\partial B}\phi(q)\frac{\partial G_k}{\partial n_q} dS(q) = \begin{cases} \phi(p), & p \in B_e, \\ \phi(p)/2, & p \in \partial B, \\ 0, & p \in B_i. \end{cases} \tag{29}$$

When p is on ∂B, the associated nonsingular integral equation can be written as

$$\phi^{\text{inc}} + \int_{\partial B}\left(\phi(q)\frac{\partial G_k}{\partial n_q} - \phi(p)\frac{\partial G}{\partial n_q}\right) dS(q) = \phi(p). \tag{30}$$

Again, a combination of Eq. (29) for $p \in B_i$ and Eq. (30) results in a unique solution when selecting effective interior points for the supplemental equations.

V. NUMERICAL RESULTS

A. Application to radiation

The first test example is the acoustic radiation from a uniformly pulsating sphere with radius a. The spherical polar coordinates (R^*, θ, υ) are related to the rectangular Cartesian coordinates (x,y,z) by the transformation $x = R^* \sin\theta\cos\upsilon$, $y = R^* \sin\theta\sin\upsilon$, and $z = R^*\cos\theta$, where $0 \leqslant \theta \leqslant \pi$ and $0 \leqslant \upsilon \leqslant 2\pi$. The radiation problem was simulated by placing a point source of unit strength at the center of the sphere. The acoustic pressure and the velocity at the body surface were then explicitly determined; this velocity was used as the known boundary condition.

An auxiliary inner sphere of radius $r^* = 0.2a$ was placed at the center of the radiating sphere when implementing Krutitskii's method. The numerical procedures first combined Eqs. (16) and (19) for solving the surface source functions σ_1 on the radiating sphere and σ_2 on the inner sphere. The 30- and 60-point Gauss–Legendre quadrature formulas were applied along θ and υ axes, respectively, for the radiating sphere, and the three-point formula was applied for the inner sphere. The resulting system of linear algebraic equations was then solved in terms of the LU decomposition method. Equation (22) finally evaluated the surface function ϕ after solving σ^* from Eq. (26). The nodal, integration, and collocation points that are usually used in the classical boundary element methods notably all imply the same thing when implementing the current singularity-free integral formulation.

In application of the CHIEF method, a combination of the modified equation (28) and the related interior integral relation (27) for $p \in B_i$ solved the surface velocity potential ϕ. A good CHIEF point must be away from the nodal surfaces. For a sphere, the characteristic frequencies are known to appear when

$$\phi(R^*, \theta, \upsilon)\big|_{R^* = a} = \begin{Bmatrix} \sin m\upsilon \\ \cos m\upsilon \end{Bmatrix}\{P_n^m(\cos\theta)\}\{j_n(ka\}\} = 0, \tag{31}$$

where P_n^m is the associated Legendre function, and j_n is the spherical Bessel function. Three CHIEF points were selected and placed at $(x,y,z) = (0,0,0)$, $(0.1a,0,0)$, and $(0,0.2a,0)$ after carefully observing the behaviors of the characteristic functions P_n^m and j_n. The resulting overdetermined system of equations was reconstructed into an ordinary one by the

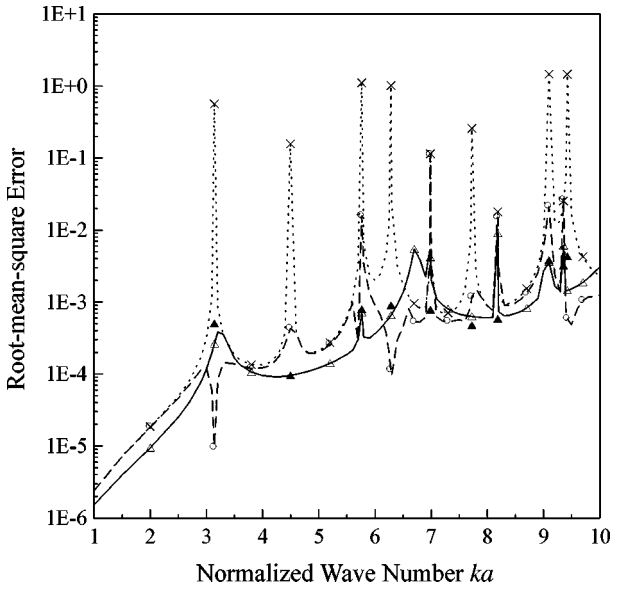


FIG. 1. Root-mean-square error in the computed surface function ϕ relative to the analytical solution for a uniformly pulsating sphere of radius a using the standard simple source method (-×-), the CHIEF method (-○-), and Krutitskii's method with $r^*=0.2a$ (-△-) and $r^*=0.1a$ (▲) where r^* is the radius of the auxiliary inner sphere.

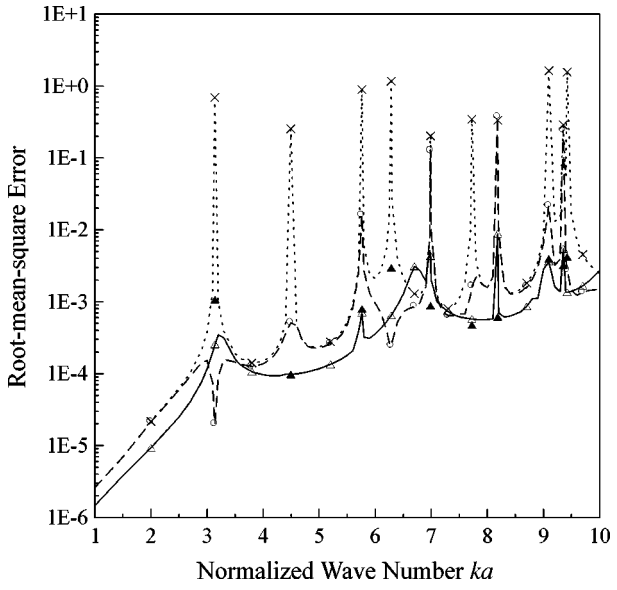


FIG. 2. Root-mean-square error in the computed surface function ϕ relative to the analytical solution for a pair of uniformly pulsating spheres of radius a and with centers at a distance $d=3a$. Other definitions are given in Fig. 1.

least-squares technique. The remaining discretization system and numerical procedures resembled that of the present method.

Figure 1 displays the root-mean-square errors in the computed surface function ϕ relative to the analytical solution using three different approaches, i.e., the standard simple source method, the CHIEF method, and Krutitskii's method. The standard simple source method implies using Eq. (16) alone. The range of normalized wave number ka varies from 1 to 10, including ten characteristic numbers $ka=3.1416$ (π), 4.4934, 5.7635, 6.2832 (2π), 6.9879, 7.7253, 8.1826, 9.095, 9.3558. 9.4248 (3π). The standard simple source formulation fails to predict the surface function at the characteristic wave numbers, as observed from Fig. 1. The CHIEF method performs extremely well at the spherically symmetric modes $ka=\pi$, 2π, 3π, and has correction effect at the nonspherically symmetric modes, except $ka=6.9879$, 8.1826, and 9.3558. Carrying out a large number of tries on CHIEF points still could not generate satisfactory results for all characteristic wave numbers. Krutitskii's method correlates well with the analytical solution at all characteristic wave numbers. Figure 1 also plots the case of using an auxiliary inner sphere of radius $r^*=0.1a$ for examining the influence of its size. The associated root-mean-square errors at the ten characteristic wave numbers are indicated by the solid triangle. The great accuracy demonstrates that the selection of the inner sphere is not critical.

Consider next two pulsating spheres each of radius a and with centers at a distance $d=3a$. A unit source was placed at the center of each sphere for modeling the radiation problem. In this case, ∂B in Eq. (7) for the velocity potential ϕ is replaced by $\partial B_1 \cup \partial B_2$, where ∂B_1 and ∂B_2 are the surfaces of the two spheres, and ∂b by $\partial b_1 \cup \partial b_2$, where ∂b_1 and ∂b_2 are the two associated interior surfaces. Reduction of weakly singular integrands to bounded integrands in integral equations for two spheres follows the method described in Sec. III. Figure 2 plots the related root-mean-square errors in the numerical results against the normalized wave number ka. Krutitskii's method with either $r^*=0.2a$ or $0.1a$ agrees well with the analytical solution. Again, the standard simple source method fails at all characteristic wave numbers and the CHIEF method fails at some nonspherically symmetric characteristic wave numbers. A large number of CHIEF points were also tested for eliminating the nonuniqueness difficulty, but none of them succeeded. Dokumaci and Sarigül[35] presented this finding in the same physical problem.

B. Application to scattering

The above three different methods were further examined by the acoustic scattering from a sphere of radius a exposed to a plane wave of unit amplitude, $\phi^{\text{inc}}=\exp(-ikz)$, incident in the direction of the negative z axis (Fig. 3). The discretization system and the numerical procedures resembled that of the previous radiation problem. Figure 3 illustrates the computed amplitudes of the surface function ϕ for two normalized characteristic wave numbers $ka=3.1416$ and 5.7635. The corresponding exact solution of the surface function is also plotted and can be written as[36]

$$\phi=\frac{i}{(ka)^2}\sum_{n=0}^{\infty}(-i)^n(2n+1)\frac{P_n(\cos\theta)}{h_n^{(1)'}(ka)}, \tag{32}$$

where P_n is the Legendre function of the first kind and $h_n^{(1)}$ is the spherical Hankel function. Figure 3 reveals that Krutitskii's method, in which $r^*=0.2a$ was used, and the CHIEF method successfully predict the acoustic scattering field. In the case of $ka=3.1416$, the root-mean-square errors are 6.8×10^{-4} and 9.6×10^{-5} for Krutitskii's method and the CHIEF method, respectively; in the case of $ka=5.7635$, the

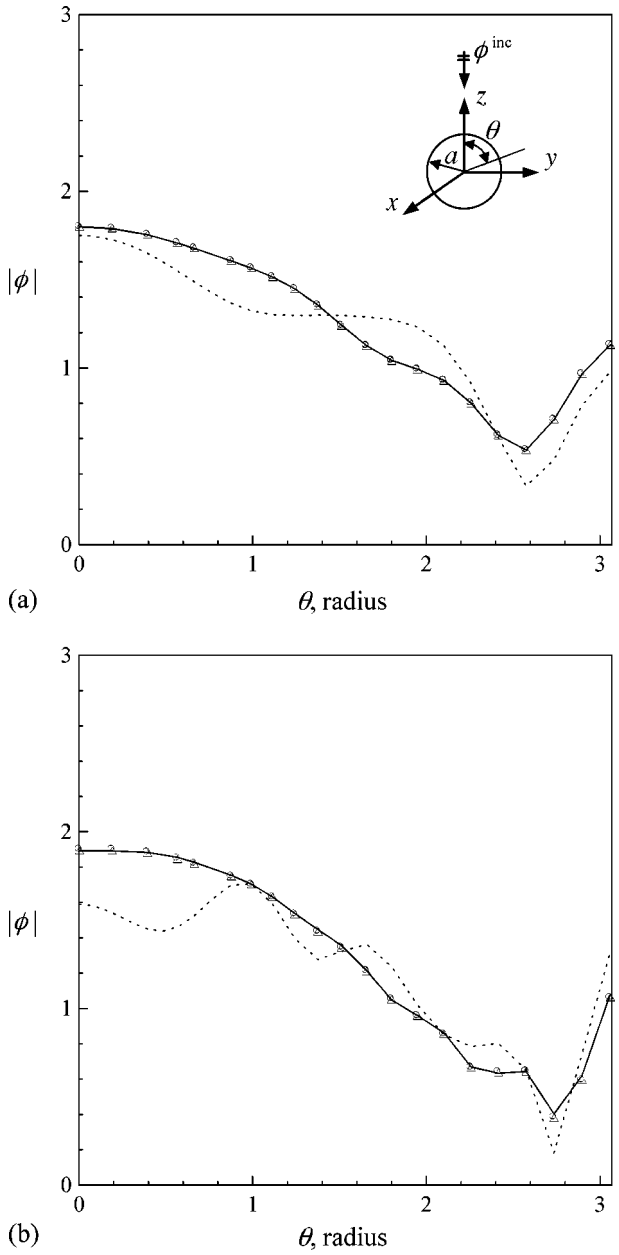


FIG. 3. Amplitude of surface function ϕ for a sphere of radius a exposed to a plane wave ϕ^{inc} with normalized wave number (a) $ka=3.1416$ and (b) $ka=5.7635$ using the standard simple source method (……), the CHIEF method (○), Krutitskii's method (……), and the analytical solution (△).

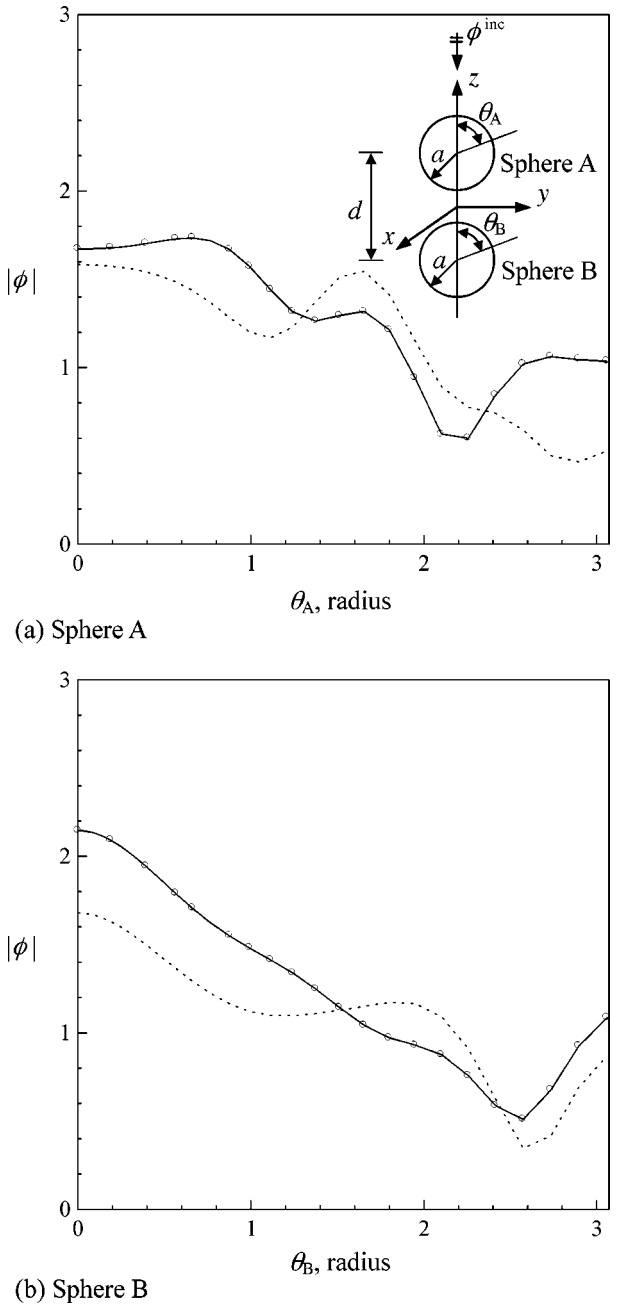


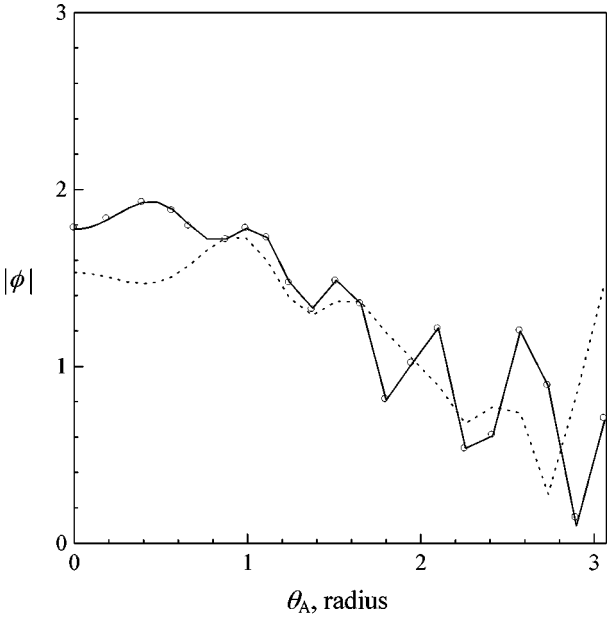
FIG. 4. Amplitude of surface function ϕ for a pair of spheres of radius a exposed to a plane wave ϕ^{inc} with normalized wave number $ka=3.1416$. The separation distance between sphere centers is $d=3a$. Other definitions are given in Fig. 3.

root-mean-square errors are 5.9×10^{-3} and 3.7×10^{-3}, respectively. The standard simple source method breaks down at the two characteristic wave numbers.

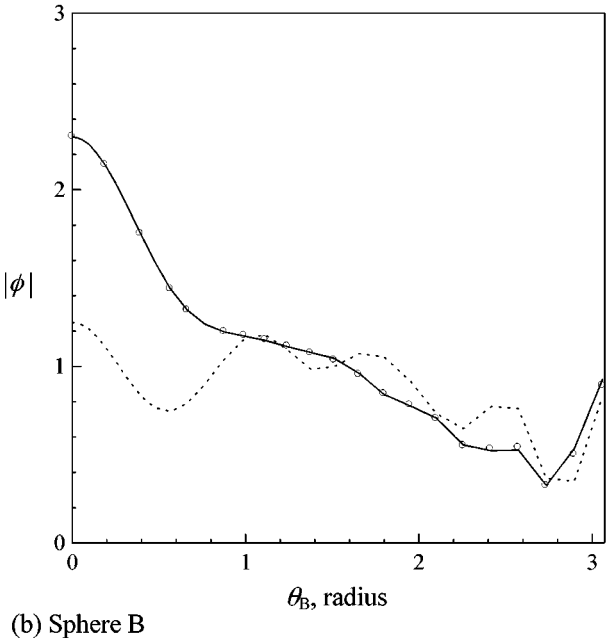
Consider finally the acoustic scattering by a pair of identical spheres (Fig. 4). Figures 4 and 5 depict the computed amplitudes of the surface function ϕ for $ka=3.1416$ and 5.7635, respectively. Figure 6 displays the mutual interaction of the two scatterers with respect to the variation of the separation distance d for the case of $ka=5.7635$. Krutitskii's method and the CHIEF method match well for $ka=3.1416$ and 5.7635, as observed from Figs. 4 and 5. According to previous test examples, the numerical results displayed in Fig. 5 for $ka=5.7635$ using the standard simple source method is not reliable. Figure 6 demonstrates that the significant interactions occur in the areas of the shadowed side of the sphere A and the illuminated side of the sphere B. Moreover, results are seen to converge to the case of single sphere with an increase in the separation distance d, and the mutual effects almost disappear for $d=20a$. Note in passing that we refer the reader to Refs. 37–40 for the analytical or approximate solutions of the acoustic scattering by a pair of spheres. However, in addition to containing certain limitation of each treatment, numerically evaluating these solutions requires unique techniques.

VI. CONCLUSIONS AND REMARKS

The foregoing illustrative examples demonstrate the availability of Krutitskii's method for solving three-

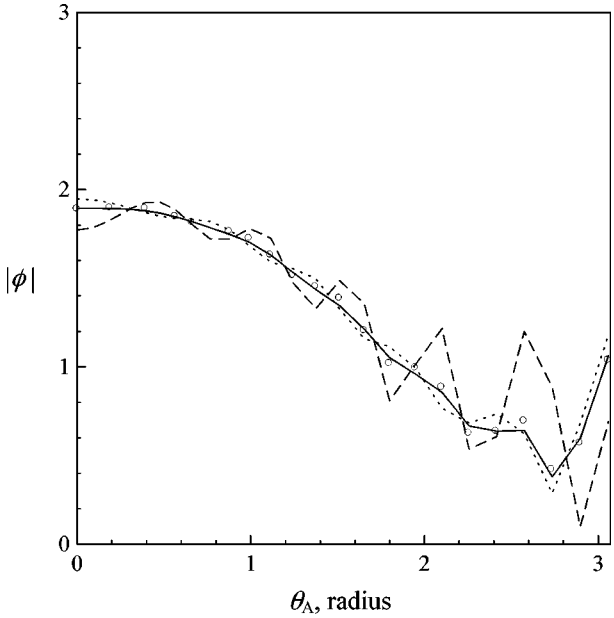


(a) Sphere A

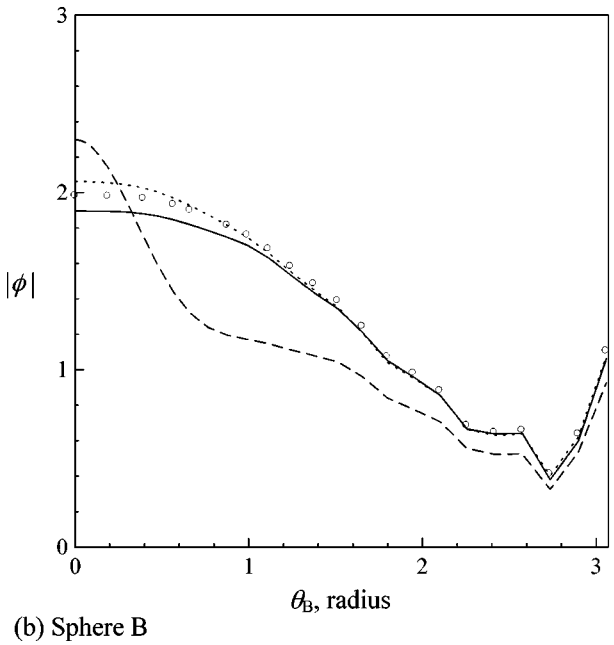


(b) Sphere B

FIG. 5. Same as Fig. 4 except $ka=5.7635$.



(a) Sphere A



(b) Sphere B

FIG. 6. Variation of computed amplitude of surface function ϕ against the separation distance $d=3a$ (---), $10a$ (. . . . .), and $20a$ (○) for a pair of spheres of radius a exposed to a plane wave ϕ^{inc} with normalized wave number $ka=5.7635$. The case of $d\rightarrow\infty$ (———) is also plotted for comparison using the analytical solution for one sphere.

dimensional acoustic radiation and scattering problems. The singularity-free formulation allows a direct use of standard quadrature without resorting to other specialized techniques. This is also referred to as a global method since the regularization is executed before the numerical implementation. In other words, the singular integrands are smoothed out over the entire integration domain before the discretization. The collocation points are exactly the positions at which the integration points are located. Standard boundary element methods usually apply, say, a polar coordinate transformation on each element, and then carry out integration by standard quadrature. This approach is referred to as a local method. The integration points are usually much more than the collocation points. Evaluation of the integrands (or construction of the matrix elements) is known to take a substantial amount of computational time in the solution. The global method is thus generally more accurate and efficient than the local one. See, for example, Ref. 41 for the detailed discussion when treating potential flow problems.

Good agreement is obtained between Krutitskii's method and the analytical solutions throughout the region of the wave number ka varying from one to ten. Further carefully examining the numerical results at the characteristic wave numbers confirms the method's reliability of overcoming the nonexistence problem. The selection of the auxiliary inner surface is explicitly determined so that previous experience in effectively selecting the interior collocation points

is not required. Moreover, only a relatively few number of interior points are sufficient to achieve an acceptable accuracy. This fact implies the high efficiency of the method.

The paper presents a promising alternative, when compared with other existing methods, for analyzing the related acoustical problems. A direct boundary integral formulation based on the proposed approach may be preferable to the indirect one since the velocity potential can be directly obtained. The use of Krutitskii's method for treating sound soft bodies is another possible extension.

ACKNOWLEDGMENTS

The author would like to thank the National Science Council, R.O.C., for partially supporting this research under Contract No. NSC91-2611-E-006-020.

[1] C. Pozrikidis, ''Boundary element grid optimization for Stokes flow with corner singularities,'' J. Fluids Eng. **124**, 22–28 (2002).
[2] J. A. Liggett and P. L.-F. Liu, *The Boundary Integral Equation Method for Porous Media Flow* (George Allen & Unwin, London, 1983).
[3] L. Morino, ''Boundary integral equations in aerodynamics,'' Appl. Mech. Rev. **46**, 445–466 (1993).
[4] M. A. Jaswon and G. T. Symm, *Integral Equation Methods in Potential Theory and Elastostatics* (Academic, London, 1977).
[5] C. Richter, *A Green's Function Time-domain BEM of Elastodynamics* (Computational Mechanics, Southampton, 1997).
[6] *Advances in Boundary Element Methods for Fracture Mechanics*, edited by M. H. Aliabadi and C. A. Brebbia (Computational Mechanics, Southampton, 1993).
[7] *Boundary Element Techniques in Geomechanics*, edited by G. D. Manolis and T. G. Davies (Computational Mechanics, Southampton, 1993).
[8] A. Farcas, L. Elliott, D. B. Ingham, D. Lesnic, and N. S. Mera, ''A dual reciprocity boundary element method for the regularized numerical solution of the inverse source problem associated to the Poisson equation,'' Inverse Probl. Eng. **11**, 123–139 (2003).
[9] A. J. Burton and G. F. Miller, ''The application of integral equation methods to the numerical solution of some exterior boundary value problems,'' Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[10] J. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (Yale U.P., New Haven, CT, 1923).
[11] G. Krishnasamy, L. W. Schmerr, T. J. Rudolphi, and F. J. Rizzo, ''Hypersingular boundary integral equations: Some applications in acoustic and elastic wave scattering,'' J. Appl. Mech. **57**, 404–414 (1990).
[12] C. C. Chien, H. Rajiyah, and S. N. Atluri, ''An effective method for solving the hypersingular integral equations in 3D acoustics,'' J. Acoust. Soc. Am. **88**, 918–937 (1990).
[13] S. A. Yang, ''A boundary integral equation method for two-dimensional acoustic scattering problems,'' J. Acoust. Soc. Am. **105**, 93–105 (1999).
[14] Y. Liu and F. J. Rizzo, ''A weakly singular form of the hypersingular boundary integral equation applied to 3D acoustic wave problems,'' Comput. Methods Appl. Mech. Eng. **96**, 271–287 (1992).
[15] S. A. Yang, ''Evaluation of 2D Green's boundary formula and its normal derivative using Legendre polynomials, with an application to acoustic scattering problems,'' Int. J. Numer. Methods Eng. **53**, 905–927 (2002).
[16] Z. Y. Yan, K. C. Hung, and H. Zheng, ''Solving the hypersingular boundary integral equation in three-dimensional acoustics using a regularization relationship,'' J. Acoust. Soc. Am. **113**, 2674–2683 (2003).
[17] P. J. Harris and K. Chen, ''On efficient preconditioners for iterative solution of a Galerkin boundary element equations for the three-dimensional exterior Helmholtz problem,'' J. Comput. Appl. Math. **156**, 303–318 (2003).
[18] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[19] A. F. Seybert and T. K. Rengarajan, ''The use of CHIEF to obtain solutions for acoustic radiation using boundary integral equations,'' J. Acoust. Soc. Am. **81**, 1299–1306 (1987).
[20] D. J. Segalman and D. W. Lobitz, ''SuperCHIEF: A modified CHIEF method,'' SANDIA Labs Report SAND-90-1266 (1990).
[21] A. F. Seybert and T. W. Wu, BEMAP *User's Manual* (Spectronics, New York, 1989), Version 2.4.
[22] T. W. Wu and A. F. Seybert, ''A weighted residual formulation for the CHIEF method in acoustics,'' J. Acoust. Soc. Am. **90**, 1608–1614 (1991).
[23] J. T. Chen, L. W. Liu, and H. K. Hong, ''Spurious and true eigensolutions of Helmholtz BIEs and BEMs for a multiply connected problem,'' Proc. R. Soc. London, Ser. A **459**, 1891–1924 (2003).
[24] W. Benthien and H. A. Schenck, ''Nonexistence and nonuniqueness problems associated with integral equation methods in acoustics,'' Comput. Struct. **65**, 295–305 (1997).
[25] J. D. Achenbach, J. D. Kechter, and Y.-L. Xu, ''Off boundary approach to the boundary element method,'' Comput. Methods Appl. Mech. Eng. **70**, 191–201 (1988).
[26] K. A. Cunefare, G. Koopmann, and K. Brod, ''A boundary element method for acoustic radiation valid for all wavenumbers,'' J. Acoust. Soc. Am. **85**, 39–48 (1989).
[27] G. H. Koopmann, L. Song, and J. B. Fahnline, ''A method for computing acoustic fields based on the principle of wave superposition,'' J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[28] D. T. Wilton, I. C. Mathews, and R. A. Jeans, ''A clarification of nonexistence problems with superposition method,'' J. Acoust. Soc. Am. **94**, 1676–1680 (1993).
[29] R. D. Miller, E. T. Moyer, Jr., H. Huang, and H. Überall, ''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).
[30] P. A. Krutitskii, ''A new approach to reduction of the Neumann problem in acoustic scattering to a nonhypersingular integral equation,'' IMA J. Appl. Math. **64**, 259–269 (2000).
[31] S. A. Yang, ''A boundary integral equation method using auxiliary interior surface approach for acoustic radiation and scattering in two dimensions,'' J. Acoust. Soc. Am. **112**, 1307–1317 (2002).
[32] D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley, New York, 1983).
[33] S. A. Yang, ''Acoustic scattering by a hard or soft body across a wide frequency range by the Helmholtz integral equation method,'' J. Acoust. Soc. Am. **102**, 2511–2520 (1997).
[34] M. A. Jaswon and G. T. Symm, *Integral Equation Methods in Potential Theory and Elastostatics* (Academic, London, 1977).
[35] E. Dokumaci and A. S. Sarigül, ''Analysis of the near field acoustic radiation characteristics of two radially vibrating spheres by the Helmholtz integral equation formulation and a critical study of the efficacy of the CHIEF overdetermination method in two-body problems,'' J. Sound Vib. **187**, 781–798 (1995).
[36] *Electromagnetic and Acoustic Scattering by Simple Shapes*, edited by J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi (Hemisphere, New York, 1987).
[37] C. Liang and Y. T. Lo, ''Scattering by two spheres,'' Radio Sci. **2**, 1481–1495 (1967).
[38] G. C. Gaunaurd, H. Huang, and H. C. Strifors, ''Acoustic scattering by a pair of spheres,'' J. Acoust. Soc. Am. **98**, 495–507 (1995).
[39] P. Gabrielli and M. Mercier-Finiori, ''Acoustic scattering by two spheres: Multiple scattering and symmetry considerations,'' J. Sound Vib. **241**, 423–439 (2001).
[40] Z. Ye and A. Alvarez, ''An analytic approach to sound scattering from a pair of scatterers with application to rigid spheres,'' J. Comput. Acoust. **9**, 311–325 (2001).
[41] S. A. Yang, ''A solution method for two-dimensional potential flow about bodies with smooth surfaces by directly use of the boundary integral equation,'' Commun. Numer. Methods Eng. **15**, 469–478 (1999).

On Webster's horn equation and some generalizations

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

(Received 17 October 2003; revised 18 May 2004; accepted 2 June 2004)

Sound waves along a rigid axisymmetric tube with a variable cross-section are considered. The governing Helmholtz equation is solved using power-series expansions in a stretched radial coordinate, leading to a hierarchy of one-dimensional ordinary differential equations in the longitudinal direction. The lowest approximation for axisymmetric motion turns out to be Webster's horn equation. Fourth-order differential equations are obtained at the next level of approximation. Comparisons with existing asymptotic theories for waves in slender tubes are made. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1775272]

PACS numbers: 43.20.Mv [LLT] Pages: 1381–1388

I. INTRODUCTION

Webster's horn equation (1919) gives a one-dimensional approximation for low-frequency sound waves along a rigid tube with a variable cross-sectional area $A(z)$. The equation itself can be written as

$$\frac{1}{A}\frac{d}{dz}\left(A(z)\frac{dP}{dz}\right)+k^2P(z)=0, \tag{1}$$

where $k=\omega/c$, ω is the frequency, c is the (constant) speed of sound, and z is the coordinate along the tube. Equation (1) can be derived by considering a thin layer of the fluid at z, perpendicular to the z-direction, with the assumption that the acoustic pressure is constant over this layer; this pressure is $P(z)$. For a succinct derivation, see (Pierce, 1989, p. 360). Alternative derivations and extensions to other problems (such as with fluid flow through the tube) are available; see, for example, (Reinstra, 2002) and (Hélie, 2003). Exact solutions of Eq. (1) are available for various specific $A(z)$; for some recent results, see (Kumar and Sujith, 1997).

In 1967, Eisner published an excellent review of early work based on Eq. (1). On p. 1127, we read: ''Eq. (1) is usually called 'Webster's horn equation,' but we see that there is little justification for this name. Daniel Bernoulli, Euler, and Lagrange all derived the equation and did most interesting work on its solution, more than 150 years before Webster.''

We are interested in giving a systematic derivation of Webster's equation, and of higher-order variants. We limit our analysis to axisymmetric tubes, and obtain Webster's equation when we seek axisymmetric solutions. We also consider nonaxisymmetric motions. To be specific, we consider a tube of finite length, and seek the frequencies of free vibration of the compressible fluid within the closed tube.

In 1916 (three years before Webster's paper), Lord Rayleigh published a paper on axisymmetric motions in an axisymmetric tube. He began by writing the general solution of the axisymmetric Helmholtz equation in cylindrical polar coordinates (r,θ,z) as

$$u(r,\theta,z)=J_0\left(r\sqrt{\frac{d^2}{dz^2}+k^2}\right)u_0(z), \tag{2}$$

where $u_0(z)=u(0,\theta,z)$ is the value of u on the axis of the tube; as the Bessel function $J_0(w)$ has a power-series expansion in integer powers of w^2, this provides meaning to the right-hand side of Eq. (2) [see Eq. (26) below]. Rayleigh then obtained an equation for $u_0(z)$ by applying the boundary condition on the rigid wall of the tube; as a first approximation, he obtained Eq. (1) [see Eq. (8) in (Rayleigh, 1916)]. He also discussed some higher-order approximations. We shall adopt a similar approach, although we do not begin with an explicit representation such as Eq. (2): we shall use a Frobenius-type power-series expansion for the radial variation of $u(r,\theta,z)$.

After formulating the problem in Sec. II, we examine the special case of circular cylinders in Secs. III and IV. The exact solution for the eigenfrequencies is recalled in Sec. III. Then, the approximate method is developed in Sec. IV. It is based on some observations of Boström (2000) for the related axisymmetric problems of elastic waves in isotropic rods. In principle, we can obtain a heirarchy of approximations: We give explicit results for the first two members of this heirarchy. Apart from the merits of explaining the method for a simple case, we are also able to give a quantitative comparison with the exact solutions from Sec. III.

In Sec. V, we consider axisymmetric, noncylindrical tubes. We change variables in the governing Helmholtz equation from r and z to ρ and ζ, where ρ is a scaled version of r chosen to that the lateral boundary is mapped to $\rho=$constant; this has the effect of complicating the partial differential equation (via the chain rule) but has the virtue that the lateral boundary condition is applied on a coordinate surface. The complications bring the shape of the boundary into the differential equation [cf. Webster's equation (1)] but they can be dealt with readily because we then use a Frobenius-type expansion in the new ''radial'' variable ρ. Again, we obtain a hierarchy of approximations. The first approximation gives a second-order ordinary differential equation; it reduces to Webster's equation for axisymmetric motions. The second approximation gives a fourth-order ordinary differential

a)Electronic mail: pamartin@mines.edu

equation. In order to assess these approximations, we compare with some results of Ting and Miksis (1983) and Geer and Keller (1983). These authors began with the governing elliptic boundary-value problem for waves in slender tubes, and then obtained various asymptotic approximations. The comparisons are made in Sec. VI. Some concluding remarks can be found in Sec. VII.

In summary, the approximate method described below has two virtues. First, the approximations can be improved. Second, the use of power series means that the basic method can be applied to much more complicated equations of motion, and to systems of such equations. For example, the propagation of elastic waves in nonuniform anisotropic rods can be studied; for some preliminary results in this direction, see (Martin, 2004).

II. FORMULATION

Consider a tube of circular cross-section and length L. Using cylindrical polar coordinates, $(\tilde{r},\theta,\tilde{z})$, the interior of the tube is specified by

$$0\leqslant\tilde{r}<a\tilde{R}(\tilde{z}),\quad 0\leqslant\theta<2\pi,\quad 0<\tilde{z}<L,$$

where $0<\tilde{R}(\tilde{z})\leqslant 1$, so that $2a$ is the maximum diameter of the tube. It will be convenient to define dimensionless variables, using L as our length scale. Thus, we put

$$r=\tilde{r}/L,\quad z=\tilde{z}/L,\quad \tilde{R}(\tilde{z})=R(z)\quad\text{and}\quad \varepsilon=a/L.$$

(Later, we shall regard ε as a small parameter.) Hence, the tube becomes

$$0\leqslant r<\varepsilon R(z),\quad 0\leqslant\theta<2\pi,\quad 0<z<1.$$

Inside the tube, the acoustic potential $U(r,\theta,z,t)$ satisfies the wave equation

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial U}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2 U}{\partial\theta^2}+\frac{\partial^2 U}{\partial z^2}=\frac{L^2}{c^2}\frac{\partial^2 U}{\partial t^2},\tag{3}$$

where c is the speed of sound. On the lateral wall of the tube, the normal derivative of U vanishes

$$\frac{\partial U}{\partial r}-\varepsilon R'(z)\frac{\partial U}{\partial z}=0\quad\text{on}\quad r=\varepsilon R(z),\quad 0<z<1.\tag{4}$$

We assume that $R(0)$ and $R(1)$ are both positive, and close the two ends of the tube with rigid circular discs, giving

$$\partial U/\partial z=0\quad\text{at}\quad z=0\quad\text{and}\quad\text{at}\quad z=1.\tag{5}$$

We seek free vibrations of the compressible fluid within the axisymmetric tube. Thus, we put $U(r,\theta,z,t)=u(r,z)\cos m\theta\cos\omega t$, where m is a non-negative integer and ω is the frequency. The wave equation becomes

$$r\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right)+r^2\frac{\partial^2 u}{\partial z^2}+(k^2r^2-m^2)u=0,\tag{6}$$

where (now) $k=\omega L/c$ is a dimensionless wave number. We are interested in determining *eigenfrequencies* ω so that there is a nontrivial u that satisfies Eq. (6) and the boundary conditions.

III. CIRCULAR TUBE: EXACT SOLUTION

Consider a circular tube with $R(z)=1$ for $0<z<1$. We can write down separated solutions of Eq. (6), $u(r,z)=J_m(\lambda r)\cos\mu z$, where J_m is a Bessel function, and λ and μ are arbitrary constants satisfying $k^2=\lambda^2+\mu^2$. The boundary conditions give $\lambda\varepsilon=j'_{m,\ell}$ and $\mu=N\pi$, where $N\geqslant 0$ and $\ell\geqslant 1$ are integers and $j'_{m,\ell}$ is the ℓ-th zero of J'_m: $J'_m(j'_{m,\ell})=0$, $\ell=1,2,\ldots,$. Hence,

$$k^2=(N\pi)^2+(j'_{m,\ell})^2\varepsilon^{-2}.\tag{7}$$

This is the exact solution of the problem. It can be found in textbooks [for example, p. 221 of (Kinsler *et al.*, 1982)] and was given by Lord Rayleigh in the first edition of volume II of *The Theory of Sound*, published in 1878 (Rayleigh, 1945).

A. Axisymmetric modes ($m=0$)

The axisymmetric problem ($m=0$) is unusual, because we can take $\lambda=0$: both Eq. (6) and the lateral boundary condition on $r=\varepsilon$ are satisfied by $u(r,z)=\cos\mu z$, so that one set of eigenfrequencies is given by

$$k=N\pi\quad\text{for all}\quad\varepsilon>0.\tag{8}$$

This set is elementary; it can be found in Sec. 62 of Lamb's book (1960).

The next set comes by using $j'_{0,2}\simeq 3.832$

$$k^2=(N\pi)^2+14.68\varepsilon^{-2},\quad N=0,1,2,\ldots,.\tag{9}$$

B. Flexural modes ($m=1$)

We are also interested in nonaxisymmetric motions. As an example, we consider the case $m=1$ (''flexural modes''). As $j'_{1,1}\simeq 1.841$ and $j'_{1,2}\simeq 5.331$, Eq. (7) gives the first two sets of eigenfrequencies as

$$k^2=(N\pi)^2+3.39\varepsilon^{-2}\quad\text{and}\quad k^2=(N\pi)^2+28.42\varepsilon^{-2}\tag{10}$$

for $N=0,1,2,\ldots$, later, we will compare the coefficients 3.39 and 28.42 with those obtained by certain approximate theories.

IV. CIRCULAR TUBE: APPROXIMATE METHOD

If the tube is slender, $\varepsilon\equiv a/L\ll 1$, we expect to be able to derive one-dimensional theories. We shall do this using power series in r. Note that we do not limit ourselves to *polynomials* in r, and so we are not limited, in principle, to very long waves. Nevertheless, it turns out that the low-order truncations obtained below work best for longer waves.

We begin, as in the method of Frobenius, by writing

$$u(r,z)=\sum_{n=0}^{\infty} r^{2n+\alpha}u_n(z),\tag{11}$$

where α and $u_n(z)$ are to be found. Substitution in Eq. (6) gives

$$(\alpha^2-m^2)u_0(z)+\sum_{n=0}^{\infty} r^{2n+2}\{\sigma_n(\alpha)u_{n+1}+u_n''+k^2u_n\}=0,$$

where $\sigma_n(\alpha)=(2n+2+\alpha)^2-m^2$. Just as in the method of Frobenius for ordinary differential equations, we require that

every coefficient of r^ℓ vanishes. For the first term to vanish, we obtain $\alpha^2=m^2$. As we want u to be bounded at $r=0$, we take $\alpha=+m$, and then we obtain

$$u_n''(z)+k^2u_n(z)=-4(n+1)(n+m+1)u_{n+1}(z),$$
$$n=0,1,2,\ldots,.$$

Notice that this procedure does not determine $u_0(z)$. However, following Boström (2000), we note that u_1, $u_2,\ldots,$ are all determined by u_0; for example, we have

$$u_1=-\frac{u_0''+k^2u_0}{4(m+1)} \tag{12}$$

and

$$u_2=-\frac{u_1''+k^2u_1}{8(m+2)}=\frac{u_0^{\mathrm{iv}}+2k^2u_0''+k^4u_0}{32(m+1)(m+2)}.$$

Then, *regardless of the choice of* u_0, the infinite series in Eq. (11) will give an exact solution of Eq. (6), assuming that the series converges.

The end boundary conditions, Eq. (5), give $\Sigma_{n=0}^{\infty}r^{2n}u_n'(z)=0$ at the two ends, whence

$$u_n'(0)=u_n'(1)=0, \quad n=0,1,2,\ldots,. \tag{13}$$

The lateral boundary condition reduces to $\partial u/\partial r=0$ on $r=\varepsilon$, and this gives

$$\sum_{n=0}^{\infty}(2n+m)\varepsilon^{2n}u_n(z)=0, \quad 0<z<1.$$

Eliminating u_n in favor of u_0, we obtain an equation for $u_0(z)$

$$\begin{aligned}0&=mu_0+(m+2)\varepsilon^2u_1+(m+4)\varepsilon^4u_2+\cdots\\&=mu_0-\frac{(m+2)\varepsilon^2}{4(m+1)}(u_0''+k^2u_0)\\&\quad+\frac{(m+4)\varepsilon^4}{32(m+1)(m+2)}(u_0^{\mathrm{iv}}+2k^2u_0''+k^4u_0)+\cdots.\end{aligned} \tag{14}$$

At this stage, no approximations have been made. We obtain various approximations by truncating Eq. (14); this is done next.

A. First approximation

If we discard all terms with powers of ε greater than 2 in Eq. (14), we obtain

$$u_0''(z)+\mathcal{E}_m^{(1)}u_0(z)=0, \tag{15}$$

where

$$\mathcal{E}_m^{(1)}(k,\varepsilon)=k^2-[4m(m+1)/(m+2)]\varepsilon^{-2}. \tag{16}$$

From Eq. (13), we see that Eq. (15) is to be solved subject to $u_0'(0)=u_0'(1)=0$.

If we look for solutions of Eq. (15) in the form

$$u_0(z)=\cos\mu z, \tag{17}$$

we find that $\mu^2=\mathcal{E}_m^{(1)}(k,\varepsilon)$. Then, in order to satisfy the boundary conditions on the two ends of the tube, we obtain $\mu=N\pi$, whence

$$k^2=(N\pi)^2+[4m(m+1)/(m+2)]\varepsilon^{-2}.$$

This can be compared with the exact solution, Eq. (7). For $m=0$, we recover the exact set of solutions given by Eq. (8). For $m=1$, we can compare with Eq. (10); when $m=1$, we have $4m(m+1)/(m+2)=8/3\simeq 2.67$ which is in error by about 20%.

B. Second approximation

If we retain the terms in ε^4 in Eq. (14), we obtain

$$u_0^{\mathrm{iv}}(z)+\mathcal{C}_mu_0''(z)+\mathcal{E}_m^{(2)}u_0(z)=0, \tag{18}$$

where

$$\mathcal{C}_m(k,\varepsilon)=2k^2-\frac{8(m+2)^2}{(m+4)\varepsilon^2}, \tag{19}$$

$$\mathcal{E}_m^{(2)}(k,\varepsilon)=k^4+\frac{32(m+1)(m+2)}{(m+4)\varepsilon^4}\left[m-\frac{(m+2)(k\varepsilon)^2}{4(m+1)}\right]. \tag{20}$$

The fourth-order Eq. (18) is to be solved subject to

$$u_0'(0)=u_0'''(0)=u_0'(1)=u_0'''(1)=0, \tag{21}$$

using Eq. (12) to express u_1' in terms of u_0''' and u_0', and Eq. (13).

Substituting Eq. (17) in Eq. (18) gives

$$\mu^4-\mu^2\mathcal{C}_m(k,\varepsilon)+\mathcal{E}_m^{(2)}(k,\varepsilon)=0, \tag{22}$$

and then a routine calculation yields

$$\mu^2=k^2-\nu_m/\varepsilon^2=\mu_m^2(k,\varepsilon), \tag{23}$$

say, where $\nu_m=\chi_m+4(m+2)^2/(m+4)$ and

$$\chi_m=\pm\left(1+\frac{1}{4}m\right)^{-1}\sqrt{(m+2)(8+4m-4m^2-m^3)}. \tag{24}$$

Finally, the end boundary conditions give

$$k^2=(N\pi)^2+\nu_m/\varepsilon^2. \tag{25}$$

1. Axisymmetric modes (m=0)

When $m=0$, Eq. (25) reduces to Eq. (8) when we take the minus sign in $\pm$. If we take the plus sign, we obtain $k^2=(N\pi)^2+8\varepsilon^{-2}$, where the coefficient 8 can be compared with the exact 14.68 in Eq. (9).

For this axisymmetric problem, we can compare with Rayleigh's approach. As $J_0(w)=1-1/4w^2+1/64w^4-\cdots$, Eq. (2) gives

$$\begin{aligned}u(r,z)&=u_0(z)-\tfrac{1}{4}r^2(u_0''+k^2u_0)\\&\quad+\tfrac{1}{64}r^4(u_0^{\mathrm{iv}}+2k^2u_0''+k^4u_0)-\cdots.\end{aligned} \tag{26}$$

If we discard the higher-order terms and apply the boundary condition, $\partial u/\partial r=0$ on $r=\varepsilon$, we obtain precisely Eq. (18) with $m=0$. This is reassuring but not surprising: The representations given by Eqs. (2) and (11) are equivalent, although the latter can be used for nonaxisymmetric problems.

2. Flexural modes (m=1)

When $m=1$, we obtain

$$k^2=(N\pi)^2+\tfrac{4}{5}(9\pm\sqrt{21})\varepsilon^{-2}. \tag{27}$$

To obtain the lowest values, we take the minus sign, giving $k^2=(N\pi)^2+3.53\varepsilon^{-2}$; the coefficient 3.53 differs from the exact 3.39 by about 4%. Thus, the fourth-order model gives good accuracy for the lowest set of eigenfrequencies.

If we take the plus sign in Eq. (27), the coefficient multiplying ε^{-2} becomes 10.87, which can be compared with the exact value of 28.42 given in the second of Eq. (10). Thus, as when $m=0$, the second set of eigenfrequencies is not approximated well by the fourth-order model.

V. NONCYLINDRICAL TUBES

Once we move away from cylinders, we no longer have exact solutions. Therefore, an approximate method will be required. For this reason, we choose to make a simple change of the independent variables, from (r,z) to (ρ,ζ), so that the new geometry is a circular cylinder. The price of this change is that the new partial differential equation is more complicated.

Thus, define new variables ρ and ζ by

$$\rho=r/R(z) \quad \text{and} \quad \zeta=z, \tag{28}$$

so that the tube is mapped onto the circular tube, given by $0\leqslant\rho<\varepsilon$, $0<\zeta<1$. (Later, we will use z in place of ζ, but it is clearer to distinguish the two variables at this stage.) The chain rule gives

$$\frac{\partial u}{\partial r}=\frac{1}{R}\frac{\partial u}{\partial\rho}, \quad \frac{\partial^2 u}{\partial r^2}=\frac{1}{R^2}\frac{\partial^2 u}{\partial\rho^2},$$

$$\frac{\partial u}{\partial z}=\frac{\partial u}{\partial\zeta}-\rho\frac{R'}{R}\frac{\partial u}{\partial\rho},$$

$$\frac{\partial^2 u}{\partial z^2}=\frac{\partial^2 u}{\partial\zeta^2}-2\rho\frac{R'}{R}\frac{\partial^2 u}{\partial\rho\partial\zeta}-\rho\frac{\partial u}{\partial\rho}\left(\frac{R'}{R}\right)'+\rho\left(\frac{R'}{R}\right)^2\frac{\partial}{\partial\rho}\left(\rho\frac{\partial u}{\partial\rho}\right).$$

Hence, Eq. (6) becomes

$$(1+\rho^2R'^2)\rho\frac{\partial}{\partial\rho}\left(\rho\frac{\partial u}{\partial\rho}\right)+\rho^3(R'^2-RR'')\frac{\partial u}{\partial\rho}+\rho^2R^2\frac{\partial^2 u}{\partial\zeta^2}-2\rho^3RR'\frac{\partial^2 u}{\partial\rho\partial\zeta}+[(k\rho R)^2-m^2]u=0, \tag{29}$$

the lateral boundary condition, Eq. (4), becomes

$$(1+\varepsilon^2R'^2)\frac{\partial u}{\partial\rho}-\varepsilon RR'\frac{\partial u}{\partial\zeta}=0 \quad \text{on} \quad \rho=\varepsilon, \quad 0<\zeta<1, \tag{30}$$

and the end boundary conditions, Eq. (5), become

$$\frac{\partial u}{\partial\zeta}-\rho\frac{R'}{R}\frac{\partial u}{\partial\rho}=0 \quad \text{at} \quad \zeta=0 \quad \text{and} \quad \text{at} \quad \zeta=1. \tag{31}$$

To solve Eq. (29), we proceed as in Sec. IV, and write

$$u(\rho,\zeta)=\sum_{n=0}^{\infty}\rho^{2n+\alpha}u_n(\zeta). \tag{32}$$

Substitution in Eq. (29) gives

$$(\alpha^2-m^2)u_0(\zeta)+\sum_{n=0}^{\infty}\rho^{2n+2}\{\sigma_n(\alpha)u_{n+1}+\Lambda_n(\zeta;\alpha)\}=0,$$

where $\sigma_n(\alpha)=(2n+2+\alpha)^2-m^2$, and

$$\Lambda_n(z;\alpha)=R^2u_n''-2RR'(2n+\alpha)u_n'+[k^2R^2+(2n+\alpha)^2R'^2+(2n+\alpha)(R'^2-RR'')]u_n.$$

As before, we take $\alpha=+m$ and then

$$\Lambda_n(z;m)=-4(n+1)(n+m+1)u_{n+1}(z), \quad n=0,1,2,\ldots,.$$

In particular, we find that

$$\begin{aligned}-4(m+1)u_1&=R^2u_0''-2mRR'u_0'+[k^2R^2+m(m+1)R'^2-mRR'']u_0,\\ -8(m+2)u_2&=R^2u_1''-2(m+2)RR'u_1'+[k^2R^2+(m+2)\{(m+3)R'^2-RR''\}]u_1.\end{aligned} \tag{33}$$

Eliminating u_1 from the last equation, using Eq. (33), gives

$$32(m+1)(m+2)u_2=\alpha_mu_0^{\mathrm{iv}}+\beta_mu_0'''+\gamma_mu_0''+\delta_mu_0'+\varepsilon_mu_0, \tag{34}$$

where

$$\begin{aligned}\alpha_m(z)&=R^4, \quad \beta_m(z)=-4mR^3R',\\ \gamma_m(z)&=2R^2\{k^2R^2+3m[(m+1)R'^2-RR'']\},\\ \delta_m(z)&=-4mR\{k^2R^2R'+R^2R'''-(m+1)R'[3RR''-(m+2)R'^2]\},\\ \varepsilon_m(z)&=k^4R^4+2mk^2R^2\{(m+1)R'^2-RR''\}-mR^3R^{\mathrm{iv}}+m(m+1)\{R^2(4R'R'''+3R''^2)-(m+2)R'^2[6RR''-(m+3)R'^2]\}.\end{aligned}$$

When Eq. (32) is substituted in the lateral boundary condition Eq. (30), together with $\alpha=m$, we obtain

$$0=mu_0(z)+\sum_{n=0}^{\infty}\varepsilon^{2n+2}\{(2n+2+m)u_{n+1}(z)+(2n+m)R'^2u_n-RR'u_n'\}. \tag{35}$$

Similarly, the end boundary conditions Eq. (31) give

$$\sum_{n=0}^{\infty}\rho^{2n}\left[u_n'(z)-\frac{R'}{R}(2n+m)u_n(z)\right]=0 \quad \text{at} \quad z=0 \quad \text{and} \quad \text{at} \quad z=1, \tag{36}$$

which immediately gives

$$R(z)u_n'(z)-R'(z)(2n+m)u_n(z)=0 \quad \text{at } z=0$$
$$\text{and} \quad \text{at } z=1, \tag{37}$$

for $n=0,1,2,\ldots,$.

Below, we shall use the following shorthand notation:

$$\mathcal{S}_1(z)=R'/R, \quad \mathcal{S}_2=R''/R, \quad \mathcal{S}_3=R'''/R, \quad \mathcal{S}_4=R^{\mathrm{iv}}/R. \tag{38}$$

A. First approximation

If we retain only the terms up to ε^2 in Eq. (35), we find that

$$0=mu_0\varepsilon^{-2}+(m+2)u_1+mR'^2u_0-RR'u_0'.$$

Upon using Eq. (33), this gives

$$u_0''(z)+D_m^{(1)}(z)u_0'(z)+E_m^{(1)}(z)u_0(z)=0, \tag{39}$$

where

$$D_m^{(1)}(z)=2\mathcal{S}_1(2-m^2)/(m+2),$$

$$E_m^{(1)}(z)=\mathcal{E}_m^{(1)}(k,\varepsilon R)+\frac{m(m+1)(m-2)}{m+2}\mathcal{S}_1^2-m\mathcal{S}_2,$$

and $\mathcal{E}_m^{(1)}$ are defined by Eq. (16). From Eq. (37), Eq. (39) is to be solved subject to $Ru_0'-mR'u_0=0$ at $z=0$ and at $z=1$.

Equation (39) can be transformed so as to eliminate the first-derivative term. Thus, put

$$u_0(z)=R^\gamma U_0(z) \quad \text{with} \quad \gamma=(m^2-2)/(m+2). \tag{40}$$

Then, we find that $U_0(z)$ solves

$$U_0''(z)+[k^2-K(z)]U_0(z)=0, \tag{41}$$

where

$$K(z)=\frac{2(m+1)}{m+2}\left\{\frac{m}{m+2}\mathcal{S}_1^2+\mathcal{S}_2+\frac{2m}{\varepsilon^2R^2}\right\},$$

subject to $(m+2)U_0'-2(m+1)\mathcal{S}_1U_0=0$ at $z=0$ and at $z=1$. Notice that Eq. (41) has oscillatory solutions when $k^2>K$ but exponential solutions when $k^2<K$.

Equation (39) and its associated boundary conditions can also be written as a regular Sturm-Liouville problem. Thus,

$$(p(z)u_0')'+[q(z)+\lambda w(z)]u_0(z)=0,$$

where $p=w=R^{-2\gamma}$, $\lambda=k^2$, γ is defined by Eq. (40), and

$$q(z)=mR^{-2\gamma}\left\{\frac{m+1}{m+2}\left[(m-2)\mathcal{S}_1^2-\frac{4}{\varepsilon^2R^2}\right]-\mathcal{S}_2\right\}.$$

1. Axisymmetric modes (m=0)

When $m=0$, Eq. (39) reduces to

$$(Au_0')'+k^2Au_0=0, \tag{42}$$

where $A(z)=\pi a^2[R(z)]^2$ is the area of the (circular) cross-section at z; Eq. (42) is recognized as Webster's horn equation Eq. (1). The appropriate boundary conditions are $u_0'(0)=u_0'(1)=0$.

2. Flexural modes (m=1)

When $m=1$, Eq. (39) reduces to

$$u_0''(z)+\frac{2}{3}\mathcal{S}_1u_0'(z)+\left[k^2-\frac{2}{3}\mathcal{S}_1^2-\mathcal{S}_2-\frac{8}{3\varepsilon^2R^2}\right]u_0(z)=0,$$

subject to $Ru_0'-R'u_0=0$ at $z=0$ and at $z=1$.

B. Second approximation

If we retain the terms up to ε^4 in Eq. (35), we obtain an equation containing u_0, u_0', u_1, u_1', and u_2. If we eliminate u_1, u_1', and u_2, using Eqs. (33) and (34), we obtain a fourth-order equation for $u_0(z)$,

$$u_0^{\mathrm{iv}}+B_m^{(2)}(z)u_0'''+C_m^{(2)}(z)u_0''+D_m^{(2)}(z)u_0'+E_m^{(2)}(z)u_0=0, \tag{43}$$

where

$$B_m^{(2)}=4\mathcal{S}_1(4-2m-m^2)/(m+4),$$

$$C_m^{(2)}=\mathcal{C}_m(k,\varepsilon R)-6m\{\mathcal{S}_2+(4-m-m^2)\mathcal{S}_1^2(m+4)\},$$

$$D_m^{(2)}=\mathcal{D}_m(k,\varepsilon R)-4m\mathcal{S}_3+\frac{4m\mathcal{S}_1}{m+4}\{(m+1)(4-m^2)\mathcal{S}_1^2+3m(m+3)\mathcal{S}_2\}, \tag{44}$$

$$E_m^{(2)}=\mathcal{E}_m^{(2)}(k,\varepsilon R)-m\mathcal{S}_4+3m(m+1)\mathcal{S}_2^2-2mk^2\left\{\frac{4-m-m^2}{m+4}\mathcal{S}_1^2+\mathcal{S}_2\right\}+\frac{4m^2(m+3)}{m+4}\mathcal{S}_1\mathcal{S}_3-\frac{6m^2(m+1)(m+2)}{m+4}\mathcal{S}_1^2\mathcal{S}_2+\frac{m(m+1)(m+2)}{m+4}(m^2-m-4)\mathcal{S}_1^4-\frac{8m(m+2)}{(m+4)\varepsilon^2R^2}\{(m+1)(m-2)\mathcal{S}_1^2-(m+2)\mathcal{S}_2\},$$

$\mathcal{C}_m$ and $\mathcal{E}_m^{(2)}$ are defined by Eqs. (19) and (20), respectively, and

$$\mathcal{D}_m(k,\varepsilon)=\frac{4\mathcal{S}_1}{m+4}\left\{(4-2m-m^2)k^2+\frac{4}{\varepsilon^2}(m+2)(m^2-2)\right\}. \tag{45}$$

Note that Eq. (43) reduces to Eq. (18) when $R(z)\equiv 1$.

Equation (43) is to be supplemented with boundary conditions. From Eq. (37), these are

$$Ru_0'-mR'u_0=0 \quad \text{and} \quad Ru_1'-(m+2)R'u_1=0 \tag{46}$$

at $z=0$ and at $z=1$. Eliminating u_1 from the second of Eq. (46), using Eq. (33), and then using the first of Eq. (46), we obtain

$$u_0'''-3m\mathcal{S}_1u_0''+\{2(m^2-1)\mathcal{S}_1^2+3\mathcal{S}_2\}u_0'-m\mathcal{S}_3u_0=0; \tag{47}$$

both this condition and $u_0' = m\mathcal{S}_1 u_0$ are to be imposed at $z=0$ and at $z=1$. Notice that the boundary conditions do not involve k^2.

1. Axisymmetric modes (m=0)

When $m=0$, Eq. (43) reduces to

$$u_0^{\rm iv} + \frac{4R'}{R}u_0''' + 2\left(k^2 - \frac{4}{\varepsilon^2 R^2}\right)u_0'' + \frac{4R'}{R}\left(k^2 - \frac{4}{\varepsilon^2 R^2}\right)u_0' + \left(k^4 - \frac{8k^2}{\varepsilon^2 R^2}\right)u_0 = 0. \tag{48}$$

From Eq. (47), we see that Eq. (48) is to be solved subject to Eq. (21). We observe that Eq. (48) involves R and R', but not higher derivatives of R. Also, Eq. (48) can be derived directly by applying Eq. (4) to Eq. (26).

2. Flexural modes (m=1)

When $m=1$, we obtain

$$B_1^{(2)}(z) = \frac{4}{5}\mathcal{S}_1, \quad C_1^{(2)}(z) = 2k^2 - \frac{12}{5}\mathcal{S}_1^2 - 6\mathcal{S}_2 - \frac{72}{5}(\varepsilon R)^{-2},$$

$$D_1^{(2)}(z) = \frac{4}{5}\mathcal{S}_1\{k^2 + 6\mathcal{S}_1^2 + 12\mathcal{S}_2 - 12(\varepsilon R)^{-2}\} - 4\mathcal{S}_3,$$

$$E_1^{(2)}(z) = k^4 - \mathcal{S}_4 + 6\mathcal{S}_2^2 - \frac{2}{5}k^2\{2\mathcal{S}_1^2 + 5\mathcal{S}_2 + 36(\varepsilon R)^{-2}\} + \frac{2}{5}\mathcal{S}_1\{8\mathcal{S}_3 - 18\mathcal{S}_1\mathcal{S}_2 - 12\mathcal{S}_1^3\} + \frac{24}{5\varepsilon^2 R^2}\left\{2\mathcal{S}_1^2 + 3\mathcal{S}_2 + \frac{8}{\varepsilon^2 R^2}\right\}.$$

Then, the differential equation Eq. (43) (with $m=1$) is to be solved subject to

$$u_0' - \mathcal{S}_1 u_0 = 0 \quad \text{and} \quad u_0''' - 3\mathcal{S}_1 u_0'' + 3\mathcal{S}_2 u_0' - \mathcal{S}_3 u_0 = 0$$

at $z=0$ and at $z=1$.

VI. COMPARISON WITH ASYMPTOTIC APPROXIMATIONS

A number of formal asymptotic theories have been developed for waves in slender tubes. In the papers by Ting and Miksis (1983) and by Geer and Keller (1983), the tubes need not have circular cross-sections and they need not be straight. Here, we shall compare the results obtained from our one-dimensional theory with those obtained by specialising the analysis in (Ting and Miksis, 1983) and (Geer and Keller, 1983) to our axisymmetric geometries.

Two asymptotic regimes are of interest to us. In one, $k=O(1)$ as $\varepsilon\to 0$, so that the wavelength is comparable to L, the length of the tube. All solutions of this kind are axisymmetric. In a second regime, $k=O(\varepsilon^{-1})$ as $\varepsilon\to 0$, so that the wavelength is comparable to $a \ll L$. Both kinds of solution are seen in the exact solutions for circular tubes (Sec. III).

A. Wavelength comparable to L

For axisymmetric motion, we obtained the fourth-order Eq. (48), which we write here as

$$(R^2 u_0')' + k^2 R^2 u_0 = \tfrac{1}{8}\varepsilon^2 R^3 (R u_0^{\rm iv} + 4R' u_0''' + 2k^2 R u_0'' + 4k^2 R' u_0' + k^4 R u_0). \tag{49}$$

In this equation, k and $u_0(z)$ are unknown; they are to be determined subject to the four boundary conditions, Eq. (21).

Let us look for solutions of the form

$$k^2 = k_0^2 + \varepsilon^2 k_2^2 + \cdots, \tag{50}$$

where k_0 and k_2 are to be found. Thus, we put $u_0(z) = v_0(z) + \varepsilon^2 v_1(z) + \cdots$ and Eq. (50) in Eq. (49). This is a classic singular perturbation, because Eq. (49) reduces to a second-order equation when $\varepsilon=0$, implying that the boundary conditions will have to be modified. Indeed, writing out the first two terms of the exact boundary condition, Eq. (36), we have

$$0 = u_0'(z) + \rho^2[u_1' - 2(R'/R)u_1] + \cdots = u_0'(z) - \tfrac{1}{4}\rho^2 R^2 (u_0'' + k^2 u_0)' + \cdots,$$

at each end of the tube, where we have used Eq. (33). We will not be able to satisfy this condition for all allowable ρ with $0 \leqslant \rho < \varepsilon$, and so we integrate over each circular end of the tube, and impose

$$0 = u_0'(z) - \tfrac{1}{8}\varepsilon^2 R^2 (u_0'' + k^2 u_0)' + \cdots \quad \text{at } z=0 \quad \text{and} \quad \text{at } z=1; \tag{51}$$

this ensures that Eq. (36) is satisfied in an average sense.

The terms in ε^0 from Eqs. (49) and (51) give

$$(R^2 v_0')' + k_0^2 R^2 v_0 = 0 \quad \text{for } 0<z<1, \quad \text{with} \quad v_0'(0) = v_0'(1) = 0. \tag{52}$$

This is Webster's horn equation again, written as a regular Sturm-Liouville problem: it is an eigenvalue problem for k_0^2 and $v_0(z)$; we normalize the solution using

$$\int_0^1 \{v_0(z)R(z)\}^2 dz = 1. \tag{53}$$

The terms in ε^2 give

$$(R^2 v_1')' + k_0^2 R^2 v_1 = -k_2^2 R^2 v_0 + V, \tag{54}$$

where

$$V = \tfrac{1}{8}R^3(R v_0^{\rm iv} + 4R' v_0''' + 2k_0^2 R v_0'' + 4k_0^2 R' v_0' + k_0^4 R v_0) = \tfrac{1}{4}R^2(R'R'' - RR''')v_0' - \tfrac{1}{2}R^3 R'' v_0', \tag{55}$$

after using Eq. $(52)_1$. Similarly, Eq. (51) gives

$$v_1'(z) = -\tfrac{1}{4}RR' v_0'' = -\tfrac{1}{4}(RR' v_0')' \quad \text{at } z=0 \quad \text{and} \quad \text{at } z=1. \tag{56}$$

Thus v_1 solves a forced version of Webster's horn equation with inhomogeneous boundary conditions. As the homogeneous form of Eq. (54) admits nontrivial solutions, Eq. (54)

will only have solutions if a consistency condition is satisfied. Thus, we multiply Eq. (54) by $v_0(z)$ and Eq. $(52)_1$ by $v_1(z)$, subtract the two, and integrate over $0 \leqslant z \leqslant 1$. This gives

$$\left[-\frac{1}{4} R^2 v_0 (RR' v_0')' \right]_{z=0}^{1} = -k_2^2 + \int_0^1 V v_0 \, dz,$$

where we have used Eqs. $(52)_2$, (53), and (56). In order to cancel the left-hand side, write $V = \{V + 1/4[R^2(RR'v_0')']'\} - 1/4[R^2(RR'v_0')']'$. Then, an integration by parts gives

$$k_2^2 = \frac{1}{4} \int_0^1 R^2 (RR' v_0')' v_0' \, dz + \int_0^1 W v_0 \, dz, \tag{57}$$

where

$$\begin{aligned} W(z) &= V + \tfrac{1}{4}[R^2(RR'v_0')']' \\ &= \tfrac{1}{4} R^3 R' v_0''' + (RR')^2 v_0'' + \tfrac{1}{2}(3R^2 R' R'' + RR'^3) v_0' \end{aligned} \tag{58}$$

and we have used Eq. (55). Integrating the first integral in Eq. (57) by parts, using Eq. $(52)_2$, then gives

$$k_2^2 = -\frac{1}{2} \int_0^1 (RR' v_0')^2 \, dz + K,$$

where $K = \int_0^1 (W v_0 - 1/4 R^3 R' v_0'' v_0') \, dz$. We are going to show that $K = 0$.

Noting the first term in Eq. (58), and making use of Eq. $(52)_1$, we have

$$\begin{aligned} v_0''' v_0 - v_0'' v_0' &= -[k_0^2 v_0' + (2R^{-1} R' v_0')'] v_0 \\ &\quad + (k_0^2 v_0 + 2R^{-1} R' v_0') v_0' \\ &= 2R^{-1} R' v_0'^2 - 2R^{-2}[(RR'' - R'^2) v_0' \\ &\quad + RR' v_0''] v_0. \end{aligned}$$

Hence, $K = 1/2 \int_0^1 \{(RR'v_0')^2 + [(RR')^2 v_0']' v_0\} dz$, and this is seen to vanish after another integration by parts. Thus,

$$k^2 = k_0^2 - \frac{1}{2} \varepsilon^2 \int_0^1 (RR' v_0')^2 \, dz + O(\varepsilon^4) \quad \text{as} \quad \varepsilon \to 0. \tag{59}$$

This elegant formula was derived in a different way by Geer and Keller (1983); see their Eq. (102).

B. Wavelength comparable to *a*

In the differential equations, Eqs. (39) and (43), put $k = \kappa/\varepsilon$ and

$$u_0(z) = E(z) w(z) \quad \text{with} \quad E(z) = \exp\left\{ \frac{i}{\varepsilon} \int_{z_0}^{z} \Phi(t) \, dt \right\}, \tag{60}$$

where $\Phi(t)$ and $w(z)$ are to be determined, and z_0 is a constant. Suppose further that

$$w(z) = w_0(z) + \varepsilon w_1(z) + \varepsilon^2 w_2(z) + \cdots. \tag{61}$$

Then, we obtain the following approximations:

$$\begin{aligned} u_0/E &= w_0 + \varepsilon w_1 + O(\varepsilon^2), \quad u_0'/E = \varepsilon^{-1} i \Phi w_0 + O(1), \\ u_0''/E &= -\varepsilon^{-2} \Phi^2 w_0 + \varepsilon^{-1}(2i\Phi w_0' + i\Phi' w_0 - \Phi^2 w_1) \\ &\quad + O(1), \\ u_0'''/E &= -\varepsilon^{-3} i \Phi^3 w_0 + O(\varepsilon^{-2}), \\ u_0^{\mathrm{iv}}/E &= \varepsilon^{-4} \Phi^4 w_0 + \varepsilon^{-3} \Phi^2 \{\Phi^2 w_1 - 2i(3\Phi' w_0 + 2\Phi w_0')\} \\ &\quad + O(\varepsilon^{-2}), \end{aligned}$$

as $\varepsilon \to 0$.

Let us begin with the second-order equation, Eq. (39). We note that $D_m^{(1)} = O(1)$ and $E_m^{(1)} = \varepsilon^{-2} \mathcal{E}_m^{(1)}(\kappa, R) + O(1)$ as $\varepsilon \to 0$. Then, the terms in ε^{-2} give

$$[\Phi(z)]^2 = \mathcal{E}_m^{(1)}(\kappa, R). \tag{62}$$

The terms in ε^{-1} give

$$2\Phi w_0' + \{\Phi' + \Phi D_m^{(1)}\} w_0 = 0,$$

which is a first-order differential equation for w_0. Rearranging gives

$$\frac{(w_0 \Phi^{1/2})'}{w_0 \Phi^{1/2}} = \frac{w_0'}{w_0} + \frac{\Phi'}{2\Phi} = -\frac{D_m^{(1)}}{2} = -q_m^{(1)} \frac{R'}{R},$$

where $q_m^{(1)} = (2 - m^2)/(m + 2)$. An integration gives

$$w_0^2 \Phi R^{2q} = \text{constant}, \tag{63}$$

where w_0, Φ, and R are all functions of z only, and $q \equiv q_m^{(1)}$.

Next, consider the fourth-order equation, Eq. (43). Substituting as before, we find that

$$\begin{aligned} &B_m^{(2)} = O(1), \quad C_m^{(2)} = \varepsilon^{-2} \mathcal{C}_m(\kappa, R) + O(1), \\ &D_m^{(2)} = \varepsilon^{-2} \mathcal{D}_m(\kappa, R) + O(1) \quad \text{and} \\ &E_m^{(2)} = \varepsilon^{-4} \mathcal{E}_m^{(2)}(\kappa, R) + O(\varepsilon^{-2}) \end{aligned}$$

as $\varepsilon \to 0$. Then, we see that the terms in ε^{-4} give

$$\Phi^4 - \Phi^2 \mathcal{C}_m(\kappa, R) + \mathcal{E}_m^{(2)}(\kappa, R) = 0, \tag{64}$$

which should be compared with Eq. (22). Thus,

$$[\Phi(z)]^2 = \mu_m^2(\kappa, R), \tag{65}$$

where μ_m is defined by Eq. (23).

The terms in ε^{-3} give

$$\begin{aligned} 0 &= \{\Phi^4 - \Phi^2 \mathcal{C}_m(\kappa, R) + \mathcal{E}_m^{(2)}(\kappa, R)\} w_1 \\ &\quad + 2i\Phi w_0' \{\mathcal{C}_m(\kappa, R) - 2\Phi^2\} + i w_0 \{\Phi \mathcal{D}_m(\kappa, R) \\ &\quad + \Phi' \mathcal{C}_m(\kappa, R) - \Phi^3 B_m^{(2)} - 6\Phi^2 \Phi'\}. \end{aligned}$$

The factor multiplying w_1 vanishes, due to Eq. (64), leaving a first-order differential equation for $w_0(z)$

$$\frac{w_0'}{w_0} + \frac{\Phi^2 \mathcal{D}_m(\kappa, R) + \Phi \Phi' \mathcal{C}_m(\kappa, R) - \Phi^4 B_m^{(2)} - 6\Phi^3 \Phi'}{2\Phi^2 [\mathcal{C}_m(\kappa, R) - 2\Phi^2]} = 0.$$

Rearranging this equation gives

$$\frac{(w_0\Phi^{1/2})'}{w_0\Phi^{1/2}} = \frac{4\Phi\Phi' - \mathcal{D}_m(\kappa,R) + \Phi^2 B_m^{(2)}}{2[\mathcal{C}_m(\kappa,R) - 2\Phi^2]} = -q_m^{(2)}\frac{R'}{R},$$

where

$$q_m^{(2)} = [4(m+2)(m^2-2) - m(m+3)\nu_m]/[(m+4)\chi_m].$$

Here, we have used $\Phi^2 = \kappa^2 - \nu_m/R^2$, and Eqs. (19), (23), (44), and (45). Hence, an integration gives Eq. (63) again, but with $q \equiv q_m^{(2)}$.

Let us compare our results (obtained by solving ordinary differential equations asymptotically) with those of Ting and Miksis (1983) and Geer and Keller (1983) (obtained by solving an elliptic boundary-value problem asymptotically). First, we obtained approximations for Φ, given by Eqs. (62) and (65), whereas the exact formula is

$$[\Phi(z)]^2 = \kappa^2 - [j'_{m,\ell}/R(z)]^2. \tag{66}$$

The errors incurred here are exactly the same as those discussed in Sec. IV for a circular tube (constant cross-section): We cannot expect to do any better for tubes of varying cross-section. Second, the analysis of Ting and Miksis (1983) and Geer and Keller (1983) leads to Eq. (63) with $q=1$ (for all m). If we take the minus sign in Eq. (24), we find that $q_0^{(2)} = 1$ and $q_1^{(2)} = 1.43$; it is not clear why this last number is not closer to 1, given that the corresponding values of Φ are close.

VII. CONCLUDING REMARKS

The method of Sec. V for waves in axisymmetric tubes leads to eigenvalue problems for ordinary differential equations. The simplest (first) approximation leads to a regular Sturm–Liouville problem. This is convenient because efficient software is readily available for solving such problems numerically (Pruess and Fulton, 1993). The next (second) approximation is expected to be more accurate, and, indeed, we have shown this in some asymptotic regimes. The second approximation leads to an eigenvalue problem for a fourth-order differential equation. However, it does not fall into the class of regular fourth-order Sturm-Liouville problems; see, for example, the review by Greenberg and Marletta (2000). In fact, we can say little about the theoretical properties of the simplest equation, namely Eq. (48), which models axisymmetric motions.

Evidently, higher-order approximations could be developed, leading to ordinary differential equations of order $2n$ with $n=3,4,...$; the necessary derivations would be expedited using software for symbolic manipulations. The method of Sec. V could also be extended to other cross-sections, using a scaling ρ that depends on the angle θ as well as on the longitudinal coordinate z.

Another possibility is to abandon power series in favor of Neumann series, which are series of Bessel functions of various orders. This would permit better representation of $u(r,z)$ for fixed z, but at the expense of additional complication.

Finally, the basic power-series method can be extended to various elastodynamic problems, generalizing the work of Boström (2000) on rods and of Boström, Johansson, and Olsson (2001) on plates; for axisymmetric motions in non-uniform anisotropic rods, see (Martin, 2004). Indeed, the fact that the power-series method is relatively insensitive to complications in the governing partial differential equations means that it may be worth developing further.

Boström, A. (**2000**). "On wave equations for elastic rods," Z. Angew. Math. Mech. **80**, 245–251.

Boström, A., Johansson, G., and Olsson, P. (**2001**). "On the rational derivation of a hierarchy of dynamic equations for a homogeneous, isotropic, elastic plate," Int. J. Solids Struct. **38**, 2487–2501.

Eisner, E. (**1967**). "Complete solutions of the 'Webster' horn equation," J. Acoust. Soc. Am. **41**, 1126–1146.

Geer, J. F., and Keller, J. B. (**1983**). "Eigenvalues of slender cavities and waves in slender tubes," J. Acoust. Soc. Am. **74**, 1895–1904.

Greenberg, L., and Marletta, M. (**2000**). "Numerical methods for higher order Sturm–Liouville problems," J. Comput. Appl. Math. **125**, 367–383.

Hélie, T. (**2003**). "Unidimensional models of acoustic propagation in axisymmetric wave-guides," J. Acoust. Soc. Am. **114**, 2633–2647.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York).

Kumar, B. M., and Sujith, R. I. (**1997**). "Exact solutions for the longitudinal vibration of non-uniform rods," J. Sound Vib. **207**, 721–729.

Lamb, H. (**1960**). *The Dynamical Theory of Sound*, 2nd ed. (Dover, New York).

Martin, P. A. (**2004**). "Waves in wood: axisymmetric waves in slender solids of revolution," Wave Motion (to be published).

Pierce, A. D. (**1989**). *Acoustics* (Acoustical Society of America, New York).

Pruess, S., and Fulton, C. T. (**1993**). "Mathematical software for Sturm-Liouville problems," ACM Trans. Math. Softw. **19**, 360–376.

Rayleigh, Lord (**1916**). "On the propagation of sound in narrow tubes of variable section," Philos. Mag., Series 6 **31**, 89–96.

Rayleigh, Lord (**1945**). *The Theory of Sound* (Dover, New York).

Reinstra, S. W. (**2002**). "The Webster equation revisited," paper AIAA 2002-2520, 8th AIAA/CEAS Aeroacoustics Conference, Breckenridge, Colorado, June 2002.

Ting, L., and Miksis, M. J. (**1983**). "Wave propagation through a slender curved tube," J. Acoust. Soc. Am. **74**, 631–639.

Webster, A. G. (**1919**). "Acoustical impedance, and the theory of horns and of the phonograph," Proc. Natl. Acad. Sci. U.S.A. **5**, 275–282.

A super-spectral finite element method for sound transmission in waveguides

Andrew Peplow[a] and Svante Finnveden[b]
Marcus Wallenberg Laboratory for Sound and Vibration Research, Department of Aeronautics and Vehicle Engineering, Royal Technical University, S-100 44, Stockholm, Sweden

(Received 17 November 2003; revised 10 June 2004; accepted 10 June 2004)

A super-spectral finite element method is developed for the study of acoustical wave propagation in nonuniform waveguides. The formulation is based on a finite-element approach using a mixture of high order element shape functions and wave solutions. The numerical method provides solutions to acoustic duct or fluid waveguide environments which may be divided into rectangular sectors. Examples of its use for infinite acoustic waveguides include sound transmission through large ambient density variations and propagation over a geometric stair-step perturbation. Computation of a *trapped mode* waveform due to a point volume source within a uniform waveguide is also presented. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1777853]

PACS numbers: 43.20.Mv, 43.20.Bi [JJM] Pages: 1389–1400

I. INTRODUCTION

In this work a numerical method is studied which provides solutions to acoustic waveguide environments divided into independent rectangular sectors. The technique used is a spectral finite element method demonstrated for arbitrary length acoustic waveguides including sound transmission through a density variation, wave propagation over a stair-step perturbation and computation of a *trapped mode* waveform due to a point volume source.

Spectral finite element methods are now an established alternative to finite difference and finite element methods to solve elliptic partial differential equations (PDE). Spectral methods are naturally chosen to solve problems in regular rectangular, cylindrical or spherical regions. However, in a general irregular region it would be unwise to turn away from the finite element method since models defined in such regions are extremely difficult to implement and solve with a spectral method. However, for many problems in fluid and ocean-wave acoustic propagation the use of the spectral finite element technique in conjunction with global domain decomposition appears to be attractive since they combine the flexibility of finite element methods with the accuracy and efficiency of spectral methods. Specifically, the spectral finite element method has been used mainly in the context of structural vibrations; its early history can be found in the comprehensive review by Williams and Wittrick.[1] The present paper, however, is concerned exclusively with acoustic wave propagation and the objective here is to generate finite element trial functions defined globally over layered acoustic media domains. The approach is original in that the method is a combination of a standard finite element technique with a wave propagation analysis.

Global domain decomposition methods have appeared previously in the solution of boundary value problems described first, and applications in fluid dynamics and wave propagation described later. Basically, the idea is to divide the domain into large elements and within each constituent element represent the solution by means of a high-order global element or a spectral approximation. The main differences between the techniques outlined in the following are in the choice of basis functions and in the treatment of the continuity or compatibility at element interfaces.

A global element method, for solving elliptic PDE developed by Delves and Hall has been presented in Ref. 2. Here, the general basis functions within each element are nonconforming or discontinuous across element interfaces. Chebyshev polynomials, defined over a square region are chosen as local basis functions and a modified functional is used to implicitly enforce an interface condition between neighboring regions. The principal difficulty with the Chebyshev polynomial approach is the high dimensionality of the numerical problem which results.

Philipps and Davies[3] have also proposed a general spectral method but for Poissons equation in rectangularly decomposable regions including semi-infinite elements. Eigenfunctions of a differential operator are chosen as trial functions within each element and are nonconforming. The solution is determined by matching the solution and its derivative across element interfaces. A similar approach of using eigenfunctions of a differential operator in the trial function space is adopted in this article where the principal field equation is the Helmholtz equation. This is the basis for the spectral super-element method.

In the spectral super-element method by Schmidt *et al.*[4] a range-dependent ocean model is divided into rectangular sectors. Horizontal dependence of the solution is expressed by contributions at vertical interfaces through a boundary integral representation in cylindrical coordinates. By combining wave number integration with a boundary integral formulation a general method for various ocean environments has been developed. Their published results are for two-dimensional plane and axisymmetric environments including a penetrable half-space.

Global decomposition methods for wave propagation

[a]Electronic mail: atpeplow@kth.se
[b]Electronic mail: svantef@kth.se

problems are well known. In Evans[5] an original coupled-mode algorithm (COUPLE) was developed to handle strongly range-dependent problems in ocean acoustics. The technique consists in subdividing the environment into a number of range segments each with range-invariant material properties. The COUPLE code was generalized by Jensen and Ferla[6] to handle a multiprofile environment and provide results in a plane or cylindrical geometry. However, generation of accurate solutions with COUPLE is a nontrivial exercise. Schmidt *et al.*[4] made a comparison of the COUPLE code to handle environments with a geometric stair-step constraint to their spectral super-element code and excellent agreement was found between the two methods. In addition, the authors Goh and Schmidt[7] also extended the super-spectral method to seismoacoustic environments which included a stair-step cylindrical geometry. It is not possible to directly compare solutions given in this article with those given in Refs. 4 and 7 because of the cylindrical geometry and half-space conditions employed.

Recently, analytic approaches have also been employed to solve acoustic waveguide problems. Namely, Callan *et al.*,[8] Evans *et al.*,[9] and Evans[10] have all studied the existence of trapped modes for a number of configurations. These wave solutions are generally associated with a situation where cutoff frequencies or nonpropagating wave solutions exist. A constructive proof is given in Ref. 10 to show existence of trapped modes in a two-dimensional waveguide. Computation of a *trapped mode* waveform due to the introduction of a finite length locally reacting strip is presented in the examples section at the end of this article.

The spectral finite element method has also been applied, in various forms, to a number of acoustic and structural waveguide problems. In particular the method used in Refs. 11 and 12 can be viewed as a merger of the dynamic stiffness method and the finite element method. The principle of the method is based on a variational formulation for nonconservative motion in the frequency domain. The SFEM has also been used to study vibration in beam frameworks,[11] and for fluid-filled pipes.[12] Use of a variational formulation for the spectral method provides a natural basis for approximations and a straightforward method for combination with standard finite elements. Indeed, the motivation behind this approach is to permit direct coupling with standard finite element software. The end product allows nonstandard problems in computational wave propagation to be resolved.

In the super-spectral finite element method (SFEM), pursued here, the region of interest is decomposed into rectangular elements. A variational formulation for the underlying equations of motion is approximated using trial functions that are conforming across element boundaries. The pressure field in each element is expressed as a combination of polynomials in the transversal z direction with wave influence functions in the longitudinal x direction. From the variational formulation eigenfunctions of a derived differential operator are chosen as the wave influence trial functions in the description of the acoustic pressure field for the local *spectral element*. Local dynamic stiffness matrices for all the elements are combined to form the global stiffness matrix for the complete problem and hence given a particular source

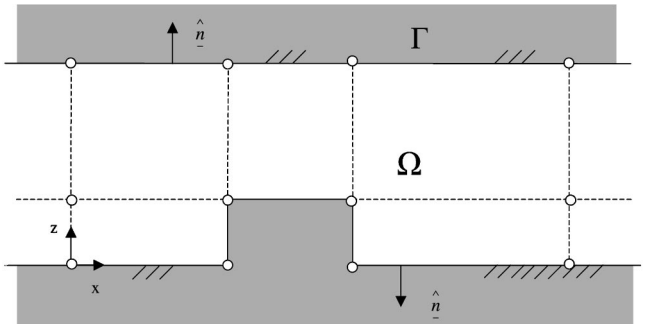


FIG. 1. Geometry of a typical waveguide.

where the pressure field may be found at any point with the fluid region. Since the spectral element does not rely on any wavelength-dependent discretization in the longitudinal direction the spectral finite element method can be applied to fairly short as well as very long wave propagation problems.

The purpose of the present study is to evaluate a computational method for predicting sound transmission through ducts or nonuniform waveguides. A Fourier series representation of the eigenfunctions is one possible method of approach for extremely simple domains, but in the present investigation a finite element method has been adopted for its versatility for more complex rectangular duct problems. A similar study is underway for nonuniform waveguides carrying elastic waves.

The article is organized as follows. In Sec. II the variational principle for nonconservative acoustic motion is stated. In Sec. III a description of the trial functions is given and associated dynamic stiffness matrices are presented. Acoustic responses for various complicated geometries are shown in Sec. IV and some conclusions are given in Sec. V.

II. ANALYSIS

The governing equations of motion for a typical acoustic rectangular duct in the $x-z$ plane lined with acoustically absorbing material is indicated in Fig. 1. It is assumed the rectangular regions of the duct as illustrated in Fig. 1 may possess different ambient fluid properties. Acoustic propagation through the inhomogeneous waveguide, Ω, with various geometries, linings, and fluid properties is then sought. Figure 2(a) shows an example of a typical geometry used in the analysis to follow:

$$\Omega=\Omega_1\cup\Omega_2\cup\ldots\cup\Omega_{J-1}\cup\Omega_J,$$

$$\Gamma=\Gamma_1\cup\Gamma_2\cup\ldots\cup\Gamma_{J-1}\cup\Gamma_J,$$

with interfaces $C_j:=\Omega_j\cap\Omega_{j+1}$, $j=1,...,J-1$. Note that on a sublocal level $\Omega_j=R_{j,1}\cup R_{j,2}\cup\ldots\cup R_{j,M}$, $m=1,...,M$, and $\Gamma_j=\Gamma_{j,1}\cup\Gamma_{j,M}$, for $j=1,...,J$. Briefly, the implementation of the numerical method is split into two stages: development of *waveguide finite elements* defined over cross sections, C_j and finally construction of *super-spectral finite elements* defined over arbitrary length regions, Ω_j.

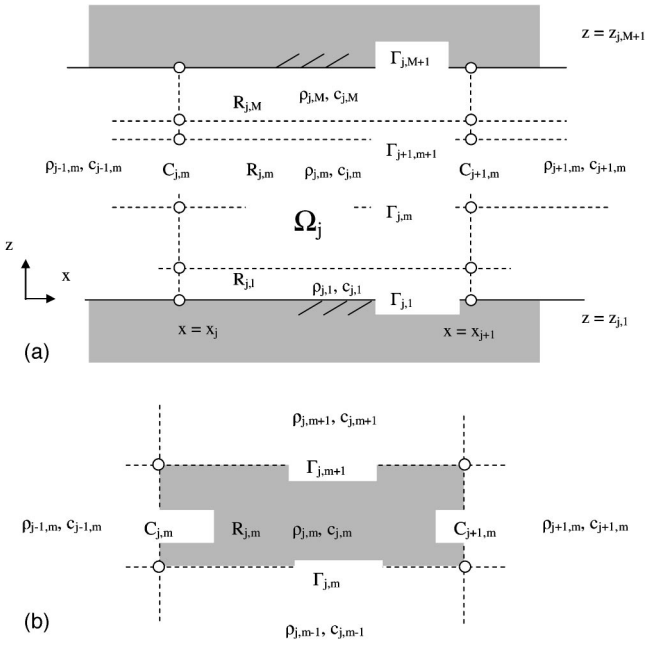


FIG. 2. (a) Geometry defining a superspectral element, (b) region at a sublocal level.

A. Governing equations

The acoustical pressure and particle displacement in a sublocal region, $R_{j,m}$ Fig. 2(a) will be denoted by $p^*(x,t)$ and $\xi^*(x,t)$, respectively. The linearized acoustical wave equation for the pressure then becomes:

$$\nabla^2 p^* = \left(\frac{1}{c_{j,m}^2}\right)\frac{\partial^2 p^*}{\partial t^2}, \quad x \in R_{j,m}, \tag{1}$$

where c is the sound speed and ∇^2 is the two-dimensional Laplacian operator, $x=(x,z)$. A relationship between particle displacement and pressure will also be required when transmission between neighboring regions and prescribed boundary conditions are considered. The linearized momentum equation provides:

$$\frac{\partial^2 \boldsymbol{\xi}^*}{\partial t^2} = -\left(\frac{1}{\rho_{j,m}}\right)\nabla p^*, \quad x \in R_{j,m}, \tag{2}$$

where ρ denotes ambient density. Note that fluid density may vary piecewise along and across the waveguide, so that ambient density is constant in each rectangular sublocal region. A time harmonic solution, of frequency $\omega = 2\pi f > 0$, is sought for an acoustic wave in the medium. Hence, the acoustical pressure and particle displacement are written in the form:

$$p^*(x,t) = p(x,z)e^{-i\omega t}, \quad \boldsymbol{\xi}^*(x,t) = \boldsymbol{\xi}(x,z)e^{-i\omega t}, \tag{3}$$

where i represents the unit imaginary number $i^2 = -1$. Equations (1) and (2) become, after introducing possible frequency dependent loss factors, in the absence of sources

$$\nabla^2 p + n_{0,(j,m)} k_{j,m}^2 p = 0, \tag{4}$$

$$\eta_{2,(j,m)} k_{j,m}^2 \boldsymbol{\xi} = \nabla p / \rho_{j,m} c_{j,m}^2, \tag{5}$$

in $R_{j,m}$ where ∇ and $\boldsymbol{\xi}$ denote two-dimensional forms, in the $x-z$ plane, of the gradient operator and particle displacement vector, respectively, and $k_{j,m} = \omega/c$ is the wave number and $n_{0,(j,m)} = \eta_{1,(j,m)}/\eta_{2,(j,m)}$. Damping factors $\eta_{1,(j,m)} = (1 - i\eta_e)^{-1}$ and $\eta_{2,(j,m)} = (1 + i\eta_v)^{-1}$ where η_e, η_v are heat and viscous loss factors, respectively, have been discussed in Ref. 12. For a sound-soft boundary the pressure-release condition

$$p \equiv 0, \quad \text{on } \Gamma_j \tag{6}$$

is adopted. For a locally reacting model the following boundary condition is used

$$\frac{\partial p}{\partial n} = \frac{i k_{j,m} \eta_{2,(j,m)}}{\zeta_j^{(U,L)}} p, \quad \text{on } \Gamma_j, \tag{7}$$

where for the specific surface impedance, $\Re e[\zeta(x)] > 0$ and $\zeta(x) = \zeta_c$, a complex constant along waveguide element boundaries with an outward pointing normal. Superscript ^{U} denotes upper face and lower face is denoted by ^{L}.

B. Variational statement

A typical waveguide airway section is shown in Fig. 2(a). We assume that the waveguide is formed by uniform constant-depth rectangular regions of acoustic fluid surrounded by a sound absorbing lining. The finite element formulation allows geometrical rectangular discontinuities such as stair steps, however, in this section we will assume that the waveguide is divided equally in the vertical direction throughout. Figure 2(b) shows a bounded rectangular region on a sublocal level, $R_{j,m}$, with neighboring regions and respective ambient densities and sound velocities $\rho_{j,m+1}$, $c_{j,m+1}, \ldots \rho_{j-1,m}$, $c_{j-1,m}$ as indicated. Let $p_{j,m+1}, \ldots, p_{j-1,m}$, and $\boldsymbol{\xi}_{j,m+1}, \ldots, \boldsymbol{\xi}_{j-1,m}$ represent corresponding pressures and displacements. In addition to the governing equations of motion, two further conditions must be satisfied by the acoustic field matching across adjacent interfaces, $C_{j,m+1}, \ldots, C_{j-1,m}$. The first is continuity of pressure across an interface

$$\begin{aligned} &p_{j,m+n}(x,z) = p_{j,m}(x,z) \quad \text{on } \Gamma_{j,m+n}, \\ &n = -1,1, \\ &p_{j,m+n}(x,z) = p_{j,m}(x,z) \quad \text{on } C_{j,m+n}. \end{aligned} \tag{8}$$

The second condition is the requirement that the normal component of displacement between regions is identical. That is, where $\hat{n}_{j,m+1} = (0,1)$, $\hat{n}_{j+1,m} = (1,0)$, $\hat{n}_{j,m-1} = (0,-1)$, $\hat{n}_{j-1,m} = (-1,0)$,

$$\hat{n}_{j+k,m+n} \cdot \boldsymbol{\xi}_{j+k,m+n} = \hat{n}_{j+k,m+n} \cdot \boldsymbol{\xi} \quad \text{on } C_{j+k,m+n}, \tag{9}$$

and using Eq. (5), condition (9) may be rewritten as

$$\begin{aligned} &\frac{1}{\eta_{2,(j+k,m+n)}} \frac{1}{\rho_{j+k,m+n}\omega^2} \nabla p_{j+k,m+n} \cdot \hat{n}_{j+k,m+n} \\ &= \frac{1}{\eta_{2,(j,m)}} \frac{1}{\rho_{j,m}\omega^2} \nabla p \cdot \hat{n}_{j+k,m+n}, \end{aligned} \tag{10}$$

where k, $n = -1$, 0, 1, and $k+n = -1$, 1.

The solution to the problem set in the original waveguide then reduces to the determination, for a given frequency of excitation and a particular source, of the solution to Eq. (4) subject to boundary condition [Eq. (7)] and constraints [Eqs. (8) and (10)]. For a given source in Ω the field equations, continuity conditions and the absorbing boundary condition may be reformulated in a single weak variational statement. That is, the problem may be expressed by the functional:

$$L_\Omega(p,p^a)=\int_\Omega\left\{\frac{\eta_1}{\rho\omega^2}k^2pp^a-\frac{\eta_2}{\rho\omega^2}\nabla p\cdot\nabla p^a\right\}d\Omega + \int_\Gamma\frac{\eta_2}{\rho\omega^2}\frac{ik}{\zeta}pp^a d\Gamma. \tag{11}$$

The superscripted symbol p^a denotes the complex conjugate of the sound pressure in the adjoint system, similar to the original system with negative damping. As shown in Ref. 11 and discussed by Finnveden,[12] the losses η_1 and η_2 may be accounted by employing a variational principle similar to Hamilton's principle whereby in the functionals describing kinetic and potential energies, the quadratic forms are replaced with symmetric bilinear forms and complex conjugates in the adjoint negatively damped system, a technique used by Gladwell.[13] With this approach the mechanisms of variational methods are retained for nonconservative systems.

Let an infinite waveguide, of uniform cross section, consist, of a sequence of rectangular elements. That is, a configuration similar to Fig. 2(a) with lower and upper boundary surface impedances ζ_j^L and ζ_j^U applied to $\Gamma_{j,1}$ and $\Gamma_{j,M+1}$, respectively, densities, wavespeeds, interfaces, damping factors, and damping factor ratios are also denoted with appropriate subscripts demonstrated in Fig. 2(a). The corresponding variational form is given by

$$L_\Omega(p,p^a)=\sum_{j=1}^{J}\left(\sum_{m=1}^{M}\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int_{R_{j,m}}(n_{0,(j,m)}k_{j,m}^2p_{j,m}p_{j,m}^a-\nabla p_{j,m}\cdot\nabla p_{j,m}^a)dR_{j,m}\right) + \sum_{j=1}^{J}\left(\frac{\eta_{2,(j,M)}}{\rho_{j,M}\omega^2}\frac{ik_{j,M}}{\zeta_j^U}\int_{\Gamma_{j,M+1}}p_{j,M}p_{j,M}^a d\Gamma_{j,M+1}+\frac{\eta_{2,(j,1)}}{\rho_{j,1}\omega^2}\frac{ik_{j,1}}{\zeta_j^L}\int_{\Gamma_{j,1}}p_{j,1}p_{j,1}^a d\Gamma_{j,1}\right). \tag{12}$$

Note that integrations are not performed over the interfaces, $\Gamma_{j,m}$ and C_i. The reason for this follows naturally from kinematic continuity outlined in the following argument.

Let $p_{j,m}$, $j=1,...,J$, $m=1,...,M$ belong to a class of continuous functions $p_{j,m}\in C(\overline{R}_{j,m})$, in each region which satisfy interface continuity, e.g., $p_{j,m}\equiv p_{j+1,m}$ on $C_{j,m}$ and $p_{j,m}\equiv p_{j,m+1}$ on $\Gamma_{j,m}$. If the functions $p_{j,m}$ are permitted to vary by quantities $\epsilon\phi_{j,m}(x,z)$ over $R_{j,m}$, so that the continuity constraint over $C_{j,m}$ and $\Gamma_{j,m}$ of acoustic pressure still holds, where $\phi_{j,m}$ are sufficiently smooth functions, the variation in L_Ω is given by

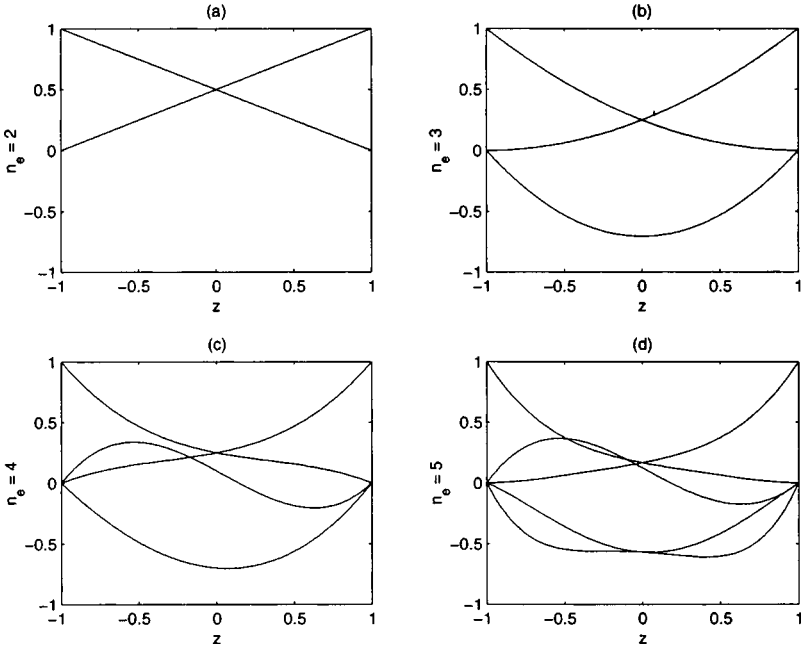


FIG. 3. Lowest four polynomials for approximation in z direction against associated number of degrees of freedom.

$$\delta L_\Omega = L_\Omega[\{p\}+\epsilon\{\phi\},\{p^a\}+\epsilon\{\phi^a\}]-L_\Omega[\{p\},\{p^a\}],$$
$$= +2\epsilon\sum_{j=1}^{J}\left\{\sum_{m=1}^{M}\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int_{R_{j,m}}(n_{0,jm}k_{j,m}^2 p_{j,m}\phi_{j,m}^a-\nabla p_{j,m}\cdot\nabla\phi_{j,m}^a)dR_{j,m}\right\}$$
$$+2\epsilon\sum_{j=1}^{J}\left\{\frac{\eta_{2,(j,M)}}{\rho_{j,M}\omega^2}\frac{ik_{j,M}}{\zeta_j^U}\int_{\Gamma_{j,M+1}}p_{j,M}\phi_{j,M}^a d\Gamma_{j,M+1}+\frac{\eta_{2,(j,1)}}{\rho_{j,1}\omega^2}\frac{ik_{j,1}}{\zeta_j^L}\int_{\Gamma_{j,1}}p_{j,1}\phi_{j,1}^a d\Gamma_{j,1}\right\}+O(\epsilon^2). \tag{13}$$

By use of the divergence therorem the integrals in the expression above become

$$\delta L_\Omega=\sum_{j=1}^{J}\left\{\sum_{m=1}^{M}\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int_{R_{j,m}}(n_{0,(j,m)}k_{j,m}^2 p_{j,m}+\nabla^2 p_{j,m})\phi_{j,m}^a dR_{j,m}\right\}$$
$$+2\epsilon\sum_{j=1}^{J}\left\{\frac{\eta_{2,(j,M)}}{\rho_{j,M}\omega^2}\int_{\Gamma_{j,M+1}}\left(\frac{ik_{j,M}}{\zeta_j^U}p_{j,M}-\nabla p_{j,M}\cdot\hat{n}_{j,M}\right)\phi_{j,M}^a dx+\frac{\eta_{2,(j,1)}}{\rho_{j,1}\omega^2}\int_{\Gamma_{j,1}}\left(\frac{ik_{j,1}}{\zeta_j^L}p_{j,1}-\nabla p_{j,1}\cdot\hat{n}_{j,1}\right)\phi_{j,1}^a dx\right\}$$
$$-2\epsilon\sum_{j=1}^{J}\sum_{m=2}^{M}\int_{\Gamma_{j,m}}\left(\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\nabla p_{j,m}\cdot\hat{n}_{j,m}\phi_{j,m}^a+\frac{\eta_{2,(j+1,m)}}{\rho_{j+1,m}\omega^2}\nabla p_{j+1,m}\cdot\hat{n}_{j+1,m}\phi_{j+1,m}^a\right)dz$$
$$-2\epsilon\sum_{j=1}^{J-1}\sum_{m=1}^{M}\int_{C_{j,m}}\left(\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\nabla p_{j,m}\cdot\hat{n}_{j,m}\phi_{j,m}^a+\frac{\eta_{2,(j,m+1)}}{\rho_{j,m+1}\omega^2}\nabla p_{j,m+1}\cdot\hat{n}_{j,m+1}\phi_{j,m+1}^a\right)dz+O(\epsilon^2). \tag{14}$$

Since $\phi_{j,m}$ are chosen arbitrarily in $R_{j,m}$ the first integral in Eq. (14) implies that if L_Ω takes a stationary value, the corresponding pressures $p_{j,m}$, must satisfy the field [Eq. (4)]. A similar argument holds for the second and third integrals in Eq. (14), that is, it implies that the boundary conditions in Eq. (7) are satisfied. The integrals over the interfaces, $C_{j,m}$ and $\Gamma_{j,m}$ in the final two terms in Eq. (14) may be rewritten as

$$\int_{\Gamma_{j,m}}\left(\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\nabla p_{j,m}\cdot\hat{n}_{j,m}\phi_{j,m}^a-\frac{\eta_{2,(j+1,m)}}{\rho_{j+1,m}\omega^2}\nabla p_{j+1,m}\cdot\hat{n}_{j,m}\phi_{j+1,m}^a\right)dz=0,\quad j=1,...,J,\quad m=2,...,M,$$
$$\int_{C_{j,m}}\left(\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\nabla p_{j,m}\cdot\hat{n}_{j,m}\phi_{j,m}^a-\frac{\eta_{2,(j,m+1)}}{\rho_{j,m+1}\omega^2}\nabla p_{j,m+1}\cdot\hat{n}_{j,m}\phi_{j,m+1}^a\right)dz=0,\quad j=1,...,J-1,\quad m=1,...,M \tag{15}$$

since $\phi_{j,m}=\phi_{j+1,m}$ on $C_{j,m}$ and $\phi_{j,m}=\phi_{j,m+1}$ on $\Gamma_{j,m}$ as explicitly stated by continuity. Now, the field equations, boundary conditions, and kinematic requirements for the original problem are satisfied by functions $p_{j,m}$ giving stationary values of the functional L_Ω Eq. (12) subject to pressure continuity over all interfaces. Due to the variational setting and the use of compact support approximating functions coupling conditions between neighboring regions are automatically satisfied. This sets the present method apart from other methods as presented in Refs. 4, 5, and 7.

Hence, Eq. (12) may now be used as a variational form for the acoustical response solution. The procedure for a finite-element idealization for the waveguide, and generation of suitable basis functions is detailed in the following section.

III. FINITE ELEMENT MODELING

The approximating functions for the super-spectral finite element method are described in detail in this section. The procedure for defining the wave influence functions is outlined and the dynamic stiffness matrices are presented.

A. Finite element interpolation

Figure 2(a) shows a typical waveguide section where $\Omega_j=R_{j,1}\cup R_{j,2}\cup\ldots\cup R_{j,M}$, $m=1,...,M$. In the finite element discretization the cross-section $C_j=[z_{j,1},z_{j,M+1}]$ is divided so that material parameters are constant within each interval. In the super-spectral method this section will define one *waveguide finite element* which sets the technique apart from standard *finite element* practise. On a sublocal level the mid-intervals $z_{j,m}\leqslant z\leqslant z_{j,m+1}$ are not subdivided as in the *h-version* finite element method, but high-order two-noded polynomials are used, i.e., the *p-version* finite element method, see Szabo and Babuska.[14] Within one *super-spectral finite element*, Ω_j for example, the approximate solution $p(\cdot)$ may be represented by an expression of the form

$$p_j(x,z)=[N_1^j(z),\ldots,N_M^j(z)][W_j(x)], \tag{16}$$

using M piecewise polynomial shape functions, compactly supported $[N_m^j(z)=0$ outside $R_{j,m}]$, in the vertical direction. Shape functions $[N_m^j(z)]$, are defined sublocally over the cross-section $C_j=[z_{j,1},z_{j,M+1}]$. The exact algebraic details of the determination of the two-noded polynomials are available elsewhere, see Finnveden.[12] Examples of shape functions chosen across an element from linear, $n_e=2$ degrees of

TABLE I. Exact and computed eigenvalues associated with wave propagation for a uniform absorbing duct; $f=200$ Hz, $h=1$ m, $\zeta^L=2+8i$, $\zeta^U\to\infty$. Increasing polynomial cross-section shape functions. An *erroneous propagating mode* wavenumber is marked with an asterisk.

$\lambda_{m,}$ (m)	Exact	$n_e=3$	$n_e=7$	$n_e=11$	$n_e=13$
1^+	$4.34+1.91E\text{-}2i$	$4.34+1.91E\text{-}2i$	$4.34+1.91E\text{-}2i$	$4.34+1.91E\text{-}2i$	$4.34+1.91E\text{-}2i$
2^+	$2.86+6.28E\text{-}2i$	$2.29+1.27E\text{-}2i$	$2.86+6.28E\text{-}2i$	$2.86+6.28E\text{-}2i$	$2.86+6.28E\text{-}2i$
3^+	$4.11E\text{-}2+4.63i$	$7.57E\text{-}2+6.66i$	$4.11E\text{-}2+4.63i$	$4.11E\text{-}2+4.63i$	$4.11E\text{-}2+4.63i$
4^+	$2.29E\text{-}2+8.42i$		$2.55E\text{-}2+8.57i$	$2.29E\text{-}2+8.42i$	$2.29E\text{-}2+8.42i$
5^+	$1.64E\text{-}2+11.8i$		$2.08E\text{-}2+12.3i$	$1.64E\text{-}2+11.8i$	$1.64E\text{-}2+11.8i$
6^+	$1.29E\text{-}2+15.1i$		$6.06E\text{-}2+27.1i$	$1.34E\text{-}2+15.2i$	$1.29E\text{-}2+15.1i$
7^+	$1.06E\text{-}2+18.4i$		$6.13E\text{-}2+36.1i$	$1.17E\text{-}2+18.5i$	$1.11E\text{-}2+18.4i$
8^+	$9.02E\text{-}3+21.6i$			$1.85E\text{-}2+24.6i$	$1.06E\text{-}2+21.9i$
9^+	$7.86E\text{-}3+24.8i$			$1.87E\text{-}2+29.3i$	$1.54E\text{-}2+27.2i$
10^+	$6.97E\text{-}3+28.0i$			$6.11E\text{-}2+70.6i$	$1.77E\text{-}2+34.3i$
11^+	$6.26E\text{-}3+31.1i$			$6.16E\text{-}2+84.6i$	$4.67E\text{-}2+60.3i$
12^+	$5.68E\text{-}3+34.3i$				$(92.7+1.17E\text{-}2i)^*$
13^+	$5.21E\text{-}3+37.5i$				$5.83E\text{-}2+99.1i$

freedom, to quartic polynomials ($n_e=5$) are illustrated in Fig. 3. Elemental shape functions $[W_j(x)]$ are defined later in this section.

It is widely accepted that the element size in element-based acoustic computations should be related to the wavelength of interest. Using the element shape functions $[N_m^j(z)]$ for a one-dimensional problem it is possible to quantify bounds for the overall error using the two-noded high order shape functions. Thompson and Pinsky[15] have also discussed this topic, investigating finite element shape functions up to polynomial degree five. To increase accuracy and reduce the effect of pollution or phase angle error the authors recommend using two quartic or quintic polynomials per wavelength. As in Ref. 14 numerical experiments proved that ill conditioning of the dynamic stiffness matrix dominated the L_2 error as the polynomial order increased beyond twelve.

B. Waveguide finite element modeling

The variational statement Eq. (12) in the preceding section is used to obtain *wave influence* basis functions $[W_j(x)]$. An approximate solution to the original problem is also found by selecting a solution $p_j(x,z)$ from a discrete set of trial functions determined by a finite element discretization of a region Ω_j Fig. 2(a).

Consider the functional L_{Ω_j}, Eq. (12), for a single arbitrary region Ω_j. Substitution of expression (16) into the resulting form, where subscript $_x$ denotes x derivative, yields the approximation

$$L_{\Omega_j}=\int \{W_j^a\}^T[K_1]\{W_j\}dx-\int \{W_j^a\}^T[K_2]\{W_j\}dx + \int \{W_j^a\}^T[K_3]\{W_j\}dx-\int \{W_j^a\}_x^T[K_4]\{W_j\}_x dx, \tag{17}$$

where, noting that the entries of matrix $[K_3]$ (of order $l_j \times l_j$, $j=1,...,M$) are explicitly stated, local submatrices are defined by

$$[K_1]_{mn}=\frac{\eta_{2,(j,m)}n_{0,(j,m)}k_{j,m}^2}{\rho_{j,m}\omega^2}\int [N_m^j N_n^j]dz, \tag{18}$$

$$[K_2]_{mn}=\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int \left[\frac{dN_m^j}{dz}\frac{dN_n^j}{dz}\right]dz, \tag{19}$$

$$[K_3^{q,r}]_{mn}=\frac{ik_{j,M}}{\zeta_j^U}\frac{\eta_{2,(j,M)}}{\rho_{j,M}\omega^2}\delta_{mM}\delta_{nM}\delta_{ql_M}\delta_{rl_M} + \frac{ik_{j,1}}{\zeta_j^L}\frac{\eta_{2,(j,1)}}{\rho_{j,1}\omega^2}\delta_{m1}\delta_{n1}\delta_{q1}\delta_{r1}, \tag{20}$$

$$[K_4]_{mn}=\frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int [N_m^j N_n^j]dz. \tag{21}$$

The final form of the matrices $[K_1],\ldots,[K_4]$ are constructed by a combination of their submatrices. Note that most *off-diagonal* submatrices are identically zero due to the compact support nature of their underlying basis functions, N_m^j and N_n^j. An example for the explicit form of the *diagonal* matrices is given in the Appendix.

C. Wave influence functions

The Euler–Lagrange equation which corresponds to Eq. (17) may be found by taking an appropriate first variation:

$$[K_4]\frac{d^2}{dx^2}\{W\}+[K_1]\{W\}-[K_2]\{W\}+[K_3]\{W\}=0. \tag{22}$$

Letting

$$[K_0]=[K_1]-[K_2]+[K_3],$$

a system of $n_e\times n_e$ second order ordinary differential equations for $[W(x)]$ follows:

$$[K_4]\frac{d^2}{dx^2}(W)+[K_0](W)=0. \tag{23}$$

The system of Eq. (23) has constant coefficients in the form of symmetric ($n_e\times n_e$) matrices $[K_4]$ and $[K_0]$. Hence, the solutions of the linear homogeneous system may be written as:

$$[W(x)]_m=c_m(\Phi)_m e^{i\lambda_m x},\quad m=1,...,n_e, \tag{24}$$

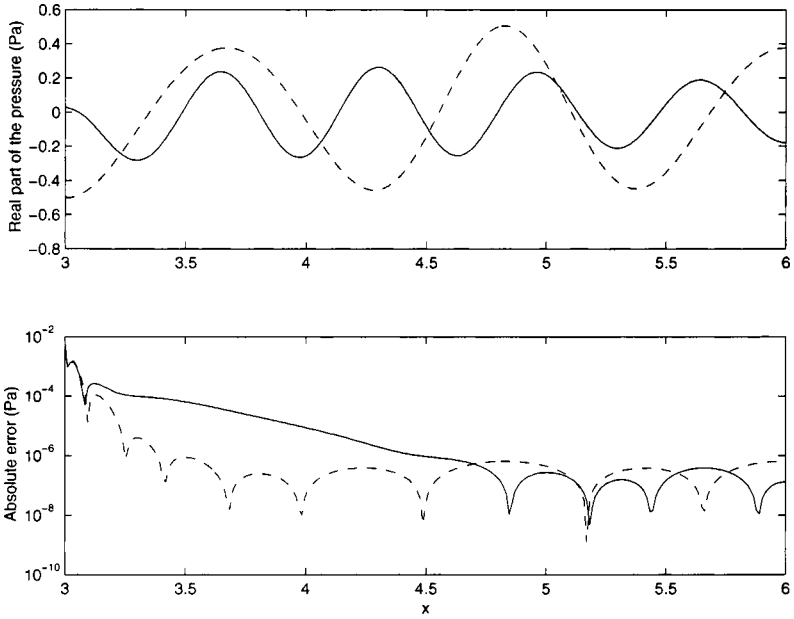


FIG. 4. Solutions and absolute errors for superspectral finite element numerical solution vs modal expansion technique via Eq. (35) for a rigid rectangular box of dimensions $(-3 \leqslant x \leqslant 3\ \mathrm{m})$ width and depth $h = 1.06\ \mathrm{m}\ (-0.53 \leqslant z \leqslant 0.53\ \mathrm{m})$. Dashed lines (--), $kh = 8.48$ and solid lines (-), $kh = 11.66$ for solutions observed along $z = 0$ m. Source strength $q = 1$. Speed of sound and density $c = 341\ \mathrm{ms}^{-1}$, $\rho = 1.21\ \mathrm{kg\,m}^{-3}$ and total number of degrees of freedom $N_e = 33$.

where $(\Phi)_m$ is a vector representing the nodal $(\Phi_{1,m}, \Phi_{n_e,m})$ and interior amplitudes $(\Phi_{j,m}, 2 \leqslant j \leqslant n_e - 1)$ and c_m is an arbitrary constant. Under this assumption Eq. (23) reduces to a nonlinear eignevalue problem, $[K(\lambda)]$, which may be cast as a linear eigenvalue problem, where

$$[K(\lambda)] = [K_0] - \lambda^2 [K_4], \tag{25}$$

of order n_e for the variable λ^2. The solution of Eq. (25) then yields eigenvalues λ which occur in pairs, $\lambda^{\pm} = \pm\lambda$, indicating that pairs of eigenmodes result with the same phase speed propagating in the positive and negative axial directions. The dimension of the eigenvalue problem is $(n_e \times n_e)$ and a finite $2n_e$ number of eigenvalues are obtained.

The resolution of the matrix eigenvalue problem Eq. (25) itself may be achieved by a number of standard computational routines. In the present analysis a QZ algorithm was used as implemented in MATLAB 6.1. This produces a complete set of cross-sectional mode shapes or waveforms and corresponding eigenvalues λ_m^2, $m = 1,...,n_e$.

Wave numbers λ_m, $m = 1,...,2n_e$, were deduced from the $(n_e \times n_e)$ generalized eigenvalue problem Eq. (25) for a 1-m-deep infinite waveguide with boundary conditions, for lower and upper surfaces, respectively, specified by the impedances $\zeta^L = 2 + 8i$ and $\zeta^U \to \infty$ (rigid) in Eq. (7) and frequency of interest $f = 200$ Hz. Exact wave numbers were found independently by a Newton–Raphson iterative procedure with initial values taken from wave numbers for a rigid waveguide. The SFEM results were computed using one set of high order polynomial basis functions in the cross-sectional direction in each case. Table I presents wave numbers λ_m^+, $m = 1,...,n_e$ corresponding to propagation in the positive x direction, see Eq. (24), to three significant figures. Although it is not shown in Table I, the second propagating mode was calculated to four decimal places for $n_e = 7$ and six decimal places with $n_e = 11$ degrees of freedom. The nonpropagating modes or evanescent waves were calculated at lower precision but it can be seen that the correspondence with exact wave numbers is close, improving as n_e increases. Note that the computed wave numbers are used within the set of basis functions described in the next section. Table I merely acts as an indicator of the calculations produced by the spectral method. An *erroneous propagating mode* wave number (marked with an asterisk) is observed for $n_e = 13$. This *erroneous* wave number, used in the set of wave influence functions, does not cause significant inaccuracy for numerical calculations but highlights the onset of numerical ill conditioning. It seems prudent, nevertheless, to take a maximum number of degrees of freedom, $n_e = 11$, for the basis functions across each individual spectral super-element.

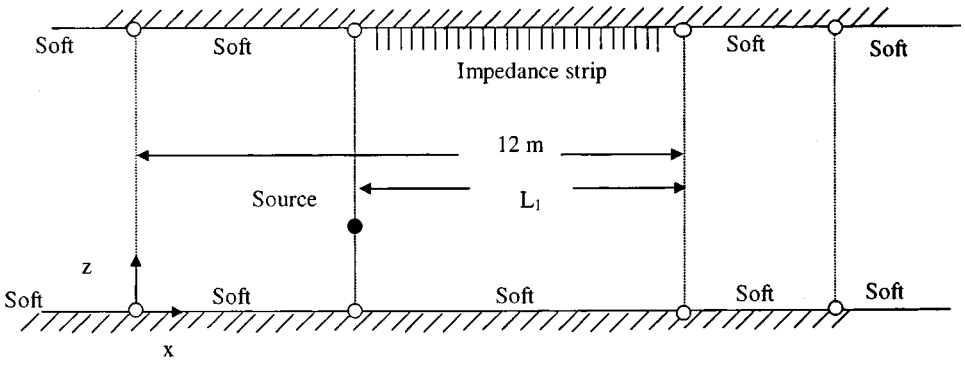


FIG. 5. Schematic for the computation of trapped modes in waveguide $-0.5 \leqslant z \leqslant 0.5$ m. Source located as illustrated with $z_0 = -0.3$ m.

D. Super-spectral element formulation

Final construction of the set of basis functions Eq. (16) and local dynamic stiffness matrix, defined over a region Ω_j as in Fig. 2(a) where $-D \leqslant x \leqslant +D$ will now be described. By consideration of the solutions in Eq. (24) it is clear that each wave influence function may be written as a linear combination of these solutions:

$$W_{jk}(x) = \sum_{l=1}^{2n_e} \overline{\Phi}_{jl} E_l(x) A_{lk} p_k, \quad j = 1,...,n_e, \quad k = 1,...,2n_e, \tag{26}$$

where entries $\overline{\Phi}_{jl}$ and E_l take the values of eigenvectors and exponential functions, respectively. Coefficients A_{lk} are determined by scaling the set of wave influence functions to unity. Now, the set of eigenvectors from Eq. (24) $\{\Phi\}_m$ is written as a partitioned matrix

$$[\overline{\Phi}] = [\Phi \quad \Phi]. \tag{27}$$

Also the vector $\{\mathbf{\Lambda}\}$, formed by the negative and positive wave number eigenvalues, is written as

$$\{\mathbf{\Lambda}\} = (-i\lambda_1,\ldots,-i\lambda_{n_e}, i\lambda_1,\ldots,i\lambda_{n_e})^T, \tag{28}$$

and introducing the new vector $\mathbf{\Lambda}^p$ defined by

$$\Lambda_m^p = \begin{cases} \Lambda_m, & \Re e(\Lambda_m) \geqslant 0, \\ -\Lambda_m, & \Re e(\Lambda_m) < 0, \end{cases} \tag{29}$$

$m = 1,...,2n_e$ the basis functions may be scaled in the longitudinal direction taking the form:

$$E_l(x) = e^{\Lambda_l x - \Lambda_l^p D}, \quad |x| \leqslant D, \quad l = 1,...,2n_e. \tag{30}$$

Projecting the basis functions, Eq. (26), onto the end points of d:

$$W_{jk}(-D) = \delta_{jk} p_k, \quad j = 1,...,n_e, \quad k = 1,...,n_e,$$

and

$$W_{(j-n_e)k}(+D) = \delta_{jk} p_k,$$

$$j = n_e + 1,...,2n_e, \quad k = n_e + 1,...,2n_e, \tag{31}$$

gives a system of $2n_e \times 2n_e$ equations determining the coefficients of the basis functions, A_{lk},

$$[X][A] = [I],$$

where

$$[X] = \begin{bmatrix} \overline{\Phi}\langle E(-D)\rangle \\ \overline{\Phi}\langle E(+D)\rangle \end{bmatrix},$$

and $[I]$ is the identity matrix and $\langle\cdot\rangle$ represents a diagonal matrix created from the original vector. Substitution of Eq. (16) into Eq. (11) over the local region Ω_j, and using the matrices as in Eqs. (18)–(21) yield a system of $2n_e \times 2n_e$ linear equations. Hence, the Lagrangian and the local dynamic stiffness matrix, $[L]_{\text{loc}}$, for a certain super-spectral element may be written as:

$$L_{\Omega_j} = \mathbf{p}^T [L]_{\text{loc}} \mathbf{p}, \tag{32}$$

where

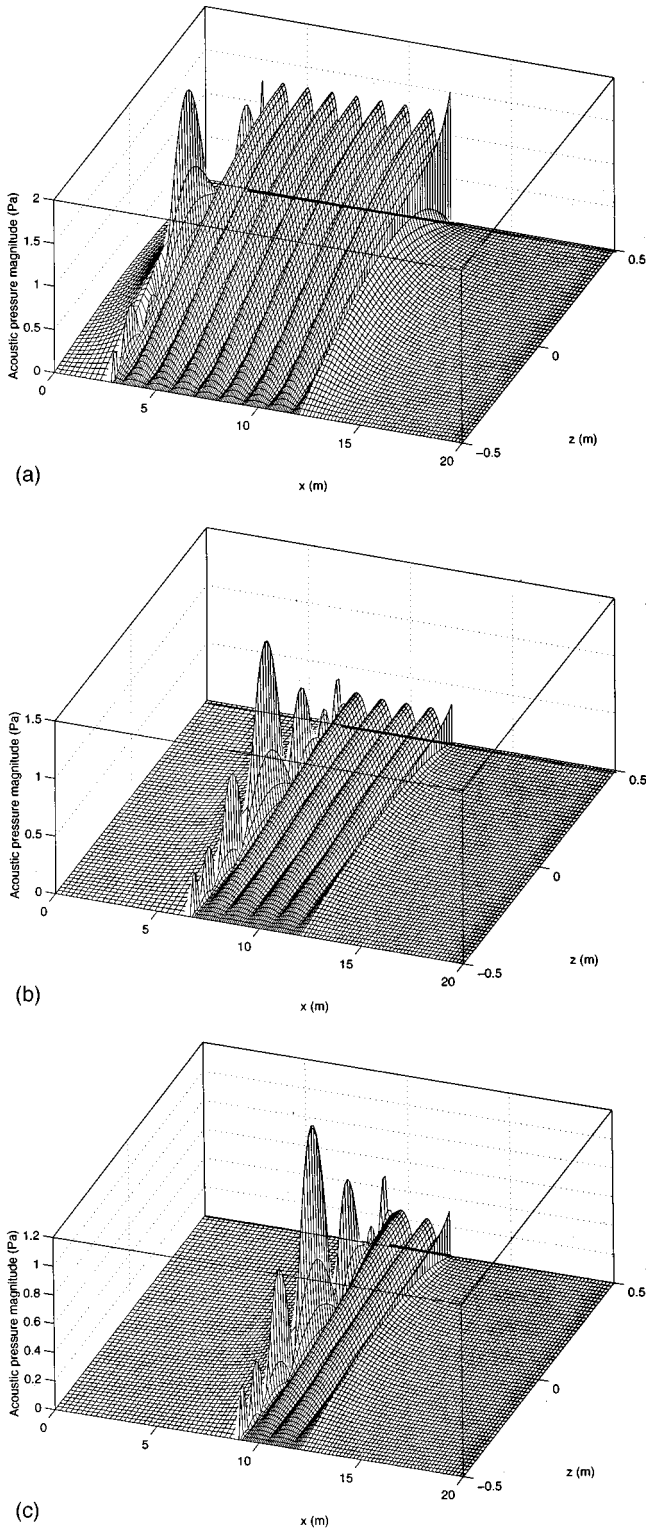


FIG. 6. Three-dimensional plots of $|p|$ at frequency $f = 166$ Hz for *infinite* sound-soft waveguide with an impedance strip $\xi^U = 10$, illustrated by solid line, at $z = 1/2$ m along lengths for (a) $2.8 \leqslant x \leqslant 12$ m. Speed of sound and density $c = 341\ \text{ms}^{-1}$, $\rho = 1.21\ \text{kg m}^{-3}$ and total number of degrees of freedom $N_e = 44$; (b) $6.6 \leqslant x \leqslant 12$ m; (c) $9 \leqslant x \leqslant 12$ m.

$$[L]_{\text{loc}} = [A]^T[([\overline{\Phi}]^T([K_1] - [K_2])[\overline{\Phi}] - \langle\mathbf{\Lambda}\rangle[\overline{\Phi}]^T[K_4] \times [\overline{\Phi}]\langle\mathbf{\Lambda}\rangle + [\overline{\Phi}]^T[K_3][\overline{\Phi}]) \cdot * [E_I]][A], \tag{33}$$

where $\cdot *$ denotes the Schur product.

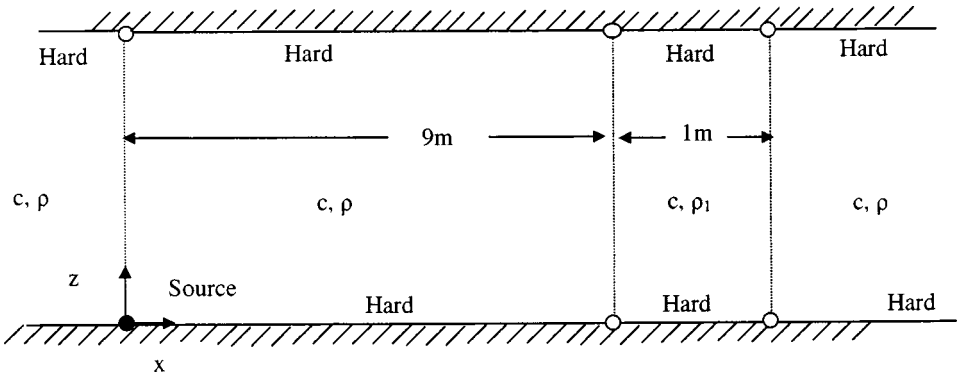


FIG. 7. Schematic for the computation of wave transmssion in an inhomogeneous waveguide $-2 \leqslant z \leqslant 2$ m; source located at $(0,-2)$.

Within the computations of local dynamic stiffness matrices use is made of a matrix generating function

$$[E_I(\{\mathbf{\Lambda}\},D)] = \int_{-D}^{D} \exp(\{\mathbf{\Lambda}\}x - \{\mathbf{\Lambda}\}^p D) \times \exp(\{\mathbf{\Lambda}\}x - \{\mathbf{\Lambda}\}^p D)^T dx, \tag{34}$$

which has an analytic form detailed in Finnveden.[12] The convention that the exp function preserves the dimension of column vectors has been adopted. Generating the E_I matrix was performed only once thus decreasing overall dynamic stiffness matrix computation substantially. For problems with *short* waveguides (small D) or *small* valued entries in $\{\mathbf{\Lambda}\}$ a Taylor expansion series for each integrated expression was performed.

The global dynamic stiffness matrix is generated by calculating local dynamic stiffness matrices for each region, Ω_j, and enforcing continuity of pressure across neighboring interfaces. For example, consider a uniform waveguide geometry consisting of three super-spectral elements, with N_e degrees of freedom across each element cross section. A $N_e \times N_e$ matrix eigenvalue problem is solved for the waveguide finite element and a $2N_e \times 2N_e$ local dynamic stiffness matrix Eq. (33) generated for each super-spectral element element. Enforcing continuity across all neighbors results in a $4N_e \times 4N_e$ global dynamic stiffness matrix. Acoustic sources are modeled as volume point sources and are introduced into the governing equations of motion Eq. (4).

Evaluation of the dynamic stiffness matrix above applies to a simple super-spectral finite element. The matrix is banded, complex and symmetric. For a combination of super-spectral finite elements, describing a general fluid domain, the corresponding dynamic stiffness matrix has a block diagonal structure; each block derived from the expression in Eq. (33). The description above and in the previous section applies to a layered waveguide where a cross section is discretized by locating appropriate interface nodes. Wave functions are subsequently found for the composed layers, by imposing pressure continuity at the interface, increasing the size of the matrix eigenvalue problem. Construction of stiffness matrices follows as with construction of stiffness matrices for one single layer waveguide.

The complete element formulation, for an arbitrary length super-spectral element, including assembling single sublocal element matrices, solving dispersion relations, and evaluating the dynamic stiffness matrix required ~0.04 s on a 1.3 GHz PC for an 11 degree-of-freedom super-spectral element.

IV. NUMERICAL EXAMPLES

In the following we illustrate how the present approach provides accurate solutions to simple canonical propagation and transmission problems. If not specified the wavespeed of air was taken to be $c = 341\ \mathrm{ms}^{-1}$ with density $\rho = 1.21\ \mathrm{kg\,m}^{-3}$ and $n_e = 11$ degrees-of-freedom across each waveguide element. Frequency values for each example were not only chosen arbitrarily but to best illustrate each case. To approximate an *infinite* waveguide extremal nodes were located at $\pm 2 \times 10^9$ m.

A. Two-dimensional rectangular domain

The spectral finite element method was tested on a two-dimensional rectangular region of width 6 m ($-3 \leqslant x \leqslant 3$ m) and depth 1.06 m ($-0.53 \leqslant z \leqslant 0.53$ m) and bounded by rigid walls. A point source was located at the center of the lower boundary. Tenth order polynomials were chosen for each of the two element sectors used in the numerical solution totalling 33 degrees-of-freedom for the model.

The solution to this problem has an analytical expression for a room of width $2D$ and depth $2h$ given by a simple modal summation:

$$P_{n,m}(x,y) = \sum_{n,m=0}^{\infty} (1+\delta_{0m})(1+\delta_{0n}) q \times \frac{\cos(\lambda_n x)\cos(\gamma_m z)\cos(\lambda_n x_0)\cos(\gamma_m z_0)}{Dh(k^2 - \lambda_n^2 - \gamma_m^2)}, \quad \lambda_n = n\pi/2D, \quad \gamma_m = m\pi/2h, \tag{35}$$

where q is the source strength.

The numerical experiment was first solved for nondimensional wave number $kh = 8.48$ and solution observed along the line $z = 0$ m and for a higher nondimensional wave number $kh = 11.66$. The numerical solutions were compared with the analytical solution termed the *exact modal solution* using $n = 2000$, $m = 90$ terms in modal expansion Eq. (35). As can be observed in Fig. 4 the results are encouraging aside from generally acceptable levels of accuracy the spectral finite element method computed the solution in a clock time of 0.02 s.

Separate computations were performed to compare the speed of the spectral finite element method and a standard modal solution. It was found that to achieve errors of order 10^{-6}, against the *exact modal solution*, a modal solution required $n=500$, $m=80$ terms. The ratio of clock times between the modal and SFEM solutions was ~7:1 for the case $kh=8.48$.

B. Computation of a trapped mode waveform

In this example an *infinite* waveguide was considered with constant cross section and a nonzero impedance condition $\zeta^U=10$ along a finite strip along the upper boundary see Fig. 5 in the vicinity of the point volume source and sound-soft boundaries elsewhere. The depth of the waveguide was taken as 1.0 m with the source excitation frequency fixed at $f=166$ Hz. For an associated nondimensional wave number $kh=3.06<\pi$ Figs. 6(a)–6(c) demonstrate the existence of *trapped mode* solutions below the cut-on frequency for various lengths of finite strip inhomogeneity. Note, that the source was located close to the absorbing boundary discontinuity in each case. It is clear that in the case where the impedance strip is longest Fig. 6(a), the trapped-mode waveform is more evident.

C. Local perturbation of ambient density

This example illustrated in Fig. 7 shows a numerical solution in an *infinite* waveguide with rigid boundary conditions, with the density of the fluid varying away from ambience in the specified region $9\leqslant x\leqslant 10$ m. Source excitation was fixed at $f=67$ Hz and located as indicated in Fig. 7.

Figure 8 shows that a density perturbation away from the density of the enclosing fluid can cause dramatic differences in wave transmission. Figure 8(a) shows wave propagation for the homogeneous case, $\rho_1=\rho=1.21\,\mathrm{kg\,m^{-3}}$, for comparative purposes. Figure 8(b) shows that a section with a significant decrease in density $\rho_1=0.021\,\mathrm{kg\,m^{-3}}$ causes a partial localization of the fluid response between the source and the inhomogeneity. However, for an increase in density Fig. 8(c) $\rho_1=10.21\,\mathrm{kg\,m^{-3}}$ near total reflection at the inhomogeneity interface is observed.

D. Local geometric perturbation

This example involves a finite stair-step located at the bottom of an otherwise constant-depth rigid waveguide. The geometry is illustrated in Fig. 9 showing a concentrated source located near the origin and frequency $f=103$ Hz. The abrupt change in the environment causes a significant amount of energy to be reflected and diffracted from the step face into the leading and shadow zones Fig. 10. This problem was solved using only seven finite element sectors, with a total of 76 degrees-of-freedom. It is important to note that for this problem, and the previous examples, it is possible to obtain computational solutions for higher nondimensional wave numbers and various density distributions and boundary conditions using the same discretization.

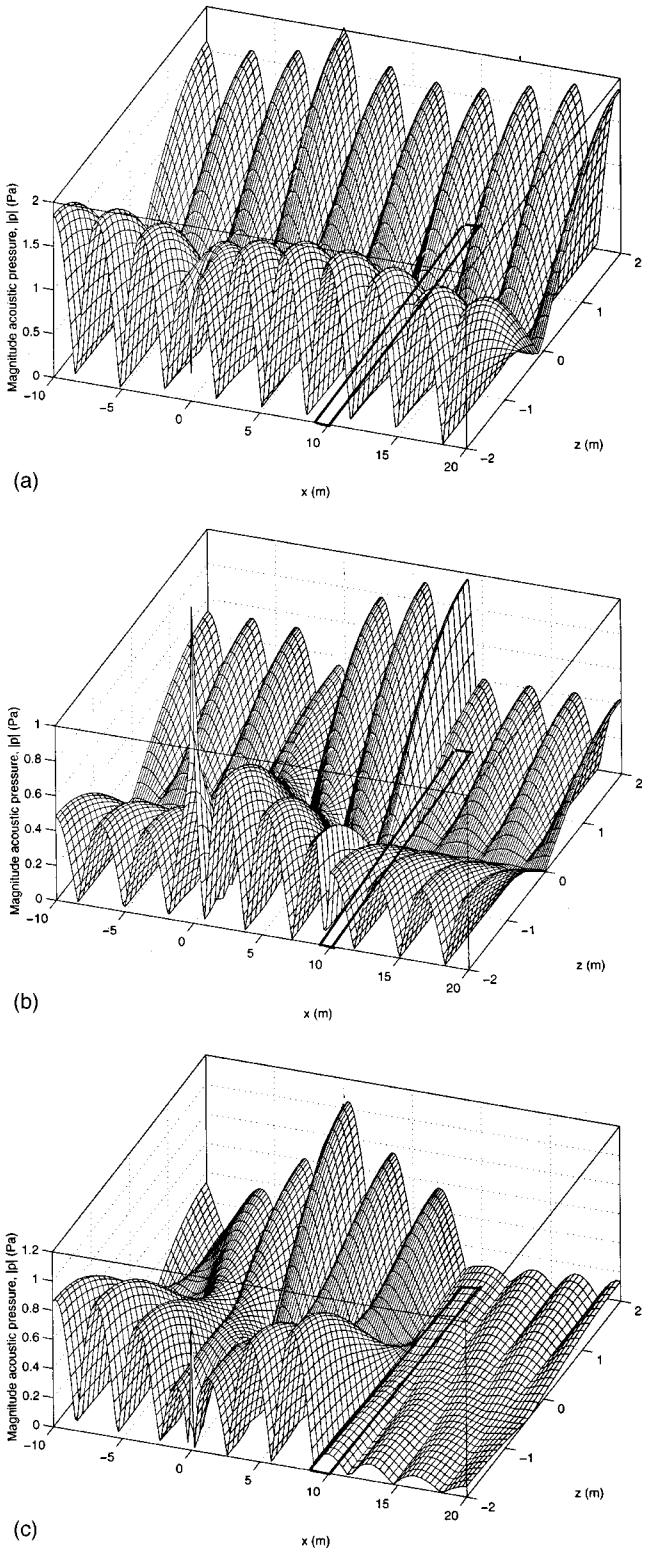


FIG. 8. Three-dimensional plots of $|p|$ at frequency $f=67$ Hz for *infinite* rigid waveguide with inhomogeneous density variation, illustrated by solid box, (a) $\rho_1=1.21\,\mathrm{kg\,m^{-3}}$. Speed of sound $c=341\,\mathrm{ms^{-1}}$ and total number of degrees of freedom $N_e=66$; (b) $\rho_1=0.021\,\mathrm{kg\,m^{-3}}$; (c) $\rho_1=10.21\,\mathrm{kg\,m^{-3}}$. Location of variation $9\leqslant x\leqslant 10$ m.

V. CONCLUSIONS

A new spectral method in the form of a finite element scheme has been used to treat the problem of sound transmission in nonuniform waveguides or ducts. A unique fea-

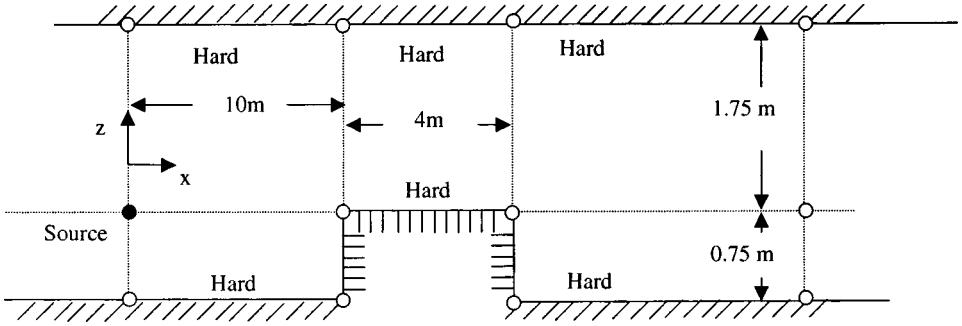


FIG. 9. Schematic for the computation of wave transmission in an stepped rigid waveguide $-1.0 \leqslant z \leqslant 1.5$ m. Source located at origin.

2ture of the super-spectral finite element approach is the use of basis functions generated from linear eigenvalue calculations. The basis functions, themselves solutions to the homogeneous reduced wave equations, may be defined over regions of arbitrary length with sound absorbing sides.

The waveguide geometries that have been given the most attention in the present study are that of a rectangular duct partially lined on one side or a geometrical nonuniform rectangular duct. These configurations are not easily amenable to analytical treatment and the use of a numerical method is appropriate here. The finite element formulation in this investigation is, however, sufficiently general that it can be extended to any nonuniform rectangular waveguide with varying linings and fluid densities, and may be coupled with a standard finite element code.

This new finite element technique extends the family of computational methods for predicting acoustic wave transmission and may possess wider applications in general wave propagation analysis for solids and in fluid–structure interaction.

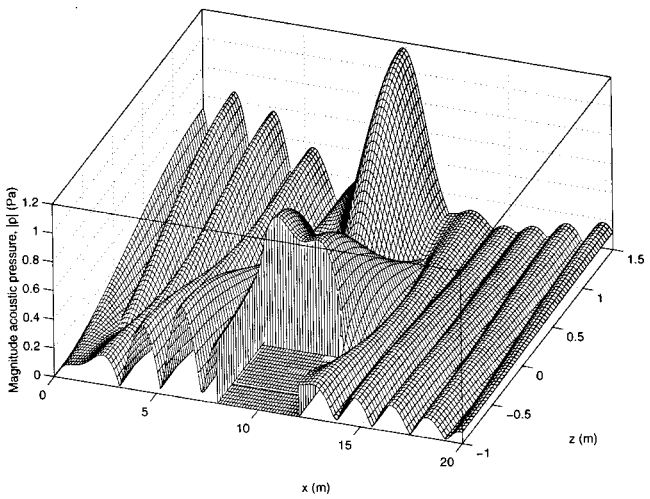


FIG. 10. Three-dimensional plot of $|p|$ at frequency $f=103$ Hz over a rigid stair-step of height 0.75 m and width 4 m in an *infinite* rigid waveguide. Speed of sound and density $c=341\ \mathrm{ms}^{-1}$, $\rho=1.21\ \mathrm{kg\,m}^{-3}$, and total number of degrees of freedom $N_e=76$.

ACKNOWLEDGMENTS

The authors wish to acknowledge support from a Marie Curie Individual Fellowship (HPMF-CT-2000-01080) and the Swedish Institute.

APPENDIX: FINITE ELEMENT MATRICES

Explicit form of the sub-local finite element matrices, derived in Section III B will be expressed. Explicit forms of the matrices $[K_1]$ and $[K_2]$ will be given. The complete set of third order polynomials is used for illustrative purposes

$$[N(\xi)]=[N^1(\xi)N^2(\xi)N^3(\xi)N^4(\xi)]$$

where

$$N^1(\xi)=\tfrac{1}{4}(2-3\xi+\xi^3),\quad N^2(\xi)=\tfrac{1}{4}(1-\xi-\xi^2+\xi^3)$$

$$N^3(\xi)=\tfrac{1}{4}(2+3\xi-\xi^3),\quad N^4(\xi)=\tfrac{1}{4}(-1-\xi+\xi^2+\xi^3).$$

Hence, the matrix $[K_1]_{mm}$ (18), defined for the interval $C_j=[a,b]$, is given by

$$[K_1]_{mm}=\frac{\eta_{2,(j,m)}n_{0,(j,m)}k_{j,m}^2}{\rho_{j,m}\omega^2}\int_a^b[N^TN]dz$$

$$=\frac{\eta_{2,(j,m)}n_{0,(j,m)}k_{j,m}^2(b-a)}{2\rho_{j,m}\omega^2}\int_{-1}^{+1}[N^T(\xi)N(\xi)]d\xi,$$

$$=\frac{\eta_{2,(j,m)}n_{0,(j,m)}k_{j,m}^2(b-a)}{210\rho_{j,m}\omega^2}$$

$$\times\begin{bmatrix}78 & 22 & 27 & -13\\ 22 & 8 & 13 & -6\\ 27 & 13 & 78 & -22\\ -13 & -6 & -22 & 8\end{bmatrix},$$

and the matrix $[K_1]_{mm}$ (19), takes the form

$$[K_2]_{mm} = \frac{\eta_{2,(j,m)}}{\rho_{j,m}\omega^2}\int_a^b \left[\frac{dN^T}{dz}\frac{dN}{dz}\right] dz$$

$$= \frac{\eta_{2,(j,m)}(b-a)}{2\rho_{j,m}\omega^2}\int_{-1}^{+1}\left[\frac{dN^T}{d\xi}\frac{dN}{d\xi}\right] d\xi,$$

$$= \frac{\eta_{2,(j,m)}(b-a)}{420\rho_{j,m}\omega^2}\begin{bmatrix} 126 & 21 & -126 & 21 \\ 21 & 56 & -21 & -14 \\ -126 & -21 & 126 & -21 \\ 21 & -14 & -21 & 8 \end{bmatrix}.$$

[1] F. W. Williams and W. H. Wittrick, "Exact buckling and frequency calculations surveyed," J. Struct. Eng. **110**, 169–187 (1983).

[2] L. M. Delves and C. A. Hall, "An implicit matching principle for global element calculations," J. Inst. Math. Appl. **23**, 223–234 (1979).

[3] T. N. Philipps and A. R. Davies, "On semi-infinite spectral elements for poisson problems with reentrant boundary singularities," J. Comput. Appl. Math. **21**, 173–188 (1988).

[4] H. Schmidt, W. Seong, and J. T. Goh, "Spectral superelement approach to range-dependent ocean acoustic modelling," J. Acoust. Soc. Am. **98**, 465–472 (1995).

[5] R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195 (1983).

[6] F. B. Jensen and C. M. Ferla, "Numerical solutions of range-dependent benchmark problems in ocean acoustics," J. Acoust. Soc. Am. **87**, 1499–1510 (1990).

[7] J. T. Goh and H. Schmidt, "A hybrid coupled wave-number integration approach to range dependent seismoacoustic modelling," J. Acoust. Soc. Am. **100**, 1409–1420 (1996).

[8] M. Callan, C. M. Linton, and D. V. Evans, "Trapped modes in two-dimensional waveguides," J. Fluid Mech. **229**, 51–64 (1991).

[9] D. V. Evans and C. M. Linton, "Trapped modes in open channels," J. Fluid Mech. **225**, 153–175 (1991).

[10] D. V. Evans, "Trapped acoustic modes," IMA J. Appl. Math. **49**, 45–60 (1992).

[11] S. Finnveden, "Exact spectral finite element analysis of stationary vibrations in a railway car structure," Acta Acust. (Beijing) **2**, 461–482 (1994).

[12] S. Finnveden, "Spectral finite element analysis of the vibration of straight fluid-filled pipes with flanges," J. Sound Vib. **199**, 125–154 (1997).

[13] G. M. L. Gladwell, "A variational formulation for damped acousto-structural problems," J. Sound Vib. **4**, 172–186 (1966).

[14] B. Szabo and I. Babuska, *Finite Element Analysis* (Wiley, New York, 1991).

[15] L. L. Thompson and P. M. Pinsky, "Complex wave number Fourier analysis of the *p*-version finite element method," Comput. Mech. **4**, 255–275 (1994).

Notes on the Gaussian beam expansion

Desheng Ding[a] and Yu Zhang
Department of Electronic Engineering, Southeast University, Nanjing 210096, People's Republic of China

(Received 22 May 2003; revised 17 April 2004; accepted 17 June 2004)

A powerful Gaussian-beam expansion technique [J. J. Wen and M. A. Breazeale, J. Acoust. Soc. Am. **83**, 1752–1756 (1988)] is widely used in calculating the Fresnel field integral. A basic ideal of this treatment is to expand the source distribution function into the superposition of a series of Gaussian functions. The evaluation of the radiation field is then greatly simplified. As a supplement, this paper presents a simple approach to determine the expansion coefficients for the sound source. The results for the ultrasonic field radiated by piston transducers are typically given, in a fairly good agreement with those calculated by numerical integration and Breazeale's. The limits of the present method are also discussed. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781619]

PACS numbers: 43.20.Rz, 43.20.Bi, 43.20.El [MO] Pages: 1401–1405

I. INTRODUCTION

In many practical situations, the sound field produced by ultrasonic transducers is reduced to the evaluation of the Fresnel field integral (or Fresnel transform), a good approximation of the Rayleigh surface integral. Generally, it is a two-dimensional and strongly oscillatory integral. When the source is distributed with the circular axial symmetry, this integral becomes a single integral, but still oscillatory. Early in 1975, Cook *et al.* first suggested the Gaussian–Laguerre (and Hermite) function method for the expansion of ultrasonic fields. The sound field is analytically described by these simple functions and the numerical evaluation of the field integral is then avoided.[1,2] Wen and Breazeale presented an alternative Gaussian expansion method in which the expansion term is purely Gaussian functions.[3] Inspired by their work, some investigators made many other extensions from various aspects, including a generalized Gaussian expansion method applied in calculation of the nonlinearly generated second-order sound field.[4–16] The nature of these approaches is based on such a fact that any *complex* source function (square integrable)[5] is aways decomposed into a certain class of *simple* functions. The word *simple* here means that the Fresnel transform (or field integral) of this class of functions may be reduced to some of the functions with mathematical simplicity. For example, the field representation of a Gaussian beam is the Gaussian function. A key problem then rises how to expand a given source function into the superposition of these simple functions. In the previous treatments, Cook *et al.* obtained the expansion coefficients of the source function by using the orthogonality relationship of the Gaussian–Laguerre functions.[1,2] For the circ function (describing the source distribution of a circular piston uniformly excited), a 65-term Gaussian–Laguerre expansion provides a satisfactory fit of it. Wen and Breazeale determined the expansion coefficients by optimization theory.[3] A better match to this function is given by only ten terms of complex Gaussian-beam functions. The results for the uniform piston field from this ten-term solution exhibit a higher precision than those of Cook *et al.* However, the expansion method by Wen and Breazeale is seemingly nonalgebraic. We ever presented a simple ''Gaussian function's polynomial'' method.[5] Thirteen terms of Gaussian functions for the expansion of the circ function were obtained. But the prediction for the uniform piston field is worse than that obtained from Breazeale's coefficients[3] and from the 65-term Gaussian–Laguerre expansion of Cook *et al.*[2]

This paper affords an improved approach to determine the expansion coefficients of the source function. The coefficients for the sources of interest in acoustics are obtained. The sound field distributions are computed and compared with the results from direct numerical integration or given by Wen and Breazeale. A good agreement is shown. It is found that the present method is very useful in the case when the distribution function is well-behaved at the edge of the source. Its limits are also discussed.

II. THEORY

A. Fresnel field integral

In the Fresnel approximation, the field radiated by an ultrasonic transducer is given by[2,5]

$$\overline{q}(\xi,\eta)=\frac{2}{i\eta}\int_0^\infty \exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right)J_0\left(\frac{2\xi\xi'}{\eta}\right)\overline{q}(\xi')\xi' d\xi' \tag{1}$$

with the dimensionless coordinates $\xi=r/a$ and $\eta=z/r_0$. Respectively, r and z are the radial and axial coordinates. Furthermore, the Rayleigh distance $r_0=ka^2/2$ is equal to π times the Fresnel distance $z_0=a^2/\lambda$, k is the wave number, λ the wavelength, and a is a characteristic radius of the source. In Eq. (1), it is assumed that the source distribution is of the circular axial symmetry. In the general case of an arbitrary function $\overline{q}(\xi')$, the field integral of Eq. (1) cannot be analytically performed, but evaluated by numerical integration technique.

B. Gaussian beam expansion

The radiation field of a Gaussian beam source with the form $e^{-B\xi^2}$ is[3,5]

[a]Electronic mail: dds@seu.edu.cn

$$G(\xi,\eta;B)=\frac{1}{1+iB\eta}\exp\left(-\frac{B\xi^2}{1+iB\eta}\right). \tag{2}$$

As stated previously, if the function $\overline{q}(\xi')$ is expressed in terms of Gaussian functions, namely,

$$\overline{q}(\xi')=\sum_{k=1}^{N} A_k e^{-B_k\xi'^2}, \tag{3}$$

the field integral of Eq. (1) is then reduced to

$$\overline{q}(\xi,\eta)=\sum_{k=1}^{N} A_k G(\xi,\eta;B_k), \tag{4}$$

i.e., calculation of the Gaussian functions. Next we shall present a simple means for determining the coefficients A_k and B_k of Eq. (3).

C. Determination of the expansion coefficients

In the fundamental paper of Wen and Breazeale,[3] the expansion and Gaussian coefficients A_k and B_k are determined by minimizing the following quantity Q, the mean square error between the original or objective function $\overline{q}(\xi)$ and Gaussian function series,

TABLE I. Expansion coefficients for a piston source supported at the edge.

k	A_k
1	$5.6474077E-1$
2	$-2.8747757E+1$
3	$4.6491142E+2$
4	$-3.4616269E+3$
5	$1.3831121E+4$
6	$-3.2052870E+4$
7	$4.4633519E+4$
8	$-3.6924454E+4$
9	$1.6753807E+4$
10	$-3.2152256E+3$

$$Q=\int_0^{\infty}\left[\overline{q}(\xi)-\sum_{k=1}^{N} A_k\exp(-B_k\xi^2)\right]^2 d\xi. \tag{5}$$

Clearly, the sum of the right-hand side in Eq. (3) approximates the original function $\overline{q}(\xi)$ in the sense of an average convergence. The coefficients A_k and B_k are finally solved based on optimization theory.[3]

It may be recognized that the method in Ref. 5 is basically equivalent to that of Wen and Breazeale. But there is an

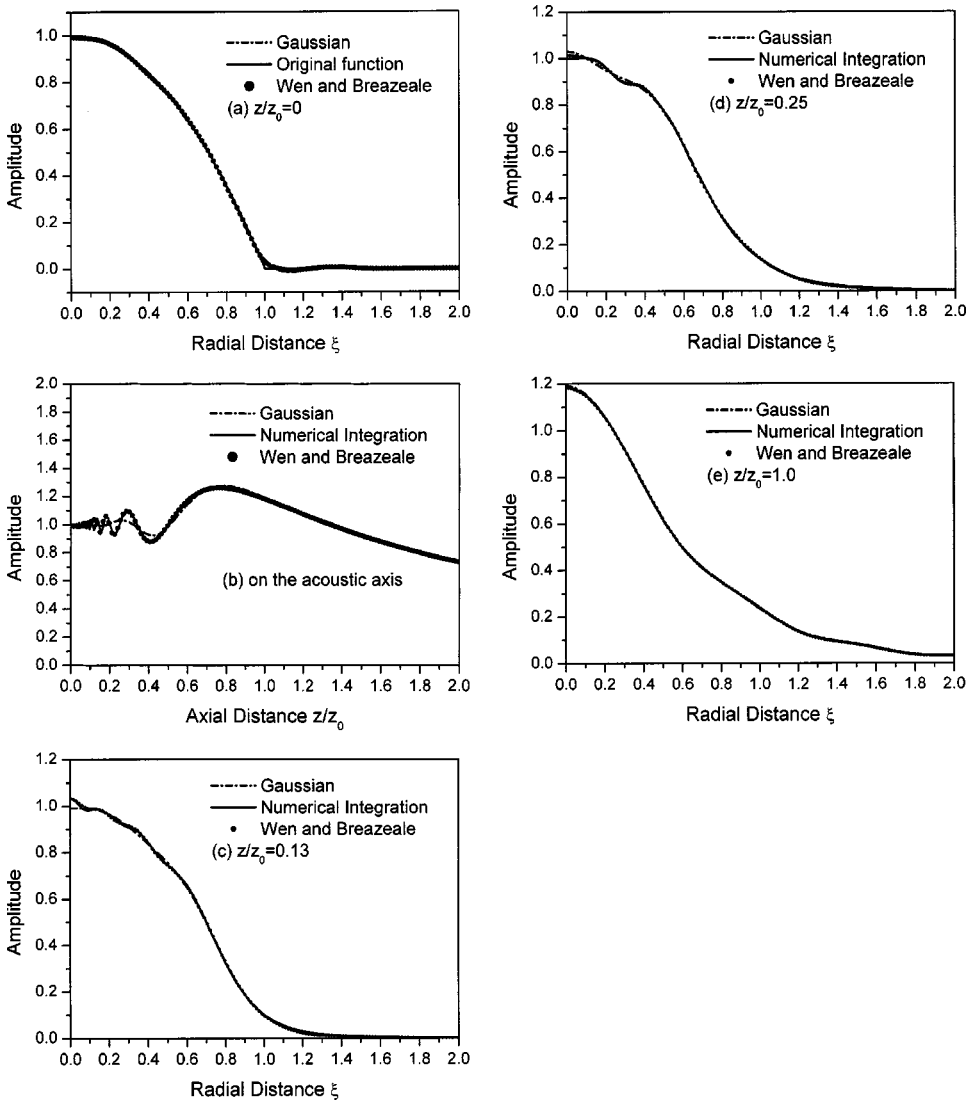


FIG. 1. Comparison of the field distributions of a piston source supported at the edge, computed by use of the present Gaussian expansion and numerical evaluation of Fresnel field integral. (a) Source function (η=0); (b) on the acoustic axis (ξ=0). And (c), (d), and (e) are at the different axial distances. Note that the axial distance is normalized to the Fresnel distance $z_0=a^2/\lambda$.

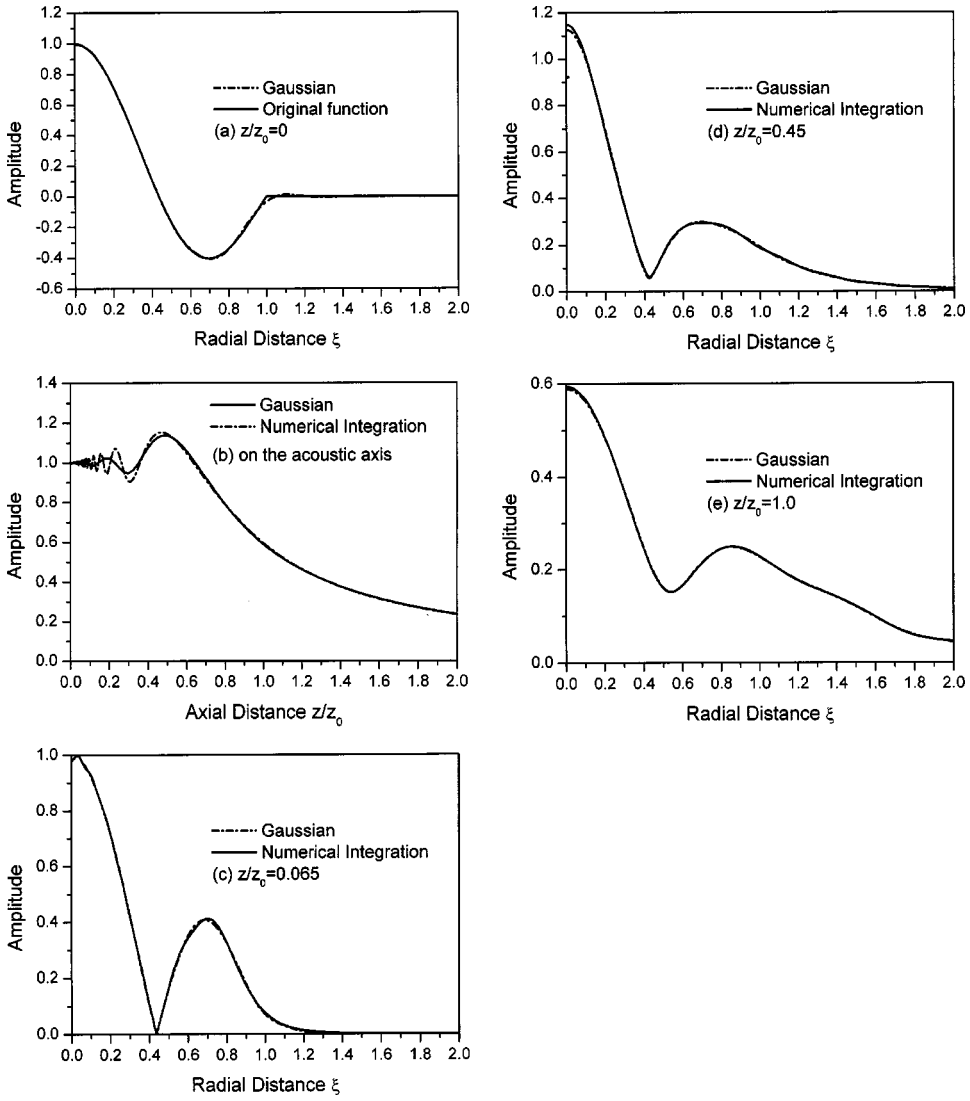


FIG. 2. Comparison of the field distributions of a truncated Bessel beam source. (a) Source function ($\eta=0$); (b) on the acoustic axis ($\xi=0$). And (c), (d), and (e) are at the different axial distances.

apparent difference that there the value of B_k is specified. It is this assumption that causes a very simple solution for the expansion coefficient A_k. Under the condition of the minimized Q, it follows

$$\frac{\partial Q}{\partial A_k}=0 \quad (k=1,2,...,N). \tag{6}$$

Therefore, the coefficient A_k can be obtained by solving a linear system of equations with respect to A_k:

$$\sum_{k=1}^{N} a_{ik}A_k=b_i \quad (i=1,2,...,N), \tag{7}$$

where

$$a_{ik}=\int_0^\infty \exp[-(B_i+B_k)\xi^2]d\xi=(\pi^{1/2}/2)(B_i+B_k)^{-1/2} \tag{8}$$

and

$$b_i=\int_0^\infty \overline{q}(\xi)\exp(-B_i\xi^2)d\xi. \tag{9}$$

D. Examples

We shall test the applicability of this method through some examples.

The first example is the sound field radiated by a piston source supported at its edge. The function of this source is defined by

$$\overline{q}(\xi)=\begin{cases}1-\xi^n, & 0\leqslant\xi\leqslant 1\\ 0, & \xi>1\end{cases} \tag{10}$$

at $n=2$. In what follows, we naturally choose $B_k=k$. Then $a_{ik}=(\pi^{1/2}/2)(i+k)^{-1/2}$ and b_i is calculated by numerical integration of Eq. (9) or in terms of the error function. The set of the expansion coefficients with ten terms is listed in Table I. Figure 1 shows the results for the original source function of Eq. (10) with $n=2$ and the radiation field fitted by these ten Gaussian functions. As above, we also apply this approach to calculate the expansion coefficients and the radiated field distributions of the other sources. One is a piston source clamped at the edge, described by Eq. (10) of $n=4$ and the other is an imagined one by Eq. (10) of $n=1$. A satisfactory result (not shown here) is again obtained and similar to these presented in Fig. 1.

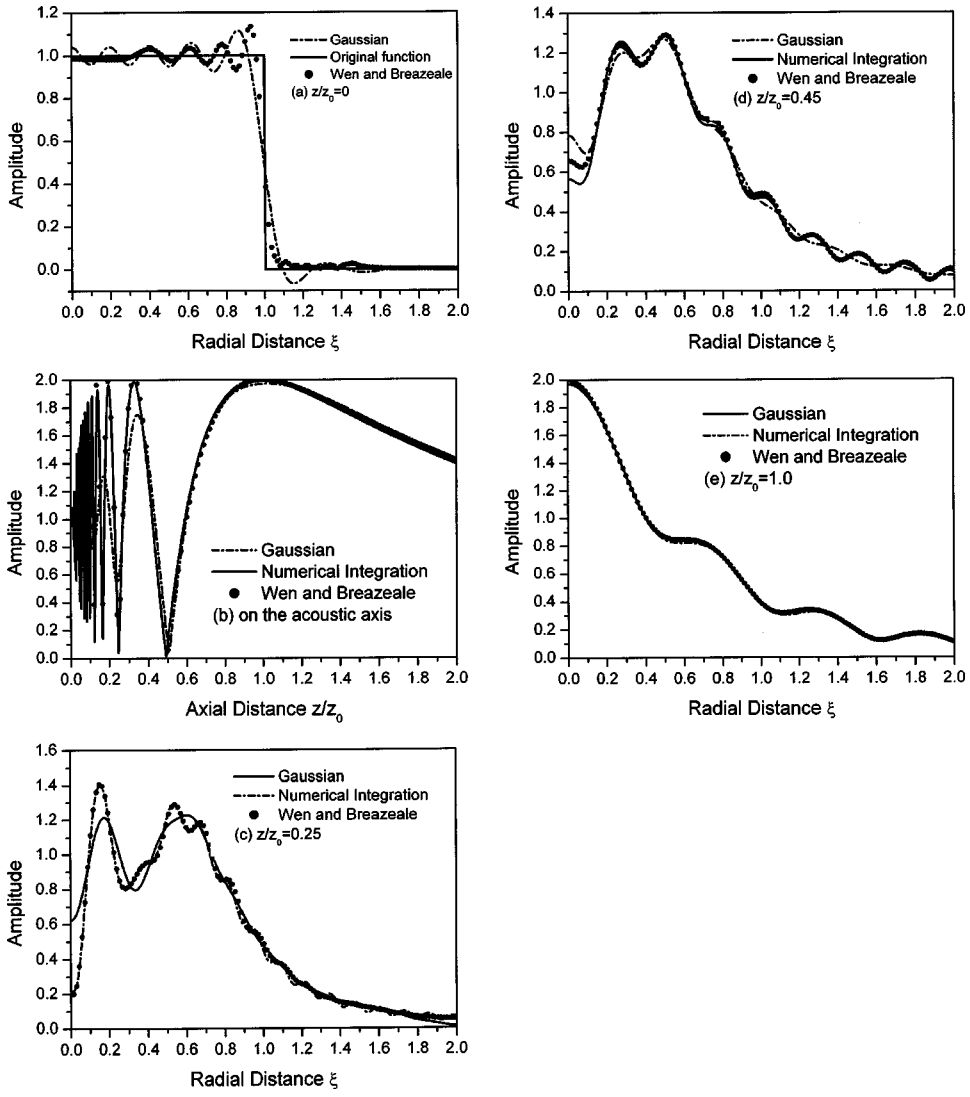


FIG. 3. Comparison of the field distributions of the uniform piston source. (a) Source function (η=0); (b) on the acoustic axis (ξ=0). And (c), (d), and (e) are at the different axial distances.

The next example is the field radiated by a truncated Bessel beam with two lobes. Its source function is given by

$$\overline{q}(\xi)=\begin{cases} J_0(\alpha\xi), & 0\leqslant\xi\leqslant 1 \\ 0, & \xi>1 \end{cases} \tag{11}$$

with α=5.5201, the second root of the Bessel function $J_0(x)$. One can see from Fig. 2 that there is a reasonably good agreement between a 15-term Gaussian function expansion and the prediction by direct numerical integration.

Finally we return to the classical case of the sound field of a uniform piston transducer. From Fig. 3, the field results calculated by a 20-term Gaussian solution are slightly improved, compared with those given by a 13-term Gaussian one based on the ''Gaussian function's polynomial'' method in Ref. 5. But within the near-field range of about 0.3 times of the Fresnel distance ($z_0=a^2/\lambda$), our results are still worse than Wen and Breazeale's.[3]

III. CONCLUDING REMARKS

From the above, we see that the simplicity is an obvious advantage of the method for determining the expansion coefficients, and their determination is solved only through the linear system of equations. However, the limits of this approach must be also noticed.

The present method is appropriate only to a limited case of the well-behaved function or source distributions. The term ''well-behaved,'' as the previous examples show, means that the source distribution does not have many points of maxima and minima, i.e., too many oscillations on the aperture, and has no sharp oscillation or jump near the edge of aperture.

The essence of this method is the same as that of Refs. 3 and 5. They are fully based on such an assumption that the source function may be expanded into a series of Gaussian functions in the sense of an average convergence. The difference is in the different procedures for determination of the coefficients A_k and B_k. In Ref. 3, the relation of these coefficients, to be all determined, is complicated and nonanalytic (algebraic) and their final determination is completed by optimization method. In our approach, the Gaussian coefficients B_k are always assumed to be known. Under this assumption, the remaining coefficients A_k are simply determined by a system of linear equations. However our simplification causes the loss of some advantages of the op-

timization theory.[3] In this frame, there is really no proper criterion to choose the Gaussian coefficient values, unlike what was done in Ref. 3. We try to change the value of Gaussian coefficients in order to get a better fit for the results from this approach and numerical integration of Eq. (1), but this effort leads to no considerable improvement.

One may see from Table I that the amplitude of expansion coefficients largely varies from the order of about $10^{-1}-10^{4}$. This implies that the more decimal numbers are needed in order to ensure the calculation precision. Presumably, the unexpected errors probably occur if the coefficient with less decimal numbers is used in the calculation of second-order fields. These problems are relevant to the convergence and stability properties of the expansion method. The convergence property has been proved that for a given (square integrable) function, averagely converging into a series of the Gaussian functions $\exp(-B_k x^2)$ in the entire interval $[0,+\infty)$, it requires the condition $\Sigma_{k=1}^{\infty} 1/B_k \to \infty$.[5,10] But this condition is of little help for the choice of coefficients in the practical case of the finite term N. A series of the Gaussian functions with form $\exp(-kx^2)$ is converging, but very slowly.

The other is the question of stability. We notice that the determinant with respect to the coefficients A_k of Eq. (7) is extremely ill-behaved. When the term or order number N becomes large, this determinant is nearly of the singularity. Hence it is difficult to obtain a highly accurate solution of the expansion coefficients. In other words, the improvement of approximation becomes immaterial by increasing the number of the expansion terms. Empirically, about twenty terms of Gaussian functions may reach a good approximation to a well-behaved function. Once the term number exceeds about 50, the error results arise. The approximate sum of the Gaussian expansion functions is strongly oscillating, and it converges no longer into the original function. It demonstrates that the present method is unstable. This instability still results from the nearly singular property of the determinant! Unfortunately, we did not find a good solution of this kind of extremely ill-behaved linear equation. We hope that our mathematicians will solve this difficult problem.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China under Grant No. 10274010.

[1] B. D. Cook and W. J. Arnoult III, "Gaussian-Laguerre/Hermite formulation for the nearfield of an ultrasonic transducer," J. Acoust. Soc. Am. **59**, 9–11 (1976).

[2] E. Cavanagh and B. D. Cook, "Gaussian-Laguerre description of ultrasonic fields-Numerical Example: Circular piston," J. Acoust. Soc. Am. **67**, 1136–1140 (1980).

[3] J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).

[4] R. B. Thompson, T. A. Gray, J. H. Rose, V. G. Kogan, and E. F. Lopes, "The radiation of elliptical and bicylindrically focused piston transducers," J. Acoust. Soc. Am. **82**, 1818–1828 (1987).

[5] D. Ding, J. Lin, Y. Shui, G. Du, and D. Zhang, "An analytical description of ultrasonic field produced by circular piston transducer," Acta Acust. (Beijing) **18**, 249–255 (1993). There a "Gaussian function's polynomial" is introduced by $G_n(x)=\Sigma_{m=1}^{n} C_{m,n} \exp(-mx^2)$, a set of orthogonal functions with the relation $\int_0^{\infty} G_n(x)G_l(x)dx=0$ $(n\neq l)$ in the interval $[0,+\infty)$. Always let $C_{1,n}=1$, and then $G_1(x)=\exp(-x^2)$. Each coefficient $C_{m,n}$ is obtained by using Schmidt orthogonalization process. For any square integrable function satisfying $\int_0^{\infty} f^2(x)dx<\infty$, it can be expanded by $f(x)=\Sigma_{n=1}^{\infty} D_n G_n(x)$ and the expansion coefficient $D_n=\int_0^{\infty} f(x)G_n(x)dx/\int_0^{\infty} G_n(x)G_n(x)dx$. A finite N-term approximation to the function is $f(x)=\Sigma_{n=1}^{N} D_n G_n(x)$, that is, $f(x)=\Sigma_{l=1}^{N} A_l \exp(-lx^2)$ with $A_l=\Sigma_{i=1}^{N} D_i C_{i,l}$. To obtain the coefficients A_l, two steps, separate determination of C_{ij} and D_i, are needed. This procedure is not so direct as that from Eq. (7) in the text, and possibly leads to the greater errors of calculation.

[6] D. Ding, Y. Shui, J. Lin, and D. Zhang, "A simple calculation approach for the second harmonic sound field generated by an arbitrary axial-symmetric source," J. Acoust. Soc. Am. **100**, 727–733 (1996).

[7] D. Ding and Z. Lu, "A simplified method to calculate the sound field of pistonlike source," Chin. J. Acoust. **15**, 213–222 (1996), also appearing in Acta Acust. (Beijing) **21**, 421–428 (1996).

[8] D. Ding and Z. Lu, "A simplified calculation for the second-order fields generated by axial-symmetric sources at bifrequency," *Nonlinear Acoustics in Perspective: Proceedings of the 14th International Symposium on Nonlinear Acoustics*, edited by R. Wei (Nanjing University Press, Nanjing, 1996), pp. 183–188.

[9] D. Ding, "A simplified algorithm for the second-order sound fields," J. Acoust. Soc. Am. **108**, 2759–2764 (2000).

[10] D. Ding and X. Liu, "Approximate description for Bessel, Bessel-Gauss and Gaussian beams with finite aperture," J. Opt. Soc. Am. A **16**, 1286–1293 (1999).

[11] M. D. Prange and R. G. Shenoy, "A fast Gaussian beam description of ultrasonic fields based on Prony's method," Ultrasonics **34**, 117–119 (1996).

[12] O. B. Matar, J. P. Rernenieras, C. Bruneel, A. Roncin, and F. Patat, "Ultrasonic sensing of vibrations," Ultrasonics **36**, 391–396 (1998).

[13] D. Huang and M. A. Breazeale, "A Gaussian finite-element method for description of sound diffraction," J. Acoust. Soc. Am. **106**, 1771–1781 (1999).

[14] M. Spies, "Transducer field modeling in anisotropic media by superposition of Gaussian base function," J. Acoust. Soc. Am. **105**, 633–638 (1999).

[15] Y. Zhang, J. Liu, and D. Ding, "Sound field calculations of elliptical pistons by the superposition of two-dimensional Gaussian beams," Chin. Phys. Lett. **19**, 1825–1827 (2002).

[16] D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gaussian beam expansion: Radiation fields of the rectangular and the elliptical transducer," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).

Acoustoelastic effects on guided waves in a fluid-filled pressurized borehole in a prestressed formation

Zhengliang Cao
School of Ocean and Earth Science, Tongji University, Shanghai, 200092, and College of Physics, Jilin University, Changchun, 130023, People's Republic of China

Kexie Wang
College of Physics, Jilin University, Changchun, 130023, People's Republic of China

Zaitian Ma
School of Ocean and Earth Science, Tongji University, Shanghai, 200092, People's Republic of China

(Received 29 October 2003; revised 4 June 2004; accepted 7 June 2004)

This paper investigates acoustoelastic effects on guided waves in a fluid-filled pressurized borehole in a biaxial prestressed formation, and develops a simplified formula that explicitly expresses the dependence of velocity dispersions on the stresses, the third-order elastic constants, and the azimuthal angle of multipole source polarization. The simplified formula shows that only in the flexural wave is there a crossover of the dispersion curves, and this phenomenon does not appear in every other guided wave. In particular, as a transformed result from the simplified expression, the square difference of flexural wave velocity dispersions is only dependent on the difference of the biaxial stresses, and neither the square sum of the flexural dispersions nor the velocity dispersions of the other guided waves is dependent on it. Computational results show that the sensitivity coefficients for the flexural dispersions induced by the biaxial stresses at low frequency asymptotically approach the corresponding coefficients of shear-waves in the formation, like the asymptotic behavior of velocities. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777857]

PACS numbers: 43.25.Ed, 43.25.Zx [MFH] Pages: 1406–1415

I. INTRODUCTION

The acoustoelastic effect, i.e., the velocity of sound in a material varies with the externally applied stress, has been investigated for more than 40 years.[1–4] The early investigations are mainly stimulated by the interest in the nondestructive evaluation of third-order elastic constants and/or residual stresses in metal samples.[5–7] In recent years, some experiments[8–10] prove that the nonlinear response of sedimentary rocks subject to applied static stresses is more obvious than that of other media. Moreover, Winkler's laboratory experiments[11,12] verify that the presence of the borehole disturbs the near hole stress field and further varies the acoustic field that is present prior to the existence of the hole. It is known that guided waves (also called normal modes or guided-wave modes)[13] excited by multipole sources (e.g., Stoneley wave by monopole or flexural wave by dipole) in a fluid-filled borehole can be used to investigate different radial depths depending on the type of transmitter and the source frequency utilized.[14] Thus the velocity dispersions of guided waves are significantly affected by the stress anisotropy near the borehole in the presence of formation stresses.[15] These phenomena spur many researchers in geophysics and seismology to study the acoustoelastic effect on guided waves as a nondestructive technique to estimate the nonlinear constants and/or the stress magnitudes in formation from acoustic logging.[16–20] The purpose of this paper is to develop the theoretical investigation of the acoustoelastic effects on guided waves based on some of the earlier results of the acoustoelasticity theory.

Acoustoelasticity is based on the hypothesis that small-disturbed fields superpose on a large nonlinear static bias, i.e., ''small-on-large'' theory, and predicts the stress dependence of acoustic velocities.[4] Acoustoelastic problems of fluid/solid systems are complicated by both the spatial inhomogeneity and the interface between the solid and the fluid. Recently, the acoustoelasticity theory is first applied by Norris and Sinha *et al.*[21] to a heterogeneous medium comprising inviscid fluid and solid constituents. They discuss the equations of motion and constitutive relations in reference, intermediate, and current coordinates, respectively, and especially develop the corresponding interface conditions including the effects of the interfacial slip. Moreover, by perturbation procedure,[22] they deduce an expression of the first-order perturbation in the eigenfrequency of Stoneley wave in a fluid-filled borehole with a pressurized fluid and prestrained solid. Sinha *et al.*[23] further studied the influence of fluid pressure in borehole on the dispersion curves of Stoneley and flexural waves through numerical calculation. In addition, Sinha *et al.*[24] calculated the change of flexural wave dispersion induced by a horizontal uniaxial stress in a formation. A significant conclusion of their works is that, because of stress concentration around a borehole, the anisotropy of formation induced by uniaxial stress will give rise to a crossover of dispersion curves corresponding to two flexural waves with polarized direction parallel and perpendicular to the applied stress direction. By the finite-difference method, Liu and Sinha[25] also proved that the crossover of dispersion curves of flexural waves can be used as an indicator of stress-induced

 © 2004 Acoustical Society of America

anisotropy distinguished from intrinsic anisotropy of formation. These theoretical predictions are subsequently verified in a scaled-borehole experiment.[15]

Since the expression of the first-order perturbation in the eigenfrequency of guided waves is expressed by volume integrals about the incremental elastic module and the displacements of unperturbed mode at a harmonic frequency [see Eq. (1) in the text],[21,23,24] it is difficult to analyze the correlation between velocity change and the azimuthal angle of multipole source polarization, and reduce the tedious calculation of sensitivity coefficients (which are defined as sensitivity of guided wave velocity to the formation stresses and the third-order elastic constants such as in Refs. 16 and 18) about different azimuthal angles.[26] On the other hand, the formation stress state is rather complicated in the vicinity of a borehole under ground. In addition to the hydrostatic pressure of the borehole fluid, there are also overburden pressure and horizontal tectonic forces around the borehole. Referring to a special stress state with a mean principle stress or the overburden pressure, the externally applied stresses may be the borehole pressurization and the biaxial stresses. So, the acoustoelastic effect on guided waves in a fluid-filled borehole induced by the complicated stresses (including the biaxial stresses and the borehole pressurization) must be further investigated. This work describes the acoustoelastic effects on guided waves in a fluid-filled pressurized borehole in a biaxial prestressed formation.

The organization of this paper is as follows: In Sec. II, we review the earlier theoretical results of acoustoelasticity about guided waves, then simplify the expression of the eigenfrequency of guided waves and deduce a simplified formula that explicitly expresses the dependence of velocity dispersions on the stresses and the azimuthal angle of multipole source polarization. In Sec. III, we numerically calculate the frequency dependence of sensitivity coefficients and velocity dispersions of Stoneley and flexural wave induced by stresses, and compare the sensitivity coefficients of flexural waves with the corresponding coefficients of shear-waves in the formation. In the end, we give a conclusion of this paper. Since the nonlinear parameter in fluid is smaller than that in rock,[11] and the acoustoelastic effect due to the borehole fluid can be separately analyzed,[23] the nonlinear effect of borehole fluid is ignored in this study.

II. FORMULATION

In acoustoelasticity theory, the kinematics of deformation of the material point can be described in three different coordinates: the reference, intermediate, and current coordinates. These coordinates correspond to the three different states of media, i.e., undeformed state, statically deformed biasing state, and elastic wave-induced deformation superposed on the static bias, respectively. Three different forms of stress tensors are defined for three states, the Cauchy stress defined in terms of current state, and the Piola–Kirchhoff stress in terms of intermediate and reference states. Three different representations of equations of motion for small dynamic fields are possible, which correspond to the three different choices of independent spatial coordinates in the reference, intermediate, or current state. In principle, one can solve wave field problems by any equations and their corresponding boundary conditions.

Generally, *in situ* stresses can be decomposed into a hydrostatical loaded state plus deviatoric stresses. The magnitude of the hydrostatic stress may be the mean principle stress or the overburden pressure at a given location. Deviatoric stresses may be the biaxial stresses and the borehole pressurization that cause stress concentration and azimuthally isotropic around the borehole, respectively. Since a hydrostatically loaded isotropic material remains isotropic, it can be served as a convenient reference state. This reference stress state can be assumed to be close to the *in situ* stress state, so that the acoustoelasticity theory can be used to solve the problem of borehole model. In this situation, the incremental static deformation of propagation medium caused by the deviatoric stresses (including the biaxial stresses and the borehole pressurization) is the static biasing state. When the biasing stresses are spatially varying in formation around the borehole, the effective elastic modules are related to the position and it is impossible to obtain a direct solution of the boundary-value problem by the equations of motion. Considering that guided waves have high magnitude in full waveform and decay away from the borehole, and applying a perturbation procedure and free boundary conditions, the first-order perturbation in the eigenfrequency ω_m can be expressed as[21,23,24]

$$\Delta\omega = \frac{\int \hat{c}_{L\gamma M\upsilon} u^m_{\upsilon,M} u^{m*}_{\gamma,L}\, dV}{2\omega_m \int \rho_0 u^m_\gamma u^{m*}_\gamma\, dV}, \tag{1}$$

where the summation convention on repeated subscripts is implied, the comma denotes partial differentiation with respect to a geometric coordinate, and the incremental elastic module in the reference state as follows,

$$\hat{c}_{L\gamma M\upsilon} = T_{LM}\delta_{\gamma\upsilon} + c_{L\gamma KM} w_{\upsilon,K} + c_{LKM\upsilon} w_{\gamma,K} + c_{L\gamma M\upsilon AB} E_{AB}, \tag{2}$$

$$T_{LM} = c_{LMRS} w_{R,S}, \tag{3}$$

$$E_{AB} = (w_{A,B} + w_{B,A})/2. \tag{4}$$

In Eqs. (1)–(4), the index m refers to the family of guided waves, for example, $m=$St and $m=$fl denote Stoneley and flexural wave, respectively. The volume V includes the fluid in borehole and the formation around it, u^m_γ is the displacement at a given wave number k_z for a selected propagating mode in reference state, * for the complex conjugate, and $c_{L\gamma M\upsilon}$ and $c_{L\gamma M\upsilon AB}$ the second- and the third-order elastic constants, respectively. Here $\delta_{\gamma\upsilon}$ is Kronecker delta, and T_{LM}, E_{AB}, and $w_{\gamma,K}$ the static biasing stresses, strains, and displacement gradients, respectively.

Without reducing generality, we assume that the formation is subjected to the borehole pressurization P_0 and biaxial stresses (S_H and S_h) along the x and y directions, respectively. The stress distribution in formation is altered by the presence of a borehole. Under the plane strain, the static biasing stresses can be obtained from the static equations of equilibrium,[27]

$$
\begin{aligned}
T_{rr} &= T_1 + T_2 \cos 2\theta, \\
T_{\theta\theta} &= T_3 + T_4 \cos 2\theta, \\
T_{r\theta} &= T_6 \sin 2\theta, \\
T_{zz} &= T_{z1} + T_{z2} \cos 2\theta, \\
T_{rz} &= 0, \quad T_{\theta z} = 0,
\end{aligned} \tag{5}
$$

where θ is the azimuthal angle relative to the x direction, and $T_i\,(i=1,2,3,4,6,z1,z2)$ is dependent on the radial location r as follows:

$$
\begin{aligned}
T_1 &= -P_0\frac{r_0^2}{r^2} + \frac{S_H+S_h}{2}\left(1-\frac{r_0^2}{r^2}\right), \\
T_2 &= \frac{S_H-S_h}{2}\left(1-\frac{4r_0^2}{r^2}+\frac{3r_0^4}{r^4}\right), \\
T_3 &= P_0\frac{r_0^2}{r^2} + \frac{S_H+S_h}{2}\left(1+\frac{r_0^2}{r^2}\right), \\
T_4 &= -\frac{S_H-S_h}{2}\left(1+\frac{3r_0^4}{r^4}\right), \\
T_6 &= -\frac{S_H-S_h}{2}\left(1+\frac{2r_0^2}{r^2}-\frac{3r_0^4}{r^4}\right), \\
T_{z1} &= \nu(S_H+S_h), \quad T_{z2} = -\nu(S_H-S_h)\frac{2r_0^2}{r^2},
\end{aligned} \tag{6}
$$

where r_0 denotes the radius of the borehole ($r \geqslant r_0$), and ν is the formation Poisson ratio. The associated biasing strains in the formation are given by

$$
\begin{aligned}
E_{rr} &= \frac{1}{2c_{66}}[(1-\nu)T_{rr} - \nu T_{\theta\theta}], \\
E_{\theta\theta} &= \frac{1}{2c_{66}}[(1-\nu)T_{\theta\theta} - \nu T_{rr}], \\
E_{r\theta} &= \frac{1}{2c_{66}}T_{r\theta}, \quad E_{zr} = E_{z\theta} = E_{zz} = 0,
\end{aligned} \tag{7}
$$

where c_{66} is the second-order elastic constant in the formation.

Note that the displacements of guided waves [see Eq. (A1) in Appendix A] exhibit the periodicity along the axial direction, moreover the incremental elastic module $\hat{c}_{L\gamma M\upsilon}$, determined by Eqs. (2)–(7), and the displacement or displacement gradients have the sine or cosine relation to the θ (see Appendix A). So that the volume integrals in Eq. (1) can be expressed in terms of surface integrals, and the surface integrals can be further separated into individual r-integrals and θ-integrals. Considering $\omega_m \Delta\omega = k_z^2 \Delta v^2$ for a given axial wave number k_z, and explicitly performing the θ-integral, Eq. (1) can be finally simplified as the following form,

$$
v^2 - v_m^2 = \mathbf{CR\Theta}. \tag{8}
$$

This simplified formula expresses the dependence of velocity dispersions of guided waves on the third-order elastic constants, the stresses, and the azimuthal angle of multipole source polarization φ relative to the x direction. Here 1×4 vector $\mathbf{C}$, 4×2 matrix $\mathbf{R}$, and 2×1 vector $\mathbf{\Theta}$ denote the matrix of the third-order elastic constants, the r-integral, and the θ-integral, respectively,

$$
\mathbf{C} = [C_{1i}] = [1, c_{155}, c_{144}, c_{123}], \quad i = 1,2,3,4, \tag{9}
$$

$$
\mathbf{R} = \frac{1}{k_z^2 D}\int_0^\infty r\,dr[I_{ij}(r)], \quad i = 1,2,3,4, \quad j = 1,2, \tag{10}
$$

$$
\mathbf{\Theta} = \begin{cases} [2\pi, 0]^T, & n = 0, \\ [\pi, \pi\cos(2\varphi)/2]^T, & n = 1, \\ [\pi, 0]^T, & n = 2,3,\ldots. \end{cases} \tag{11}
$$

In Eq. (9) $c_{155} = (c_{111} - c_{112})/4$, c_{123}, and $c_{144} = (c_{112} - c_{123})/2$ are three independent third-order elastic constants. In Eq. (10), D is the integral in the denominator of Eq. (1) and includes the effect of the borehole fluid and the solid formation referring to Eq. (A7). The elements of the matrix $I_{ij}(r)$, expressed in Eq. (A9), are related to the borehole pressurization P_0 and the biaxial stresses (S_H and S_h). In Eq. (11), because the θ-integral can be explicitly performed, $\mathbf{\Theta}$ is dependent on the azimuthal angle of multipole source polarization φ. We find from Eq. (11) that only the flexural wave excited by the dipole source ($n=1$) is dependent on φ, while the axial symmetric Stoneley wave and the other multipole waves are independent of φ. Because of the crossover of flexural dispersion curves as a result of the dependency of the source orientation, there is not any split and crossover of dispersion curves of multipole modes ($n>1$) along any two angles of wave polarization.

Note that the difference term of biaxial stresses $S_H - S_h$ does not appear in $I_{i1}(r)$, and P_0 as well as $S_H + S_h$ do not appear in $I_{i2}(r)$, so that the velocity dispersions of guided waves excited by monopole and the other multipoles ($n>1$) are dependent on both P_0 and $S_H + S_h$. But for the flexural wave, all of P_0, $S_H + S_h$, and $S_H - S_h$ affect the velocity dispersions except $\varphi = 45°$ and $\varphi = 135°$. When the radial component of the flexural wave is parallel or normal to the x-direction, i.e., $\varphi = 0°$ or $\varphi = 90°$, we obtain that $\mathbf{\Theta}_{\varphi=0°} = [\pi, \pi/2]^T$ or $\mathbf{\Theta}_{\varphi=90°} = [\pi, -\pi/2]^T$. Substituting them into Eqs. (8)–(11), we can obtain the square difference and the square sum of velocity dispersions of flexural waves parallel and normal to the x direction, respectively,

$$
v_{\varphi=0°}^2 + v_{\varphi=90°}^2 = 2v_{fl}^2 + \mathbf{CR\Theta}_+, \tag{12}
$$

$$
v_{\varphi=0°}^2 - v_{\varphi=90°}^2 = \mathbf{CR\Theta}_-, \tag{13}
$$

where v_{fl} is the flexural dispersion in the reference state calculated from the solution of linear equation of motion with a borehole boundary, and

$$
\mathbf{\Theta}_+ = [2\pi, 0]^T, \tag{14}
$$

$$
\mathbf{\Theta}_- = [0, \pi]^T. \tag{15}
$$

From Eqs. (13) and (15), we conclude that the square difference of velocity dispersions for the two orthogonally polar-

TABLE I. Material properties of the fluid and formation.

ρ (kg/m^3)	v_p (m/s)	v_s (m/s)	c_{111} (GPa)	c_{112} (GPa)	c_{123} (GPa)
2062	2320	1500	−21 217	−3044	2361

ized flexural waves is only dependent on the difference of biaxial stresses $S_H - S_h$. From Eqs. (12) and (14), the square sum is not dependent on $S_H - S_h$ but on both P_0 and $S_H + S_h$.

III. NUMERICAL RESULTS

Computational results are presented for the frequency dependence of sensitivity coefficients and velocity dispersions due to the presence of stresses. In Sec. III A, we calculate the sensitivity coefficients and velocity dispersions of Stoneley wave. In Sec. III B, we discuss the sensitivity coefficients and velocity dispersions of the two orthogonally polarized flexural waves, and further compare them with the corresponding coefficients of shear-wave velocities along the same polarization in the formation.

The results are computed for a borehole of radius 0.1016 m in a formation whose material properties are listed in Table I (referred to Ref. 24). The borehole fluid has a P-wave velocity $v_f = 1500$ m/s and a mass density $\rho_f = 1000$ kg/m^3. With no loss in generality, we select the biaxial horizontal stresses $S_H = -5$ MPa, $S_h = 0$ MPa, and the borehole pressurization $P_0 = -2$ MPa in the formation, where minus denotes pressure acting on the formation.

A. Stoneley dispersion

Since the velocity dispersion of Stoneley wave is dependent on both P_0 and $S_H + S_h$, Eq. (8) can be reduced to

$$v^2 - v_{St}^2 = Q^P P_0 + Q^S (S_H + S_h), \tag{16}$$

where v_{St} is the Stoneley dispersion in the reference state calculated from the solution of linear equation of motion with a borehole boundary, and

$$Q^P = C_0^P + C_1^P c_{155} + C_2^P c_{144} + C_3^P c_{123}, \tag{17}$$

$$Q^S = C_0^S + C_1^S c_{155} + C_2^S c_{144} + C_3^S c_{123}, \tag{18}$$

where the coefficients C_i^{P} $(i = 0,1,2,3)$ denote frequency sensitivity of Stoneley dispersion to the third-order elastic constants and the borehole pressurization alone, while the coefficients C_i^{S} $(i = 0,1,2,3)$ denote frequency sensitivity to the sum of the biaxial stresses alone. They are calculated by dot product of $\mathbf{R}$ and $\boldsymbol{\Theta}$ with unit borehole pressurization alone or with unit sum of the biaxial stresses alone. It is important to note that these sensitivity coefficients multiplied by $c_{66}/(2v_{St})$ are the same as those defined in Ref. 18 as $v \approx v_{St}$.

Figures 1 and 2 show frequency sensitivity of these coefficients to the borehole pressurization alone and to the sum of the biaxial stresses alone, i.e., C_i^{P} and C_i^{S} $(i = 0,1,2,3)$, respectively. In Fig. 1, we plot the frequency dependent of C_1^{P}, C_2^{P}, and C_3^{P} displayed by a solid line with y-axis labeling on the left and C_0^{P} displayed by the dashed line on the right. From Appendix A, because $E_1 + E_2 = 0$ with P_0 alone, C_3^{P} is zero in every frequency so that, with the borehole pressurization alone, the change of Stoneley dispersion is determined by two of the three independent third-order elastic constants, i.e., c_{155} and c_{144}. This result is consistent with results in Ref. 18. Moreover, the frequency dependencies of the sensitivity coefficients in Fig. 1 are similar to those in Fig. 5 in Ref. 18. Different from the borehole pressurization causing it to become azimuthally isotropic, the biaxial stresses induce stress concentration, so that frequency sensitivities of Stoneley dispersion to the sum of the biaxial stresses alone are more complicated as shown in Fig. 2 (its label as in Fig. 1). In this case, the sensitivity coefficient C_3^{S} gradually increases from zero with frequency. While the sensitivity coefficient C_1^{S} is significantly large at frequencies lower than 6 kHz, the coefficient C_2^{S} is significantly large at frequencies higher than 6 kHz, and C_0^{S} is significantly large in the frequency range of 1 to 5 kHz.

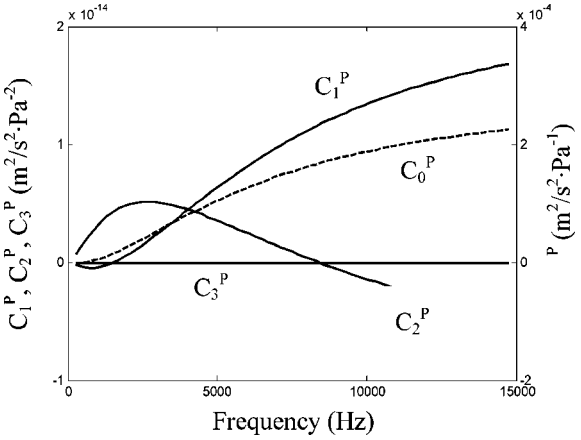


FIG. 1. Sensitivity of Stoneley dispersion to the borehole pressurization alone where the solid line and dashed line with *y*-axis labeling are on the left and the right, respectively.

According to Eq. (16) the Stoneley dispersion induced by stresses can be calculated as

$$v = \sqrt{v_{St}^2 + Q^P P_0 + Q^S (S_H + S_h)}. \tag{19}$$

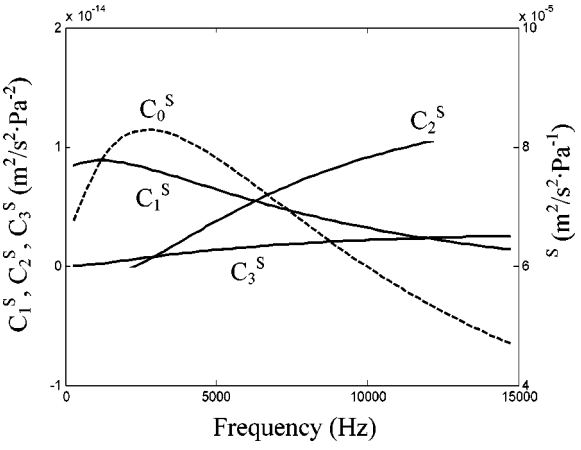


FIG. 2. Sensitivity of Stoneley dispersion to the sum of the biaxial stresses alone where the solid line and dashed line with *y*-axis labeling are on the left and the right, respectively.

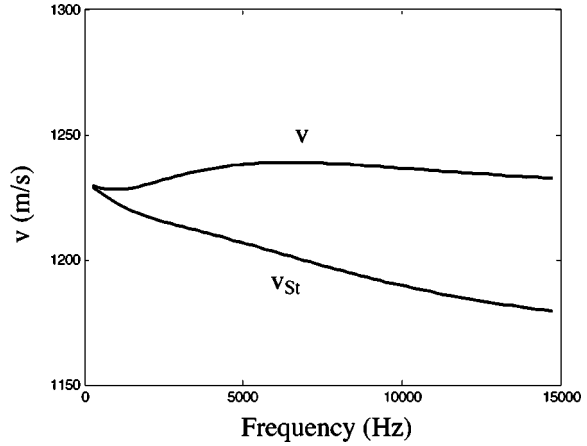


FIG. 3. Stoneley dispersions both in the reference state and the presence of the borehole pressurization alone.

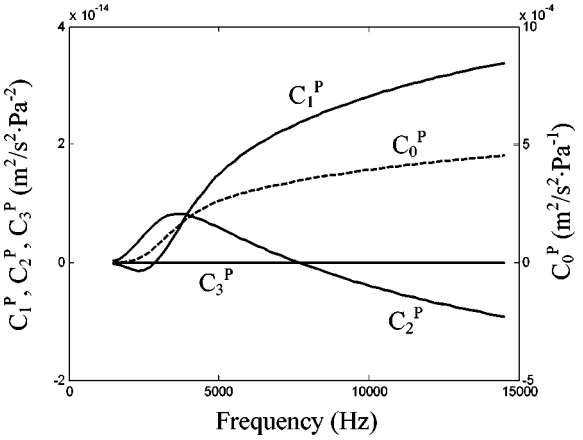


FIG. 5. Sensitivity of the square sum of flexural dispersions to the borehole pressurization alone where the solid line and dashed line with *y*-axis labeling are on the left and the right, respectively.

With $S_H+S_h=0$ or $P_0=0$, we can calculate the Stoneley wave dispersion curves with the borehole pressurization P_0 alone or the sum of the biaxial stresses S_H+S_h alone. In Fig. 3, we show the Stoneley wave velocity dispersions both in the reference state and the presence of the borehole pressurization P_0 alone. Because azimuthally isotropic stress distribution by the borehole pressurization is localized within about three times the borehole radius in the formation, the change in the Stoneley velocities at lower frequencies is rather small, and the change is larger at higher frequencies where the radial depth of investigation is rather shallow. These results are very similar to those reported earlier by Sinha *et al.* in Ref. 23. Figure 4 shows the Stoneley wave velocity dispersions both in the reference state and in the presence of the sum of the biaxial stresses S_H+S_h alone. These results are consistent with results shown in Fig. 6 in Ref. 16.

B. Flexural dispersion

The square sum of velocity dispersions for the two orthogonally polarized flexural waves is independent of S_H-S_h, thus Eqs. (12) and (14) can be reduced to the following forms,

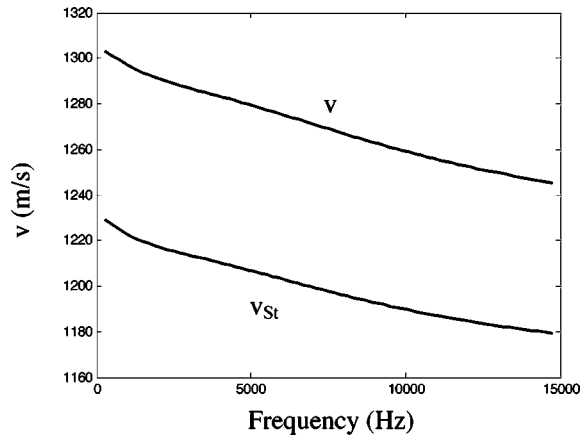


FIG. 4. Stoneley dispersions both in the reference state and the presence of the sum of the biaxial stresses alone.

$$v_{\varphi=0^\circ}^2+v_{\varphi=90^\circ}^2-2v_{fl}^2=Q^P P_0+Q^S(S_H+S_h), \tag{20}$$

$$Q^P=C_0^P+C_1^P c_{155}+C_2^P c_{144}+C_3^P c_{123}, \tag{21}$$

$$Q^S=C_0^S+C_1^S c_{155}+C_2^S c_{144}+C_3^S c_{123}. \tag{22}$$

As defined in Eqs. (17)–(19), the coefficients C_i^{P} $(i=0,1,2,3)$ denote frequency sensitivity of flexural dispersion to the third-order elastic constants and the borehole pressurization alone, while the coefficients C_i^{S} $(i=0,1,2,3)$ denote frequency sensitivity to the sum of the biaxial stresses alone. They are calculated by the same method described above.

From Appendix A, C_3^{P} is also zero in every frequency because $E_1+E_2=0$ with P_0 alone. The sensitivity coefficients C_i^{P} $(i=0,1,2,3)$ are shown in Fig. 5 (its label as in Fig. 1). In contrast to the results for Stoneley dispersion in Fig. 1, the frequency dependencies of C_i^{P} $(i=0,1,2,3)$ are similar to those of Stoneley dispersion, but different from jumping-off points of frequency and large variation levels. The result that C_3^{P} is zero for flexural dispersion induced by the borehole pressurization alone is consistent with results in Ref. 18.

From Eqs. (13) and (15), the square difference of velocity dispersions is only dependent on S_H-S_h, and we can obtain the square difference of flexural dispersions for the two orthogonally polarized directions as

$$v_{\varphi=0^\circ}^2-v_{\varphi=90^\circ}^2=Q^S(S_H-S_h), \tag{23}$$

$$Q^S=C_0^S+C_1^S c_{155}+C_2^S c_{144}+C_3^S c_{123}. \tag{24}$$

As defined in Eq. (22), the coefficients C_i^{S} $(i=0,1,2,3)$ denote frequency sensitivity of flexural dispersion to the third-order elastic constants and the difference of the biaxial stresses alone, and are calculated by dot product of $\mathbf{R}$ and $\boldsymbol{\Theta}$ with unit difference of the biaxial stresses.

Since the square difference of velocity dispersions is independent of the borehole pressurization, from Eqs. (20) and (21), we can directly calculate the flexural dispersion with the borehole pressurization alone. Figure 6 shows the flexural wave velocity dispersions both in the reference state and the presence of the borehole pressurization alone. At low frequency, flexural waves also have large radial depths of investigation, and they are insensitive to the stress distribu-

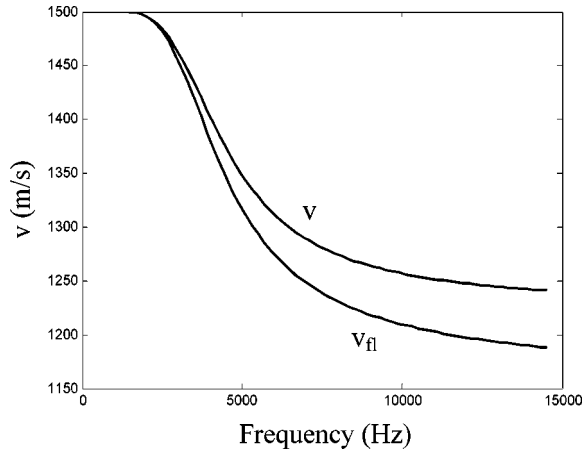


FIG. 6. Flexural dispersions both in the reference state and the presence of the borehole pressurization alone.

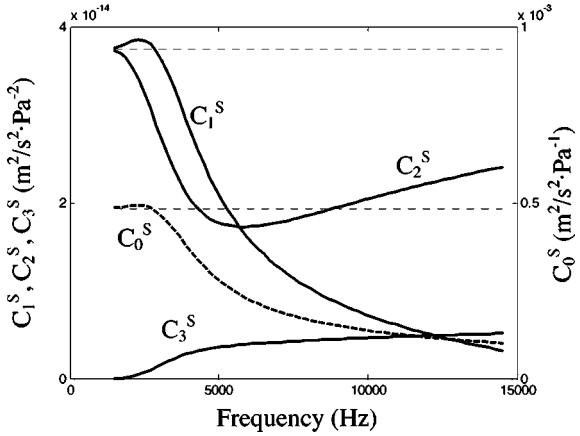


FIG. 7. Sensitivity of the square sum of flexural dispersions to the sum of the biaxial stresses alone and the corresponding coefficients of the shear-wave where the solid line and dashed line with *y*-axis labeling are on the left and the right, respectively.

tion around a pressurized borehole. However, at higher frequencies flexural waves have shallow radial depths of investigation, and they are more affected by the borehole pressurization. These results are very similar to those reported earlier by Sinha *et al.* in Ref. 23.

Referring to the analysis of shear stress parameter in Ref. 24, we can obtain the square sum and square difference of far-field shear-wave velocities with polarizations in the *x* and y directions, respectively, as follows (detailed derivation in Appendix B):

$$v_x^2+v_y^2-2v_0^2=Q^S(S_H+S_h), \tag{25}$$

$$Q^S=\frac{1}{\rho}+\frac{1-2\sigma}{2\rho c_{66}}c_{155}+\frac{1-2\sigma}{2\rho c_{66}}c_{144}, \tag{26}$$

$$v_x^2-v_y^2=Q^S(S_H-S_h), \tag{27}$$

$$Q^S=\frac{1}{\rho}+\frac{1}{2\rho c_{66}}c_{155}-\frac{1}{2\rho c_{66}}c_{144}, \tag{28}$$

where $v_0=\sqrt{c_{66}/\rho}$ is shear-wave velocity in the reference state. In Eqs. (25) and (26), the square sum is independent of S_H-S_h such as in Eq. (20), and it is also independent of the third-order elastic constant c_{123}. Because $c_{456}=(c_{155}-c_{144})/2$, Eqs. (27) and (28) can also be expressed as

$$v_x^2-v_y^2=\frac{(1+c_{456}/c_{66})}{\rho}(S_H-S_h). \tag{29}$$

This expression is the same as the expression of the formation shear stress parameter in Eq. (8) in Ref. 24.

Figure 7 shows the frequency dependence of sensitivity coefficients for the square sum of the flexural dispersions induced by the sum of the biaxial stresses S_H+S_h alone (labeling as in Fig. 1). In addition, the corresponding coefficients of the square sum of shear-wave velocities in Eq. (26) are displayed by dotted lines. The sensitivity coefficient C_0^S gradually decreases with frequency from a cutoff frequency of about 1.5 kHz. While the sensitivity coefficient C_1^S rapidly decreases from 2 kHz, C_2^S rapidly decreases to 6 kHz and then gradually increases with frequency. Note that the sensitivity coefficient C_3^S gradually increases from zero with frequency. This implies that the square sum of flexural dispersions is insensitive to the third-order elastic constant c_{123} at low frequencies. At low frequencies close to the cutoff frequency, all of the sensitivity coefficients of the flexural dispersions asymptotically approach the corresponding coefficients of shear-wave velocities.

The frequency dependence of sensitivity coefficients for the square difference of flexural dispersions induced by the difference of the biaxial stresses alone is shown in Fig. 8 (its label as in Fig. 1). Note that the sensitivity coefficient C_0^S decreases more rapidly than that of the square sum in Fig. 7. In addition, C_1^S is almost opposite to C_2^S, but the level of C_2^S is larger than that of C_1^S at higher frequencies. The sensitivity coefficient C_3^S slowly decreases from zero with frequency. This implies that the square difference of flexural dispersions is insensitive to the third-order elastic constant c_{123} at all frequencies. Similarly, all of the sensitivity coefficients of the flexural dispersions asymptotically approach the corresponding coefficients of shear-wave velocities at low frequencies.

According to Eqs. (20)–(24), we can obtain flexural dispersions $v_{\varphi=0^\circ}$ and $v_{\varphi=90^\circ}$, shown in Fig. 9. In the figure, we also display the flexural dispersion in the reference state v_{fl} and the shear-wave velocities (denoted by dotted lines)

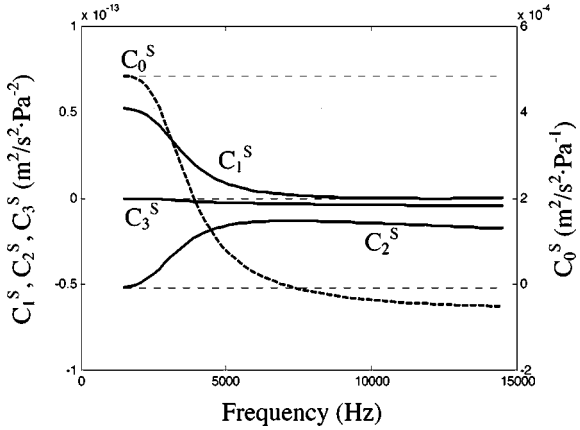


FIG. 8. Sensitivity of the square difference of flexural dispersions to the difference of the biaxial stresses alone and the corresponding coefficients of shear-wave where the solid line and dashed line with *y*-axis labeling are on the left and the right, respectively.

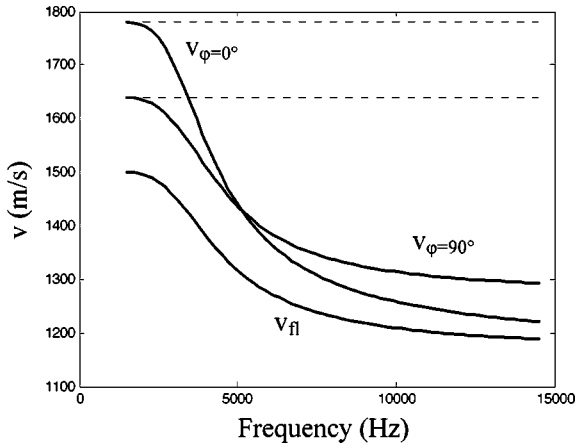


FIG. 9. Flexural dispersions in the reference state and the presence of the biaxial stresses where the subscript $\varphi=0°$ and $\varphi=90°$ denote flexural waves with radial polarization parallel and normal to the x direction, respectively.

calculated by Eqs. (25)–(28). At low frequencies, the effect of biaxial stresses on $v_{\varphi=0°}$ is larger than that on $v_{\varphi=90°}$, but it is reversed at high frequencies so that there is a crossover of dispersion curves for the two orthogonally polarized flexural waves induced by the unequal biaxial stresses. These results are consistent with results similar to those reported earlier by Sinha *et al.* in Ref. 24. In the figure, the flexural wave velocities asymptotically approach the shear-wave velocities in the formation with the same polarization at low frequencies.

IV. CONCLUSIONS

The presence of a borehole alters the existing stress state and induces complicated stress distribution in formation. As a result, the velocity dispersions of guided waves excited by multipole sources in a fluid-filled borehole are significantly affected by this stress distribution. Referring to the special reference (hydrostatically loaded) state, we discuss the acoustoelastic effect on guided waves in a fluid-filled pressurized borehole in a biaxial prestressed formation.

Based on the expression of the first-order perturbation in the eigenfrequency,[21,23,24] we develop a simplified formula that explicitly expresses the dependence of velocity dispersions of guided waves on the stresses, the third-order elastic constants, and the azimuthal angle of multipole source polarization. From the simplified formula, we find that only flexural waves excited by a dipole source are dependent on the source orientation or wave polarization, while axial symmetric Stoneley mode and the other multipole modes are independent of it, so that there are not any split and crossover of dispersion curves of multipole modes ($n>1$) along any two angles of wave polarization. In addition, according to the simplified expression of the flexural wave, the square difference of velocity dispersions for the two orthogonally polarized flexural waves is only dependent on the difference of biaxial stresses; however, the square sum of the flexural dispersions and the velocity dispersions of the other guided waves is independent of it.

On the other hand, we numerically discuss the frequency dependence of sensitivity coefficients for Stoneley and flexural waves induced by the borehole pressurization and biaxial stresses, respectively, with a set of formation parameters. Computational results show that the sensitivity coefficient C_3^S for both Stoneley and flexural wave velocity dispersions induced by the biaxial stresses varies from zero with frequency. This implies that Stoneley and flexural wave velocity dispersions are insensitive to the third-order elastic constant c_{123} at low frequencies. Further, the square difference of flexural dispersions is insensitive to the third-order elastic constant c_{123} at all frequencies, because the sensitivity coefficient C_3^S for it slowly decreases from zero with frequency. Note that C_1^S is almost opposite to C_2^S at low frequencies; the square difference of the flexural dispersions is mainly affected by the third-order elastic constant $c_{456}=(c_{155}-c_{144})/2$. In addition, such as the asymptotic behavior of velocities at low frequencies, all of the sensitivity coefficients of the flexural dispersions asymptotically approach the corresponding coefficients of shear-wave velocities in the formation with the same polarization. Understanding these will be helpful for the inversion of borehole guided wave dispersions for the nonlinear constants and/or the stress magnitudes in formation from acoustic logging.

ACKNOWLEDGMENTS

The authors are very grateful to Professor B. K. Sinha for the helpful discussions by email, and thank Professor Jingzhong Cao and Professor Hengshan Hu for revising this article. This work was supported in part by National Natural Science Foundation of China Grant No. 60303009, Grant No. 40074032, and Grant No. 4989419-04.

APPENDIX A: SIMPLIFICATION OF EQ. (1)

At a harmonic eigenfrequency frequency ω, the displacements of guided waves in the fluid and the formation expressed in cylindrical coordinate (r,θ,z) as follows,[28]

$$
\begin{aligned}
u_z^f(r,\theta,z) &= A_n[ik_z I_n(k_f r)]\cos n(\theta-\varphi)e^{ik_z z},\\
u_r^f(r,\theta,z) &= A_n\left[\frac{n}{r}I_n(k_f r)+k_f I_{n+1}(k_f r)\right]\\
&\quad\times\cos n(\theta-\varphi)e^{ik_z z},\\
u_\theta^f(r,\theta,z) &= -A_n\left[\frac{n}{r}I_n(k_f r)\right]\sin n(\theta-\varphi)e^{ik_z z},\\
u_z(r,\theta,z) &= [iB_n k_z K_n(k_p r)-C_n k_s^2 K_n(k_s r)]\\
&\quad\times\cos n(\theta-\varphi)e^{ik_z z},\\
u_r(r,\theta,z) &= \left\{B_n\left[\frac{n}{r}K_n(k_p r)-k_p K_{n+1}(k_p r)\right]\right.\\
&\quad+D_n\left[\frac{n}{r}K_n(k_s r)\right]+iC_n k_z\left[\frac{n}{r}K_n(k_s r)\right.\\
&\quad\left.\left.-k_s K_{n+1}(k_s r)\right]\right\}\cos n(\theta-\varphi)e^{ik_z z},
\end{aligned}
\tag{A1}
$$

$$u_\theta(r,\theta,z) = -\left\{ B_n \frac{n}{r} K_n(k_p r) + D_n \left[\frac{n}{r} K_n(k_s r) - k_s K_{n+1}(k_s r) \right] + iC_n k_z \frac{n}{r} K_n(k_s r) \right\} \times \sin n(\theta-\varphi) e^{ik_z z},$$

where superscript f denotes in the borehole fluid, $n = 0,1,2,...$ is the order of multipole source, and φ is the source orientation. I_n and K_n are the modified Bessel functions of the first and second kinds of order n. A_n, B_n, C_n, and D_n are the weighting coefficients. Here k_z is axial wave number, $k_f = \sqrt{k_z^2 - \omega^2/v_f^2}$, $k_p = \sqrt{k_z^2 - \omega^2/v_p^2}$, and $k_s = \sqrt{k_z^2 - \omega^2/v_s^2}$. The corresponding displacement gradients in the formation can be expressed as

$$\begin{aligned}
u_{r,r} &= V_{rr} \cos n(\theta-\varphi) e^{ik_z z}, \\
u_{r,\theta} &= V_{r\theta} \sin n(\theta-\varphi) e^{ik_z z}, \\
u_{\theta,r} &= V_{\theta r} \sin n(\theta-\varphi) e^{ik_z z}, \\
u_{r,z} &= V_{rz} \cos n(\theta-\varphi) e^{ik_z z}, \\
u_{z,r} &= V_{zr} \cos n(\theta-\varphi) e^{ik_z z}, \\
u_{z,\theta} &= V_{z\theta} \sin n(\theta-\varphi) e^{ik_z z}, \\
u_{\theta,z} &= V_{\theta z} \sin n(\theta-\varphi) e^{ik_z z}, \\
u_{\theta,\theta} &= V_{\theta\theta} \cos n(\theta-\varphi) e^{ik_z z}, \\
u_{z,z} &= V_{zz} \cos n(\theta-\varphi) e^{ik_z z},
\end{aligned} \tag{A2}$$

with

$$\begin{aligned}
V_{rr} &= B_n \left[\left(\frac{n^2-n}{r^2} + k_p^2 \right) K_n(k_p r) + \frac{k_p}{r} K_{n+1}(k_p r) \right] + D_n \left[\frac{n^2-n}{r^2} K_n(k_s r) - \frac{n}{r} k_s K_{n+1}(k_s r) \right] + iC_n k_z \left[\frac{n}{r} K_n(k_s r) - k_s K_{n+1}(k_s r) \right], \\
V_{r\theta} &= B_n \left[\frac{n-n^2}{r^2} K_n(k_p r) + \frac{n}{r} k_p K_{n+1}(k_p r) \right] + D_n \left[\frac{n-n^2}{r^2} K_n(k_s r) - \frac{k_s}{r} K_{n+1}(k_s r) \right] + C_n ik_z \left[\frac{n-n^2}{r^2} K_n(k_s r) + \frac{n}{r} k_s K_{n+1}(k_s r) \right], \\
V_{\theta r} &= B_n \left[\frac{n-n^2}{r^2} K_n(k_p r) + \frac{n}{r} k_p K_{n+1}(k_p r) \right] + D_n \left[\left(\frac{n-n^2}{r^2} - k_s^2 \right) K_n(k_s r) - \frac{k_s}{r} K_{n+1}(k_s r) \right] + C_n ik_z \left[\frac{n-n^2}{r^2} K_n(k_s r) + \frac{n}{r} k_s K_{n+1}(k_s r) \right], \\
V_{rz} &= B_n ik_z \left[\frac{n}{r} K_n(k_p r) - k_p K_{n+1}(k_p r) \right] + D_n \frac{ik_z n}{r} \times K_n(k_s r) - C_n k_z^2 \left[\frac{n}{r} K_n(k_s r) - k_s K_{n+1}(k_s r) \right], \\
V_{zr} &= B_n ik_z \left[\frac{n}{r} K_n(k_p r) - k_p K_{n+1}(k_p r) \right] + C_n k_s^2 \left[-\frac{n}{r} K_n(k_s r) + k_s K_{n+1}(k_s r) \right], \\
V_{z\theta} &= -B_n ik_z \frac{n}{r} K_n(k_p r) + C_n k_s^2 \frac{n}{r} K_n(k_s r), \\
V_{\theta z} &= B_n ik_z \left[-\frac{n}{r} K_n(k_p r) \right] + D_n ik_z \left[-\frac{n}{r} K_n(k_s r) + k_s K_{n+1}(k_s r) \right] - C_n k_z^2 \left[-\frac{n}{r} K_n(k_s r) \right], \\
V_{\theta\theta} &= B_n \left[\frac{n-n^2}{r^2} K_n(k_p r) - \frac{k_p}{r} K_{n+1}(k_p r) \right] + D_n \left[\frac{n-n^2}{r^2} K_n(k_s r) + \frac{n}{r} k_s K_{n+1}(k_s r) \right] + C_n ik_z \left[\frac{n-n^2}{r^2} K_n(k_s r) - \frac{k_s}{r} K_{n+1}(k_s r) \right], \\
V_{zz} &= -B_n k_z^2 K_n(k_p r) - C_n ik_z k_s^2 K_n(k_s r).
\end{aligned} \tag{A3}$$

Insertion of Eqs. (5) into Eq. (7) then yields the static biasing strains as

$$\begin{aligned}
E_{rr} &= E_1 + E_2 \cos 2\theta, \\
E_{\theta\theta} &= E_3 + E_4 \cos 2\theta, \\
E_{r\theta} &= E_6 \sin 2\theta, \\
E_{zr} &= E_{z\theta} = E_{zz} = 0,
\end{aligned} \tag{A4}$$

where

$$\begin{aligned}
E_1 &= \frac{1}{2c_{66}} [(1-\nu)T_1 - \nu T_3], \\
E_2 &= \frac{1}{2c_{66}} [(1-\nu)T_2 - \nu T_4], \\
E_3 &= \frac{1}{2c_{66}} [(1-\nu)T_3 - \nu T_1], \\
E_4 &= \frac{1}{2c_{66}} [(1-\nu)T_4 - \nu T_2].
\end{aligned} \tag{A5}$$

Under the plane strain assumption, the radial-azimuthal shear displacement gradients are

$$W_{r\theta}=\left[E_6+\frac{(k+1)}{4}(E2+E_4)\right]\sin 2\theta,$$
$$W_{\theta r}=\left[E_6-\frac{(k+1)}{4}(E2+E_4)\right]\sin 2\theta, \tag{A6}$$

where $k=3-4\nu$, and $E_6=(1/2c_{66})T_6$. $T_i\,(i=1,2,3,4,6)$ are denoted in Eq. (6).

Note that the displacement gradients exhibit the periodicity along the axial direction in Eq. (A2), and the terms of θ in Eqs. (A3)–(A6) are sine or cosine formats. Substituting Eqs. (A1)–(A6) into (1), and according to individual term about the third-order elastic constants, the stresses, and the azimuthal angle of multipole source polarization, respectively, we can reduce Eq. (1) to Eqs. (8)–(11), where

$$D=Q_1\left[\rho_f\int_0^{r_0}(\hat{u}_r^f\hat{u}_r^{f*}+\hat{u}_z^f\hat{u}_z^{f*})r\,dr+\rho\int_0^{\infty}(\hat{u}_r\hat{u}_r^*+\hat{u}_z\hat{u}_z^*)r\,dr\right]+Q_2\left[\rho_f\int_0^{r_0}\hat{u}_\theta^f\hat{u}_\theta^{f*}r\,dr+\rho\int_0^{\infty}\hat{u}_\theta\hat{u}_\theta^*r\,dr\right], \tag{A7}$$

with $\hat{u}_i^f=u_i^f(r)$, $\hat{u}_i=u_i(r)$, $i=r,\theta,z$, and

$$Q_1=\int_0^{2\pi}\cos^2 n(\theta-\varphi)d\theta,$$
$$Q_2=\int_0^{2\pi}\sin^2 n(\theta-\varphi)d\theta. \tag{A8}$$

The elements of the 4×2 matrix, $I_{ij}(r)$, are denoted as follows:

$$\begin{aligned}
I_{11}&=T_{z1}(R_{zzzz}+R_{rzrz}+R_{\theta z\theta z})+T_1(R_{zrzr}+R_{rrrr}+R_{\theta r\theta r})\\
&\quad+T_3(R_{\theta\theta\theta\theta}+R_{z\theta z\theta}+R_{r\theta r\theta})+2E_1[\lambda(R_{rrzz}+R_{rr\theta\theta})\\
&\quad+(\lambda+2\mu)R_{rrrr}+\mu(R_{rzzr}+R_{rzrz}+R_{r\theta\theta r}+R_{r\theta r\theta})]\\
&\quad+2E_3[\lambda(R_{\theta\theta zz}+R_{\theta\theta rr})+(\lambda+2\mu)R_{\theta\theta\theta\theta}\\
&\quad+\mu(R_{z\theta\theta z}+R_{\theta z\theta z}+R_{\theta r\theta r}+R_{r\theta\theta r})],\\
I_{12}&=T_{z2}(R_{zzzz}+R_{rzrz}-R_{\theta z\theta z})+T_2(R_{zrzr}+R_{rrrr}-R_{\theta r\theta r})\\
&\quad+T_4(R_{\theta\theta\theta\theta}-R_{z\theta z\theta}-R_{r\theta r\theta})+2E_2[\lambda(R_{rrzz}+R_{rr\theta\theta})\\
&\quad+(\lambda+2\mu)R_{rrrr}+\mu(R_{rzzr}+R_{rzrz}-R_{r\theta\theta r}-R_{r\theta r\theta})]\\
&\quad+2E_4[\lambda(R_{\theta\theta zz}+R_{\theta\theta rr})+(\lambda+2\mu)R_{\theta\theta\theta\theta}-\mu(R_{z\theta\theta z}\\
&\quad+R_{\theta z\theta z}+R_{\theta r\theta r}+R_{r\theta\theta r})]+2T_6(R_{r\theta rr}+R_{\theta r\theta\theta}+R_{z\theta zr})\\
&\quad+2W_{r\theta}[\mu(R_{rzz\theta}+R_{\theta zrz}+R_{rr\theta r})+\lambda R_{r\theta zz}\\
&\quad+(\lambda+\mu)R_{rrr\theta}+(\lambda+2\mu)R_{r\theta\theta\theta}]\\
&\quad+2W_{\theta r}[\mu(R_{rz\theta z}+R_{\theta zzr}+R_{r\theta\theta\theta})+\lambda R_{\theta rzz}\\
&\quad+(\lambda+\mu)R_{\theta\theta\theta r}+(\lambda+2\mu)R_{rr\theta r}],\\
I_{21}&=4[E_1(\tilde{R}_{rrrr}+\tilde{R}_{rzrz})+E_3(\tilde{R}_{\theta\theta\theta\theta}+\tilde{R}_{\theta z\theta z})\\
&\quad+(E_1+E_3)\tilde{R}_{r\theta r\theta}],\\
I_{22}&=4[E_2(\tilde{R}_{rrrr}+\tilde{R}_{rzrz})+E_4(\tilde{R}_{\theta\theta\theta\theta}-\tilde{R}_{\theta z\theta z})\\
&\quad-(E_2+E_4)\tilde{R}_{r\theta r\theta}+2E_6(\tilde{R}_{r\theta rr}+\tilde{R}_{\theta zrz}+\tilde{R}_{r\theta\theta\theta})],\\
I_{31}&=2[(E_1+E_3)(\tilde{R}_{rrrr}+\tilde{R}_{zzzz}+\tilde{R}_{\theta\theta\theta\theta}+2\tilde{R}_{rr\theta\theta})\\
&\quad+2E_1(\tilde{R}_{rrzz}+\tilde{R}_{\theta z\theta z})+2E_3(\tilde{R}_{\theta\theta zz}+\tilde{R}_{rzrz})],\\
I_{32}&=2[(E_2+E_4)(\tilde{R}_{rrrr}+\tilde{R}_{\theta\theta\theta\theta}+\tilde{R}_{zzzz}+2\tilde{R}_{rr\theta\theta})\\
&\quad+2E_2(\tilde{R}_{rrzz}-\tilde{R}_{\theta z\theta z})+2E_4(\tilde{R}_{\theta\theta zz}+\tilde{R}_{rzrz})\\
&\quad+4E_6(\tilde{R}_{r\theta zz}-\tilde{R}_{\theta zrz})],\\
I_{41}&=(E_1+E_3)[\tilde{R}_{rrrr}+\tilde{R}_{\theta\theta\theta\theta}+\tilde{R}_{zzzz}\\
&\quad+2(\tilde{R}_{rr\theta\theta}+\tilde{R}_{\theta\theta zz}+\tilde{R}_{rrzz})],\\
I_{42}&=(E_2+E_4)[\tilde{R}_{rrrr}+\tilde{R}_{\theta\theta\theta\theta}+\tilde{R}_{zzzz}\\
&\quad+2(\tilde{R}_{rr\theta\theta}+\tilde{R}_{\theta\theta zz}+\tilde{R}_{rrzz})],
\end{aligned} \tag{A9}$$

where

$$R_{ijkl}=R_{klij}=\mathrm{Real}(V_{ij}V_{kl}^*),$$
$$\tilde{R}_{ijkl}=\tilde{R}_{klij}=\mathrm{Real}(\epsilon_{ij}\epsilon_{kl}^*), \tag{A10}$$

and

$$\epsilon_{ij}=\tfrac{1}{2}(V_{ij}+V_{ji}),\quad i,j,k,l=r,\theta,z.$$

APPENDIX B: DERIVATION OF EQS. (25)–(28)

The plane wave velocities for waves propagating along the z direction in an isotropic medium subject to homogeneous biasing normal stresses and strains are given by[27]

$$\rho v_x^2=c_{66}+T_{zz}+(2c_{66}+c_{155})(E_{zz}+E_{xx})+c_{144}E_{yy}, \tag{B1}$$

$$\rho v_y^2=c_{66}+T_{zz}+(2c_{66}+c_{155})(E_{zz}+E_{yy})+c_{144}E_{xx}, \tag{B2}$$

where v_x and v_y is shear wave velocity along the x and y directions, respectively. Under plane strain assumption, the biasing stresses and strains induced by biaxial stresses, S_H and S_h, are expressed in Cartesian coordinates as

$$E_{xx}=\frac{1}{2c_{66}}[(1-\nu)S_H-\nu S_h],$$
$$E_{yy}=\frac{1}{2c_{66}}[-\nu S_H+(1-\nu)S_h], \tag{B3}$$
$$T_{zz}=\nu(S_H+S_h),\quad E_{zz}=0.$$

Substituting Eq. (B3) into Eqs. (B1) and (B2), and adding or subtracting the two equations, we obtain the square sum and square difference of shear-wave velocities in the formation with the orthogonal polarizations, as follows,

$$v_x^2+v_y^2-2v_0^2=\left(\frac{1}{\rho}+\frac{1-2\sigma}{2\rho c_{66}}c_{155}+\frac{1-2\sigma}{2\rho c_{66}}c_{144}\right)\times(S_H+S_h), \tag{B4}$$

$$v_x^2 - v_y^2 = \left(\frac{1}{\rho} + \frac{1}{2\rho c_{66}} c_{155} - \frac{1}{2\rho c_{66}} c_{144} \right) (S_H - S_h), \quad \text{(B5)}$$

where $v_0 = \sqrt{c_{66}/\rho}$ is shear wave velocity in the reference state.

[1] D. S. Hughes and J. L. Kelly, ''Second-order elastic deformation of solids,'' Phys. Rev. **92**, 1145–1149 (1953).
[2] R. A. Toupin and B. Bernstein, ''Sound wave in deformed perfectly elastic materials. Acoustoelastic effect,'' J. Acoust. Soc. Am. **32**, 216–225 (1961).
[3] R. N. Thurston and K. Brugger, ''Third-order elastic constants and the velocity of small amplitude elastic waves in homogeneously stressed media,'' Phys. Rev. **133**, A1604–A1610 (1964).
[4] A. N. Norris, ''Finite-amplitude waves in solid,'' in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), pp. 263–276.
[5] D. M. Egle and D. E. Bray, ''Measurement of acoustoelastic and third-order elastic constants for rail steel,'' J. Acoust. Soc. Am. **60**, 741–744 (1976).
[6] R. N. Thurston, ''Waves in solids,'' in *Mechanics of Solids*, edited by C. Truesdell (Spinger-Verlag, Berlin, 1984), Vol. 4, pp. 109–308.
[7] Y.-H. Pao, W. Sachse, and H. Fukuoka, ''Acoustoelasticity and ultrasonic measurements of residual stresses,'' in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1984), Vol. XVII, pp. 61–143.
[8] P. A. Johnson, G. D. Meegan, K. McCall, B. P. Bonner, and T. J. Shankland, ''Finite amplitude wave studies in earth materials,'' J. Acoust. Soc. Am. **91**, 2350 (1992).
[9] G. D. Meegan, P. A. Johnson, R. A. Guyer, and K. R. McCall, ''Observation of nonlinear elastic wave behavior in sandstone,'' J. Acoust. Soc. Am. **94**, 3387–3391 (1993).
[10] K. W. Winkler and L. Xingzhou, ''Measurements of third-order elastic constants in rocks,'' J. Acoust. Soc. Am. **100**, 1392–1398 (1996).
[11] K. W. Winkler, ''Azimuthal velocity variations by borehole stress concentrations,'' J. Geophys. Res. **101**, 8615–8621 (1996).
[12] K. W. Winkler, ''Acoustic evidence of mechanical damage surrounding stressed borehole,'' Geophysics **62**, 16–22 (1997).
[13] F. L. Paillet and J. E. White, ''Acoustic modes of propagation in the borehole and their relationship to rock properties,'' Geophysics **47**, 1215–1228 (1982).
[14] A. L. Kurkjian and S. K. Chang, ''Acoustic multipole sources in fluid-filled boreholes,'' Geophysics **51**, 148–163 (1986).
[15] K. W. Winkler, B. K. Sinha, and T. J. Plona, ''Effects of borehole stress concentrations on dipole anisotropy measurements,'' Geophysics **63**, 11–17 (1998).
[16] B. K. Sinha, ''Inversion of borehole dispersion for formation stresses,'' Proc. 1997 IEEE International Ultrasonic Symposium, 5–8 October, 1997, IEEE Catalog No. 97CH36118, pp. 781–786.
[17] T. J. Plona, B. K. Sinha, M. R. Kane, K. W. Winkler, and B. Frignet, ''Stress-induced dipole anisotropy: Theory, experiment and field data,'' SPWLA 40th Annual Logging Symposium, paper RR (1999).
[18] B. K. Sinha and K. W. Winkler, ''Formation nonlinear constants from sonic measurements at two borehole pressures,'' Geophysics **64**, 1890–1900 (1999).
[19] X. Huang, B. K. Sinha, M. N. Toksöz, and D. R. Burns, ''Formation stress estimation using standard acoustic logging,'' 74th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstract, 53–56 (1999).
[20] B. K. Sinha, M. R. Kane, and B. Frignet, ''Dipole dispersion crossover and sonic logs in a limestone reservoir,'' Geophysics **65**, 390–407 (2000).
[21] A. N. Norris, B. K. Sinha, and S. Kostek, ''Acoustoelasticity of solid/fluid composite systems,'' Geophys. J. Int. **118**, 439–446 (1994).
[22] B. K. Sinha, ''Elastic waves in crystals under a bias,'' Ferroelectrics **41**, 61–73 (1982).
[23] B. K. Sinha, S. Kostek, and A. N. Norris, ''Stoneley and flexural modes in pressurized borehole,'' J. Geophys. Res. **100**, 22375–22381 (1995).
[24] B. K. Sinha and S. Kostek, ''Stress-induced azimuthal anisotropy in borehole flexural waves,'' Geophysics **61**, 1899–1907 (1996).
[25] Q. H. Liu and B. K. Sinha, ''Multipole acoustic waveforms in fluid-filled boreholes in biaxially stressed formations: a finite-difference method,'' Geophysics **65**, 190–201 (2000).
[26] Z. Cao, K. Wang, G. Li, R. Xie, J. Liu, and X. Lv, ''Dipole flexural waves splitting induced by borehole pressurization and formation stress concentration,'' Chin. J. Geophys. **46**, 1021–1030 (2003).
[27] H. Reismann and P. S. Pawlik, *Elasticity, Theory and Applications* (Wiley, New York, 1980), Chap. 5, pp. 185–198.
[28] Z. Cao, ''Forward and Inverse Study on Estimation of Abnormal Formation Stress from Cross-dipole Acoustic Logging,'' Ph.D. thesis, Jilin University, 2002.

Leading edge noise from thick foils in turbulent flows

Jonathan Gershfeld[a)]
Carderock Division, Naval Surface Warfare Center, Code 725, W. Bethesda, Maryland 20817-5700

(Received 20 April 2004; accepted for publication 20 June 2004)

The prediction of dipole sound from the diffraction of turbulence by the leading edge of a rigid foil is made using acoustic analogies for homogeneous flows with and without mean shear. The Green's function accounts for the foil thickness and leading edge shape. A comparison between the theory and published measurements suggests that the foil thickness exponentially attenuates the dipole sound pressure power spectrum by the product of the convection wavenumber and half the maximum section thickness. The ratio of the predicted dipole sound from the mean shear source to the source without mean shear is determined to be proportional to the ratio of the mean shear to the frequency of the sources. [DOI: 10.1121/1.1780575]

PACS numbers: 43.28.Ra [MSH] Pages: 1416–1426

I. INTRODUCTION

An airfoil in a turbulent flow generates a net distribution of acoustic dipoles primarily at the leading and trailing edges. In these regions, the dipole surface normal stresses induced by the turbulence are not canceled by their images due to the surface curvature (Powell, 1960; Meecham, 1965). Sears (1941) modeled the incompressible unsteady lifting force associated with the dipoles by representing the foil as an acoustically compact two-dimensional strip. His analysis included a trailing edge Kutta condition and showed for fluctuating velocities much smaller than the free stream velocity that the unsteady lifting force was proportional to the product of the mean free stream velocity and the fluctuating ''upwash'' velocity. The application of the Kutta condition essentially removed this edge's contribution to the incompressible part of the total unsteady lift for reduced frequencies greater than one (based on the chord length) (Howe, 2001b).

When the airfoil chord length is acoustically noncompact, the dipole sound produced from the diffraction of the turbulence by either the leading or trailing edge is subsequently ''back-scattered'' by the opposite edge of the foil. Adamczyk (1974), Amiet (1975), and Martinez and Widnall (1980) accounted for the back-scattering by the trailing edge of the leading edge dipole sound for a surface without thickness. Roger, Moreau, and Wang (2002) extended Amiet's (1975) solution with a second scattering correction term for an application to a trailing edge noise problem. Howe (2001a) constructed a Green's function to account for the multiple back-scattering by both edges. The Green's function also included the dependence of the shape of the edge for the diffraction of the incident turbulence (Howe, 1975) for trailing edge noise estimation. In this paper, this Green's function (Howe, 2001a) is used to predict leading edge noise from a foil with thickness.

In practice, the use of a Green's function to predict the dipole sound produced by the diffraction of a turbulent flow requires that the trajectories of the incident sources are in accordance with a potential flow past the diffracting surface. This ''rapid distortion'' approximation attributes the diffraction to the Green's function and does not account for any interaction between the incident sources and their images (Howe, 1999). This approximation is satisfactory as long as the turbulence fluctuations are less than 10% of the free stream velocity (Grace, 2001).

The acoustic analogy of Howe (1975) has been utilized with a shape dependent Green's function to estimate the dipole sound from the diffraction of the sources within turbulent boundary layers by trailing edges of various contours (Howe, 1988, 1998a, 2000). Calculations of the shape dependence of the unsteady lift for turbulence encountering the leading edge have been restricted to numerical models with idealized representations of the flow acoustic source. The spatial domain formulation of the blade vortex interaction (BVI) models of Martinez and Rudzynsky (1997) and Grace (2001) predict a reduction in an unsteady lift with increasing frequency for a thick foil relative to a foil without thickness. These trends agree with the leading edge dipole sound measurements made by Olsen and Wagner (1982) with airfoils of varying thickness. The BVI calculations are also substantiated by the difference between the measured dipole sound spectra from a NACA 0012 airfoil and the sound spectra predicted by a zero thickness theory (Paterson and Amiet, 1976, 1977) as illustrated in Fig. 1. Paterson and Amiet attributed the discrepancy between their theory and data to the thickness of the foil for convection wavelengths of the incident turbulence that were smaller than the foil thickness. Numerical calculations of the unsteady lift in the wavenumber domain have not captured the thickness effect. In particular, Grace (2001) showed that calculations performed in the wave number domain of the unsteady lift response of thick foils to ''gusts'' incident to the leading edge recover the zero thickness result.

The current work uses acoustic analogies and a shape dependent Green's function to predict the lifting dipole sound from the leading edge diffraction of homogeneous turbulence by a thick, symmetric foil at a zero angle of attack to the mean flow. The foil thickness, h, is restricted to being acoustically compact, ($hk_0 \ll 1$) while the chord length, c, is greater than an acoustic wavelength, ($ck_0 > 1$).

In Sec. II, the portion of the Howe (1975) aero-acoustic

a)Electronic mail: GershfeldJL@nswccd.navy.mil

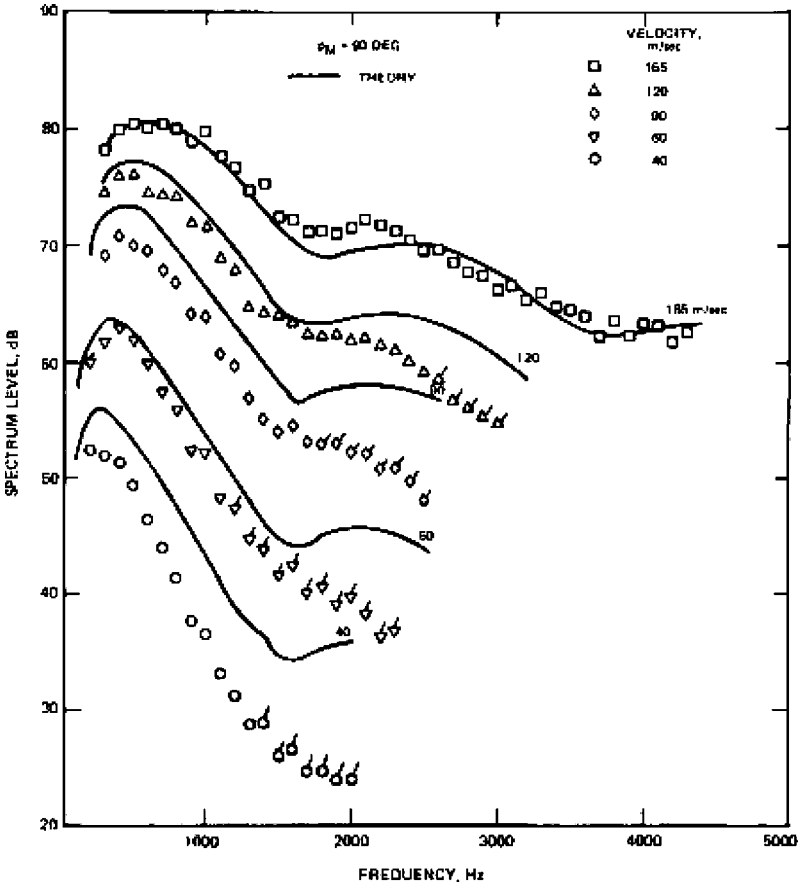


FIG. 1. A comparison of the measured (points) and predicted (lines) far-field dipole sound. Originally published in ''Noise and surface pressure response of an airfoil to incident turbulence,'' by R. W. Paterson and R. K. Amiet, J. Aircr. **14**, 729–736, Fig. 4. Copyright © 1977 by The American Institute of Aeronautics and Astronautics, Inc. Reprinted with permission.

source that generates the lifting component of the dipole sound is expressed in terms of the mean free stream velocity and the fluctuating up-wash velocity that also arises in Sears' (1941) analysis. The source is tailored to be linear in fluctuating variables, without mean shear (Blake, 1986; Howe, 1989) and spatially homogeneous. The near field of the quadrupole radiation on a rigid plane is then obtained in terms of the stagnation enthalpy from the volume integral of this source with the rigid plane Green's function.

In Sec. III, the Kirchhoff–Helmholtz integral equation for rigid surface diffraction of the stagnation enthalpy models the leading edge lifting dipole sound produced by a foil. The Green's function incorporates the effect of the foil thickness (Howe, 1998a) and the acoustic noncompactness of the airfoil's chord length (Howe, 2001a). The incident surface stagnation enthalpy is expressed in terms of the mean velocity and the fluctuating up-wash velocity spectrum obtained from Sec. II. This formula, along with the measurements of the incident flow turbulence reported by Paterson and Amiet (1976, 1977), are used to predict the measured dipole sound for a NACA0012 airfoil ($h/c=0.12$) and the corresponding zero thickness theory ($h/c=0$) for a variety of reported flow speeds.

In Sec. IV, the acoustic source is tailored to be linear with respect to fluctuating terms and to contain the mean shear (Kraichnan, 1956; Chase, 1980; Blake, 1986). A velocity potential cannot describe the mean shear flow. However, the near field component of the Green's function accounts for the diffraction of the distributed shear source as if it ''rapidly distorts'' as it convects past the leading edge as a potential flow. The shear source is specified to be spatially homogeneous, and the imposed surface pressure spectrum from the near field of the quadrupole radiation on a rigid plane surface is obtained. The calculated surface pressure spectrum for the mean shear flow, when combined with the results of Sec. III, yield a prediction formula for the leading edge lifting dipole sound of a thick, acoustically noncompact foil cutting through a shear layer.

II. WALL STAGNATION ENTHALPY SPECTRUM FROM HOMOGENEOUS TURBULENCE WITHOUT MEAN SHEAR

The stagnation enthalpy spectrum, B, imposed on a rigid plane $(y_1,y_2=0,y_3)$ will be calculated from the acoustic near field from a spatially homogeneous turbulence field as illustrated in Fig. 2. The turbulent flow is specified to have

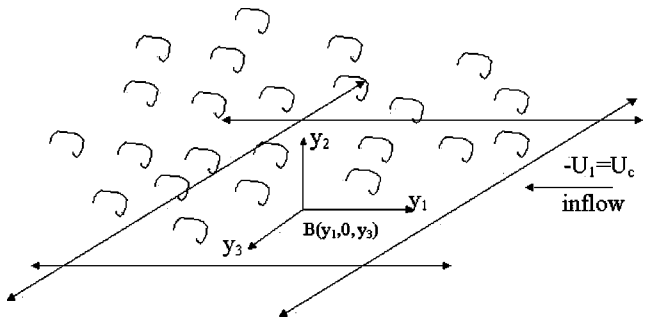


FIG. 2. Homogeneous turbulence over a rigid plane and imposed wall stagnation enthalpy, $B(y_1,0,y_3)$.

no mean shear so that the resultant wall stagnation enthalpy spectrum will serve as the incident field that will be used in Sec. III to calculate the leading edge noise from homogeneous turbulence in the absence of mean shear. The radiating stagnation enthalpy, **B**, is described by the acoustic analogy of Howe (1975) for vortex sound:

$$\nabla^2\mathbf{B}-\frac{1}{c_0^2}\frac{\partial^2\mathbf{B}}{\partial t^2}=-\nabla\cdot(\boldsymbol{\Omega}\times\mathbf{U}), \tag{1}$$

where $\boldsymbol{\Omega}$ and **U** represent, respectively, the sum of the mean and fluctuating vorticity and velocity. The stagnation enthalpy is expressed in terms of the acoustic pressure, p, the fluid density, ρ, and the fluctuating velocity, u:

$$B=\int dp/\rho+0.5u^2. \tag{2}$$

The component of the aero-acoustic source that contributes to the lifting dipoles with axes aligned normal to the free streamlines (Blake, 1986; Howe, 1989) shall be used to model the incident source,

$$\boldsymbol{\nabla}\cdot(\Omega'\times\mathbf{U}')=\frac{\partial}{\partial y_2}(\Omega_3'\times U_1)(y,t). \tag{3}$$

The primed ($'$) variables are in the time domain, and the subscripts 1, 3, and 2 denote respectively the streamwise, spanwise, and surface normal directions. The source in Eq. (3) is linear with respect to the fluctuating terms, and only the spanwise vorticity, Ω_3, contains the fluctuating components of the source. The subsequent planar wave number spectrum of the near field radiation is given by Ffowcs-Williams (1965),

$$B(k;\omega)=\frac{1}{(2\pi)^2}\int\int_V\int dV(y)\,\frac{\partial(\Omega_3\times U_1)(y,\omega)}{\partial y_2}\times\int\int_S d^2x_{1,3}\frac{e^{-ik_{1,3}\cdot x_{1,3}+ik_0r}}{r}, \tag{4}$$

where the unprimed variables have been Fourier transformed from time to frequency and $k_{1,3}=(k_1,k_3)$, $x_{1,3}=(x_1,x_3)$, $d^2x_{1,3}=dx_1\,dx_3$. When the in-plane Fourier transform of the Green's function,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int\frac{e^{ik_0r}}{r}e^{-ik_{1,3}\cdot x_{1,3}}d^2x_{1,3}=i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}e^{-ik_{1,3}\cdot y_{1,3}}, \tag{5}$$

is substituted into Eq. (4), the stagnation enthalpy wave number spectrum becomes

$$B(k,\omega)=\frac{1}{(2\pi)^2}\int_V\int\int dV(y)\,\frac{\partial(\Omega_3\times U_1)}{\partial y_2}\times i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}e^{-ik_{1,3}\cdot y_{1,3}}, \tag{6}$$

where

$$\gamma(k)=\sqrt{k_0^2-k_1^2-k_3^2},$$

and $k_0=\omega/c_0$, is the acoustic wave number. Rearranging the terms of Eq. (6),

$$B(k,\omega)=\frac{1}{(2\pi)^2}\int\int_V\int dy_2\,d^2y_{1,3}\,e^{-ik_{1,3}\cdot y_{1,3}}\times\frac{\partial(\Omega_3\times U_1)(y_1,y_2,y_3,\omega)}{\partial y_2}i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}, \tag{7}$$

while taking the Fourier transform of the source variables with respect to the in-plane coordinates (denoted by the overbar):

$$\overline{(\Omega_3U_1)}(k_{1,3},y_2,\omega)=\int_{-\infty}^{+\infty}\int d^2y_{1,3}\frac{1}{(2\pi)^2}e^{-ik_{1,3}\cdot y_{1,3}}(\Omega_3U_1)(y_1,y_2,y_3,\omega), \tag{8}$$

and upon substitution of Eq. (8) into Eq. (7), in the absence of mean shear, yields the following expression for the wall stagnation enthalpy wave number spectrum:

$$B(k,\omega)=\int_0^\infty dy_2\,U_1\overline{\frac{\partial(\Omega_3(k_{1,3},y_2,\omega))}{\partial y_2}}i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}. \tag{9}$$

Howe's (1975) acoustic source in Eq. (9) can be written in terms of the fluctuating up-wash velocity, u_2, and free stream velocity, U_1 from Sears' (1941) analysis. The solenoidal velocity wave number spectrum can be expressed in terms of the vorticity wave number spectrum (Howe, 1998a):

$$\hat{u}(k)=ik\times\hat{\Omega}(k)/|k|^2,\quad |k|^2=k_1^2+k_2^2+k_3^2. \tag{10}$$

The up-wash velocity wave number spectrum becomes

$$\hat{u}_2(k)=i(k_1\times\hat{\Omega}_3(k))_2/|k|^2. \tag{11}$$

Rearranging Eq. (11),

$$\frac{|k|^2}{k_1}U_1\hat{u}_2(k)=-i(\hat{\Omega}_3(k)\times U_1)_2, \tag{12}$$

and taking the inverse Fourier transform, Eq. (12), with respect to k_2,

$$\int_{-\infty}^{+\infty}dk_2\,e^{ik_2y_2}\left(i\frac{|k|^2}{k_1}U_1\hat{u}_2(k_1,k_2,k_3,\omega)\right)=\overline{\Omega_3(k_1,y_2,k_3,\omega)}U_1, \tag{13}$$

and then substituting Eq. (13) into Eq. (9), yields the wall stagnation enthalpy wave number spectrum in terms of the free stream velocity and fluctuating up-wash velocity:

$$B(k,\omega)=\int_0^\infty dy_2\frac{\partial}{\partial y_2}\times\left\{\int_{-\infty}^{+\infty}dk_2\,e^{ik_2y_2}\left(i\frac{|k|^2}{k_1}U_1\hat{u}_2(k_1,k_2,k_3,\omega)\right)\right\}\times i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}. \tag{14}$$

In the absence of mean shear, Eq. (14) becomes

$$B(k,\omega)=\int_0^{\infty} dy_2\, i\,\frac{e^{iy_2\gamma(k)}}{\gamma(k)}\left\{\int_{-\infty}^{+\infty} dk_2(ik_2)e^{ik_2y_2}\times\left(i\,\frac{|k|^2}{k_1}U_1\hat{u}_2(k_1,k_2,k_3,\omega)\right)\right\}. \quad (15)$$

The complex conjugate (*) of the wall stagnation enthalpy,

$$B^*(k',\omega')=\int_0^{\infty} dy_2'(-i)\,\frac{e^{-iy_2'\gamma^*(k')}}{\gamma^*(k')}\times\left\{\int_{-\infty}^{+\infty} dk_2'(-ik_2')e^{-ik_2'y_2'}\times\left(-i\,\frac{|k'|^2}{k_1'}U_1\hat{u}_2^*(k_1',k_2',k_3',\omega')\right)\right\}, \quad (16)$$

with Eq. (15), enables the wall stagnation enthalpy wave number power spectrum to be determined:

$$\Phi_{BB}(k,\omega)\,\delta(k-k')\,\delta(\omega-\omega')=\langle B(k,\omega)B^*(k',\omega')\rangle, \quad (17)$$

where the angled parentheses ($\langle\ \rangle$) signify an ensemble average. Substituting Eqs. (15) and (16) into (17) gives the wall stagnation enthalpy wave number power spectrum,

$$\Phi_{BB}(k_{1,3},\omega)=\int_0^{+\infty} dy_2\int_0^{+\infty} dy_2'\times\left[\int_{-\infty}^{+\infty} dk_2\, e^{ik_2(y-y_2')}U_1^2\frac{k_2^2}{k_1^2}(k_1^2+k_2^2+k_3^2)^2\times\Phi_{u_2u_2}(k_{1,2,3},\omega)\right]\frac{e^{iy_2\gamma(k)-iy_2'\gamma^*(k)}}{|\gamma(k)|^2}, \quad (18)$$

where the up-wash velocity power spectrum is also obtained as in Eq. (17):

$$\Phi_{u_2u_2}(k,\omega)\,\delta(k-k')\,\delta(\omega-\omega')=\langle\hat{u}_2(k,\omega)\hat{u}_2^*(k',\omega')\rangle. \quad (19)$$

In order to facilitate the integration of Eq. (18), the up-wash velocity wave number frequency power spectrum is expressed in a separable form,

$$\Phi_{u_2u_2}(k_1,k_2,k_3,\omega)=\Phi_{u_2u_2}(\omega)\,\phi_1(k_1)\,\phi_2(k_2)\,\phi_3(k_3), \quad (20)$$

where all of the wave number spectra are normalized accordingly:

$$\int_{-\infty}^{+\infty} dk_j\,\phi_j(k_j)=1. \quad (21)$$

The streamwise (1) and vertical (2) wave number spectra will use Corcos' (1963) representation:

$$\phi_1(k_1)=\frac{l_1/\pi}{1+l_1^2(k_1-k_c)^2};\quad \phi_2(k_2)=\frac{l_2/\pi}{1+l_2^2k_2^2};$$
$$l_1\approx 9U_c/\omega,\quad l_2\approx 1.4U_c/\omega. \quad (22)$$

Referring to Fig. 2, the convection wave number $k_c=-\omega/U_1$ for $\omega>0$. The sign of the convection velocity has no bearing on the value of the stagnation enthalpy power spectrum. This will become apparent since all of the subsequent expressions for the incident surface stagnation enthalpy power spectra involve even powers of the streamwise wave number, k_1. The spanwise (3) wave number spectrum will be approximated with a Gaussian form used by Blake (1971):

$$\phi_3(k_3)=l_3\frac{e^{-(k_3l_3/2)^2}}{2\sqrt{\pi}};\quad l_3\approx 1.4U_c/\omega. \quad (23)$$

Substituting Eqs. (22) into Eq. (18) and integrating Eq. (18) with respect to k_2 yields

$$\begin{aligned}\Phi_{BB}(k_{1,3},\omega)&=\int_0^{+\infty} dy_2\int_0^{+\infty} dy_2'\,U_1^2\Phi_{u_2u_2}(k_{1,3},\omega)\\&\times\frac{e^{iy_2\gamma(k)-iy_2'\gamma^*(k)}}{|\gamma(k)|^2}\,\frac{1}{k_1^2\pi}\{l_2[l_2^{-7}(e^{-(y_2-y_2')/l_2}\\&\times(-1+2k_1^2l_2^2+2k_3^2l_2^2-k_1^4l_2^4-2k_1^2k_3^2l_2^4-k_3^4l_2^4)\pi)\\&+2(-1+k_1^2l_2^2+k_3^2l_2^2)^2\pi\delta(y_2-y_2')l_2^{-6}\\&-2(-1+k_1^2l_2^2+k_3^2l_2^2)\pi\delta''(y_2-y_2')l_2^{-4}\\&+2\pi\delta^{iv}(y_2-y_2')l_2^{-2}]\},\end{aligned} \quad (24)$$

where δ^n is the nth order derivative of the delta function. The relationship between the convolution of the nth order derivative of the delta function, with a function, f, and the nth order derivative of f is given in Eq. (25) (Lighthill 1958):

$$\int_0^{+\infty} dy_2\, f(y_2)\,\delta^n(y_2-y_2')=(-1)^n f^n(y_2'),\quad 0<y_2'<\infty. \quad (25)$$

If the integral scale is much smaller than the spatial extent of the turbulence, then the vertical (2) correlation function can be represented as a product of the integral scale and the delta function:

$$e^{-|y_2-y_2'|/l_2}=l_2\,\delta(y_2-y_2');\quad l_2\approx 1.4U_c/\omega=1.4k_c^{-1}. \quad (26)$$

Substituting Eq. (26) into the expression for the wall stagnation enthalpy wave number power spectrum [Eq. (24)] yields

$$\Phi_{BB}(k_{1,3},\omega)$$
$$= \int_0^{+\infty} dy_2 \int_0^{+\infty} dy_2' U_1^2 \Phi_{u_2u_2}(k_{1,3},\omega)$$
$$\times \frac{e^{iy_2\gamma(k)-iy_2'\gamma^*(k)}}{|\gamma(k)|^2} \frac{1}{k_1^2\pi} \{l_2[l_2^{-7}(\delta(y_2-y_2')$$
$$\times l_2(-1+2k_1^2l_2^2+2k_3^2l_2^2-k_1^4l_2^4-2k_1^2k_3^2l_2^4-k_3^4l_2^4)\pi)$$
$$+2(-1+k_1^2l_2^2+k_3^2l_2^2)^2\pi\delta(y_2-y_2')l_2^{-6}$$
$$-2(-1+k_1^2l_2^2+k_3^2l_2^2)\pi\delta''(y_2-y_2')l_2^{-4}$$
$$+2\pi\delta^{iv}(y_2-y_2')l_2^{-2}]\}. \quad (27)$$

The major contribution to the integrand in Eq. (27) is from small values of k_3. Accordingly, for $k_1>k_0$ and $k_1>k_3$, the exponential phase in Eq. (27) may be approximated as

$$\gamma(k)\approx ik_1.$$

Integrating Eq. (27) with respect to y_2,

$$\Phi_{BB}(k_{1,3},\omega)=\int_0^{+\infty} dy_2' U_1^2 \Phi_{u_2u_2}(k_{1,3},\omega)\frac{e^{-2k_1y_2'}}{|\gamma(k)|^2}\frac{1}{k_1^2\pi}$$
$$\times\{l_2[l_2^{-7}(l_2(-1+2k_1^2l_2^2+2k_3^2l_2^2-k_1^4l_2^4$$
$$-2k_1^2k_3^2l_2^4-k_3^4l_2^4)\pi)+2(-1+k_1^2l_2^2$$
$$+k_3^2l_2^2)^2\pi l_2^{-6}-2(-1+k_1^2l_2^2$$
$$+k_3^2l_2^2)\pi 4k_1^2l_2^{-4}+2\pi 16k_1^4l_2^{-2}]\}, \quad (28)$$

and then integrating Eq. (28) with respect to y_2' yields the wall stagnation enthalpy frequency wave number power spectrum:

$$\Phi_{BB}(k_{1,3},\omega)=U_1^2\Phi_{u_2u_2}(k_{1,3},\omega)\frac{1}{|\gamma(k)|^2}\frac{1}{k_1^2\pi}\cdot\frac{1}{2k_1}$$
$$\times\{l_2[l_2^{-7}(l_2(-1+2k_1^2l_2^2+2k_3^2l_2^2-k_1^4l_2^4$$
$$-2k_1^2k_3^2l_2^4-k_3^4l_2^4)\pi)+2(-1+k_1^2l_2^2$$
$$+k_3^2l_2^2)^2\pi l_2^{-6}-2(-1+k_1^2l_2^2$$
$$+k_3^2l_2^2)\pi 4k_1^2l_2^{-4}+2\pi 16k_1^4l_2^{-2}]\}. \quad (29)$$

The wall stagnation enthalpy frequency power spectrum is determined by integrating Eq. (29) with respect to k_1 and k_3. Using approximate forms of the wave number spectra of the up-wash velocity facilitates the integration. Accordingly, since the streamwise wave number spectrum of the up-wash velocity [Eq. (22)] is peaked at the convection ridge, and since the rest of the integrand, denoted by $f(k_1,k_3)$ of Eq. (29), is varying slowly in this region, the streamwise wave number spectrum shall be represented as a delta function. Thus

$$\int_{-\infty}^{+\infty} dk_1 f(k_1,k_3)\phi_1(k_1)$$
$$\approx \int_{-\infty}^{+\infty} dk_1 f(k_1,k_3)\delta(k_1-k_c)=f(k_c,k_3). \quad (30)$$

The wall stagnation enthalpy power spectrum becomes

$$\Phi_{BB}(k_c,k_3,\omega)=U_1^2\Phi_{u_2u_2}(\omega)\frac{l_3e^{-(k_3l_3/2)^2}}{2\sqrt{\pi}}\frac{1}{k_c^2+k_3^2}\frac{1}{k_c^2\pi}$$
$$\cdot\frac{1}{2k_c}\times\{l_2[l_2^{-7}(l_2(-1+2k_c^2l_2^2+2k_3^2l_2^2$$
$$-k_c^4l_2^4-2k_c^2k_3^2l_2^4-k_3^4l_2^4)\pi)+2(-1$$
$$+k_c^2l_2^2+k_3^2l_2^2)^2\pi l_2^{-6}-2(-1+k_c^2l_2^2$$
$$+k_3^2l_2^2)\pi 4k_c^2l_2^{-4}+2\pi 16k_c^4l_2^{-2}]\}. \quad (31)$$

The integration of Eq. (31) with respect to k_3 yields the component of the wall stagnation enthalpy power spectrum that contributes to the lifting dipole sound expressed in terms of the mean flow velocity and the fluctuating up-wash velocity:

$$\Phi_{BB}(\omega)\approx 6.4U_1^2\Phi_{u_2u_2}(\omega). \quad (32)$$

III. DIPOLE SOUND FROM A THICK, SYMMETRIC FOIL IN A TURBULENT FLOW WITHOUT MEAN SHEAR

In Sec. II, the surface stagnation enthalpy, B, was calculated on a rigid plane due to the near field acoustic radiation from the component of Howe's (1975) acoustic source that gives rise to the lifting components of the dipoles. The source was specified to be without mean shear and was expressed in terms of the free stream velocity and the fluctuating up-wash velocity. In this section, the dipole sound, p_{rad}, due to the diffraction of these sources by the leading edge of a thick, symmetric foil, will be calculated using a rigid surface Green function, $G(x,y,\omega)$ and incident "blocked" surface stagnation enthalpy, B, that satisfies the Kirchhoff–Helmholtz integral equation,

$$B_{\mathrm{rad}}(x,\omega)=-\frac{1}{2}\int_s\int\frac{\partial B(y_1,y_2=0,y_3,\omega)}{\partial y_n}$$
$$\times G(x,y,\omega)dS(y), \quad (33)$$

where B_{rad} is the far field radiated stagnation enthalpy and y_n is normal to the surface and is directed into the fluid. At low flow Mach numbers, $M_c=U/c_0<1$, the radiated stagnation enthalpy, B_{rad}, is related to the acoustic pressure in the far field, p_{rad} (Howe, 1978),

$$P_{\mathrm{rad}}=B_{\mathrm{rad}}\cdot\rho/(1+M_c(x_1/R)), \quad (34)$$

where $M_c(x_1/R)=M_{OR}$ is the flow Mach number component in the observer direction. The acoustic pressure can then be determined in the far field from Eqs. (33) and (34):

$$P_{\mathrm{rad}}(x,\omega)=-\frac{\rho}{2(1+M_c(x_1/R))}$$
$$\times\int_s\int\frac{\partial B(y_1,y_2=0,y_3,\omega)}{\partial y_n}G(x,y,\omega)dS(y). \quad (35)$$

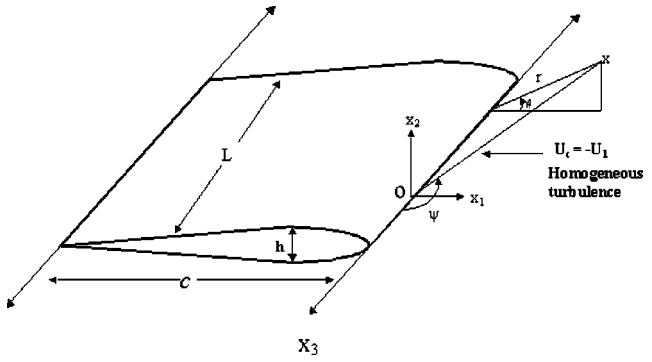


FIG. 3. Coordinates for the acoustic field of the two-dimensional NACA 0012 airfoil with wetted span length, L, thickness, h, chord length, c, and convection velocity, $U_c=-U_1$.

The coordinates for the radiation field are illustrated for a foil of thickness, h, chord length, c, and infinite span in Fig. 3. Only a spanwise length, L, is wetted by the turbulent flow. The foil encounters homogeneous turbulence with convection velocity, $U_c=-U_1$. The cross section of the foil (Fig. 4) shows that the turbulence flows past both sides of the leading edge of the foil for $|x_3|<L/2$. The half-plane ''slab'' scattering geometry that is used to model the leading edge diffraction of the incident stagnation enthalpy [Eq. (35)] is given in Fig. 5. The incident stagnation enthalpy is imposed by the turbulence on both sides of the foil. A leading edge unsteady Kutta condition is not applied and the incident vortical sources follow the potential flow streamlines associated with the foil geometry and the free stream velocity indicated in the figures. The surface of the foil is defined to have no discontinuities in the surface curvature. Even though the sources accelerate due to the curvature of the potential streamlines in the vicinity of the leading edge, it will still be assumed that the incident source wave number spectrum is peaked at the convection ridge associated with the velocity upstream of the leading edge so that the frozen flow approximation can be made.

The dipole sound spectrum due to the diffraction of the incident turbulence by the leading edge may also be modeled using a surface pressure scattering formulation with Howe's (1998a) trailing edge Green's function for a half-plane rigid surface with thickness, h and rounded edge using Eq. (33) with p replacing B:

$$\Phi_{PP}^{\mathrm{rad}}(x,\omega)\approx \frac{\omega L\,\sin^2(\theta/2)\sin\psi}{\pi^2 c_0|x|^2}\,\frac{\pi}{k_c} \times e^{-\omega h/(2U_c)}\Phi_{PP}^{\mathrm{surf}}(\omega)l_3(\omega)/\pi. \tag{36}$$

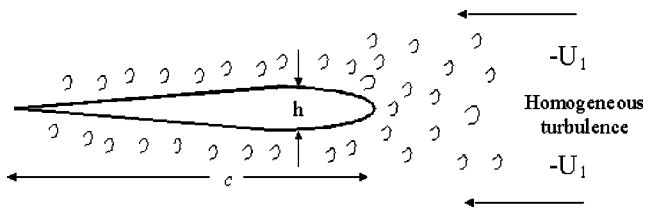


FIG. 4. A NACA 0012 airfoil with chord length, c, thickness, h, ($h/c=0.12$) and a homogeneous, turbulent flow with velocity, $-U_1$.

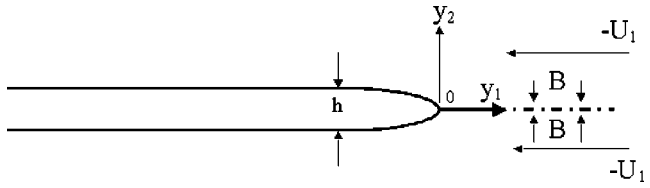


FIG. 5. The representation of the foil as a semi-infinite slab with thickness, h, and rounded leading edge for an estimation of the leading edge force due to the leading edge diffraction of the incident surface stagnation enthalpy, B.

$\Phi_{PP}^{\mathrm{rad}}(x,\omega)$ and $\Phi_{PP}^{\mathrm{surf}}(\omega)$ denote, respectively, the radiated acoustic pressure power spectrum and the incident surface pressure power spectrum, π/k_c contains the square of the magnitude of the high frequency limit to the Sears' function, and $e^{-\omega h/(2U_c)}$ is the thickness correction. The right hand side of the equation has been multiplied by a factor of two in order to account for the presumed statistical independence of the turbulent flows around each side of the airfoil's leading edge. The thickness correction to Sears' function couples only to $\phi_1(k_1)$ components of the incident sources. This implies that the unsteady lift is independent of $\phi_2(k_2)$ for k_2 not equal to zero. Referring to Eq. (28), incident sources at the convection wave number, $k_c=\omega/U_c$, and location, y_2, normal to the surface have weighted contributions to the surface pressure spectrum that are exponentially attenuated by their product:

$$e^{-2\omega y_2/(U_c)}. \tag{37}$$

The thickness correction, in Eq. (36),

$$e^{-\omega h/(2U_c)}, \tag{38}$$

suggests that the sound generated by a foil with thickness, h, in a spatially homogeneous turbulence field can be alternatively represented by a foil with zero thickness where the incident sources are off-set from the surface in the surface normal direction by a distance equal to one quarter of the maximum section thickness, $|y_2|>h/4$.

Equations (32), (33), (35) and (36) yield an expression for the leading edge dipole sound pressure power spectrum due to the diffraction of the homogeneous turbulence in terms of the mean velocity and fluctuating up-wash velocity:

$$\Phi_{PP}^{\mathrm{rad}}(x,\omega)\approx \frac{\omega L\,\sin^2(\theta/2)\sin\psi}{(1+M_{OR})^2\pi^2 c_0|x|^2}\,\frac{\pi}{k_c} \times e^{-\omega h/(2U_c)}6.4\rho^2U_1^2\Phi_{u_2u_2}(\omega)l_3(\omega)/\pi. \tag{39}$$

The multiple ''back-scattering'' of the acoustic waves generated by the leading edge diffraction of the turbulence by the edges of the foil occurs when the chord length is finite but acoustically noncompact. Howe's (2001a) normalized Green's function, G_T, shall be used to model this effect. This yields an equation for the dipole sound from the thick but acoustically noncompact airfoil in terms of the mean velocity and the up-wash velocity spectrum and spanwise integral scale:

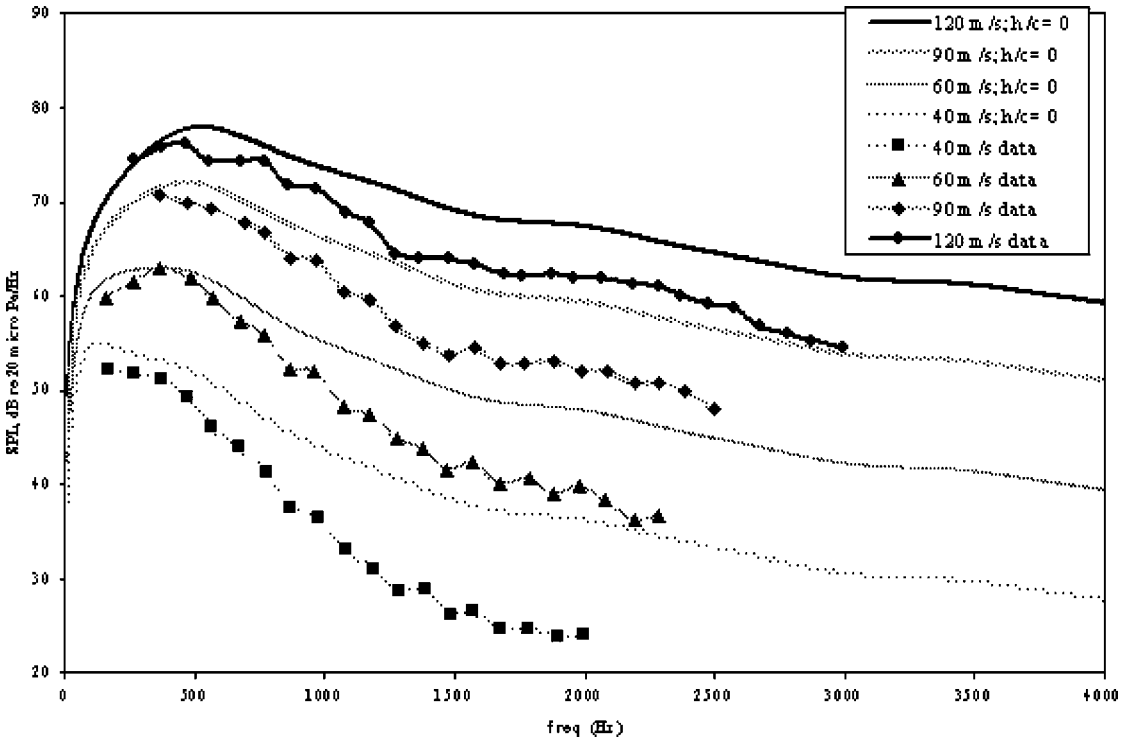


FIG. 6. A comparison of the NACA 0012 ($h/c=0.12$) dipole sound pressure measurements (dB re (20 micro Pa)^2/Hz) of Paterson and Amiet (1976, 1977) (points), and the thickness theory [Eq. (40)] with $h/c=0.0$ (line) for free stream velocities of 40 m/s, 60 m/s, 90 m/s, and 120 m/s.

$$\Phi_{PP}^{\mathrm{rad}}(x,\omega)\approx\frac{|G_T(k_0c\sin(\psi))|^2}{(1+M_{OR})^2}\frac{\omega L\sin^2(\theta/2)\sin\psi}{\pi^2 c_0|x|^2}\frac{\pi}{k_c}\times e^{-\omega h/(2U_c)}6.4\rho^2U_1^2\Phi_{u_2u_2}(\omega)l_3(\omega)/(\pi). \tag{40}$$

The normalized Green's function is given by

$$G_T(x,y,\omega)=1+\frac{G_{LE}(x,y,\omega)+G_{TE}(x,y,\omega)}{G_{1/2\mathrm{plane}}(x,y,\omega)}, \tag{41}$$

where G_{LE}, G_{TE}, $G_{1/2\mathrm{plane}}$ represent, respectively, the contributions from the leading edge, the trailing edge and the Green's function for the "half-plane slab" diffracting surface used in Eq. (39). These are given by Howe (2001a) with (Fig. 3) $\theta=\psi=\pi/2$:

$$G_{TE}(x,y,\omega)\approx\frac{-\varphi^*(y)e^{ik_0(|x'|)}}{\pi^2\sqrt{2ic}|x|[1+e^{2ik_0c}/(2\pi ik_0c)]}\times F(2\sqrt{k_0c\cos^2(\pi/4)/\pi}), \tag{42}$$

$$G_{LE}(x,y,\omega)\approx\frac{\sqrt{k_0}\sin^{1/2}(\psi)\varphi^*(y)e^{ik_0(|x'|)}}{i\pi^{3/2}|x|[1+e^{2ik_0c}/(2\pi ik_0c)]}\times F(2\sqrt{k_0c\cos^2(\pi/4)/\pi}). \tag{43}$$

$F(x)=f(x)+ig(x)$ are the Fresnel integral auxiliary functions whose approximations are given by

$$f(x)=\frac{1+0.926x}{2+1.792x+3.104x^2};$$

$$g(x)=\frac{1}{2+4.142x+3.492x^2+6.67x^3}. \tag{44}$$

The predicted dipole sound from Eq. (40) shall be compared with the measured sound reported by Paterson and Amiet (1976) in Fig. 1. Accordingly, the current predictions will use the reported up-wash velocity power spectrum with the von Karman formulation for isotropic turbulence with an integral scale, $\lambda_f=0.03$ m, and a turbulence intensity, $u/U=0.045$,

$$\Phi_{u_2u_2}(\omega)=\frac{\overline{u^2}\lambda_f}{2\pi U}\frac{[1+\frac{8}{3}(k_c4\lambda_f/3)^2]}{[1+(k_c4\lambda_f/3)^2]^{11/6}}. \tag{45}$$

The spanwise correlation scale, $l_3(\omega)$, will also use the von Karman formulation,

$$l_3(\omega)=\frac{8}{3}\left[\frac{\Gamma(1/3)}{\Gamma(5/6)}\right]^2\lambda_f\frac{(4\lambda_fk_c/3)^2}{[3+8(4\lambda_fk_c/3)^2]\sqrt{1+(4\lambda_fk_c/3)^2}}, \tag{46}$$

yielding the same result as the Corcos model or the Gaussian model for frequencies above 300 Hz. Γ is the Gamma function with $\Gamma(1/3)=2.68$, and $\Gamma(5/6)=1.13$.

The airfoil chord length is $c=0.23$ m, the span is $L=0.53$ m. The thickness (h) to chord (c) ratio of the NACA foil is $h/c=0.12$. The dipole sound in Fig. 1 was measured at a distance $x_2=2.25$ m from the mid-chord of the foil. The Mach number correction in Eq. (34) becomes, $M_{OR}=0.05\,M_c$ and $\theta\cong\psi\cong\pi/2$. In Fig. 6 for $h/c=0$ and in Fig. 7 for $h/c=0.12$, the predicted dipole sound spectra are plotted using Eq. (40) at velocities of 40 m/s, 60 m/s, 90 m/s, and 120 m/s. These predictions are compared at each velocity to the measured dipole sound from Paterson and Amiet (1976, 1977).

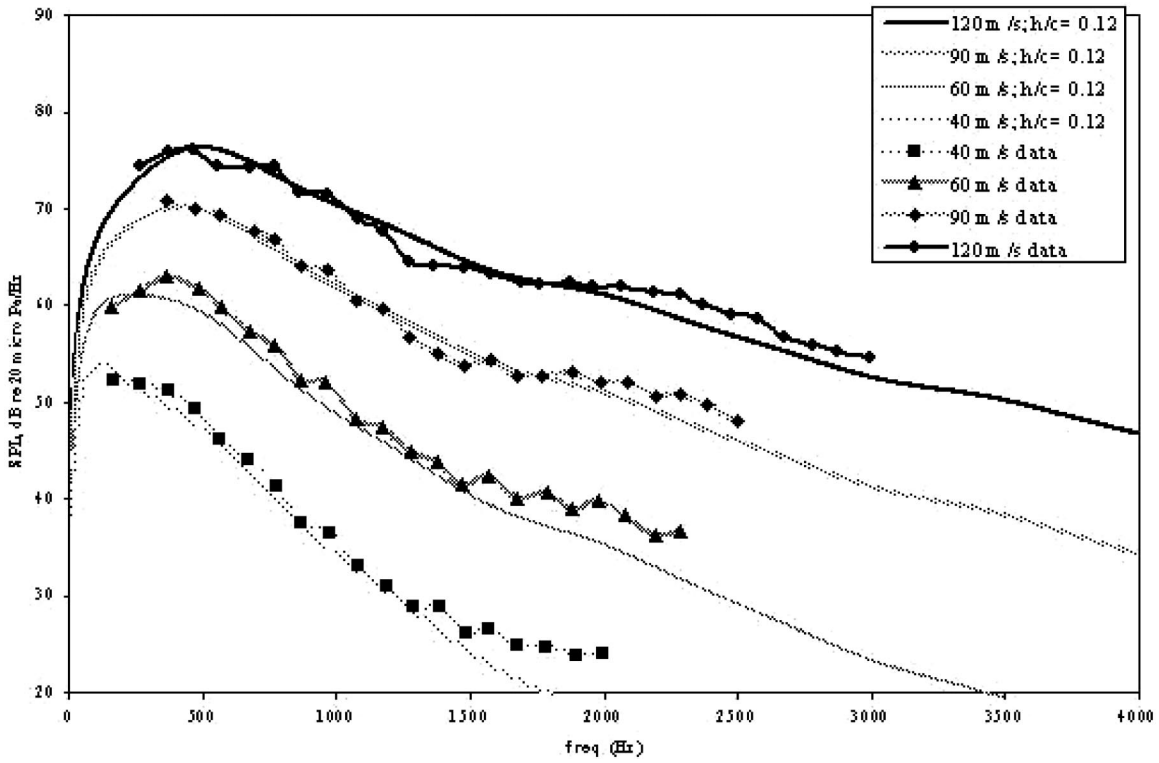


FIG. 7. A comparison of the NACA 0012 ($h/c=0.12$) dipole sound presure measurements (dB re (20 micro Pa)^2/Hz) of Paterson and Amiet (1976, 1977) (points), and the thickness theory [Eq. (40)] with $h/c=0.12$ (line) for free stream velocities of 40 m/s, 60 m/s, 90 m/s, and 120 m/s.

IV. DIPOLE SOUND FROM A THICK, SYMMETRIC FOIL IN A TURBULENT FLOW WITH MEAN SHEAR

In Sec. II, the turbulence quadrupole source was specified to be without mean shear. These types of turbulent flows are typically generated by turbulent wake flows and jet induced shear layers that have had sufficient time to diffuse or by the outer portion of a turbulent boundary layer (TBL) flow. Conversely, wake, and jet flows have regions of high mean shear before sufficient diffusion occurs, as do turbulent boundary layer flows near the bounding surfaces. The portion of the acoustic source that contains the mean shear-turbulent (MS-T) source has been used to describe the acoustic near field produced from turbulent shear flows that occur in TBL flows (Kraichnan, 1956; Chase 1980; Blake, 1986) and jets (Moon and Zelazny, 1975). In this section, the dipole sound due to a MS-T source incident to the leading edge of a thick foil will be modeled.

Referring to Fig. 8, a foil of thickness, h, and chord length, c, encounters a turbulent shear layer with constant transverse mean shear $dU_1(y_2)/dy_2=$constant. The shear layer is taken to be much thicker than the foil's thickness so that the incident source from such a flow shall be considered to be spatially homogeneous. The force induced by the diffraction of the turbulent shear layer by the leading edge is formulated as a Kirchhoff–Helmholtz integral equation for the scattering of the incident surface pressure by a ''half-plane slab'' with thickness, h, and rounded leading edge (Fig. 9). The approach is identical to that described in Secs. II and III. In this section, the wall pressure spectrum imposed on a rigid plane will be calculated from the near field quadrupole radiation of the aero-acoustic source that has been specialized to be linear in fluctuating terms and to contain the mean shear. This pressure spectrum can be used with the results of Sec. III to calculate the leading edge noise for a thick foil cutting through a shear layer that will be specialized to be spatially homogeneous.

At low Mach numbers, the acoustic pressure, p, generated by the MS-T source is described by the following acoustic analogy:

$$\frac{1}{c_0^2}\frac{\partial^2 \mathbf{p}}{\partial t^2}-\nabla^2 \mathbf{p}\approx 2\rho_0 \frac{\partial U_1}{\partial y_2}\frac{\partial \mathbf{u}_2'}{\partial y_1}. \tag{47}$$

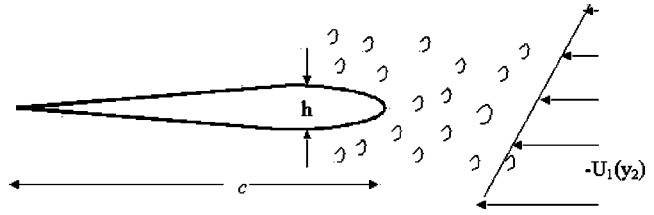


FIG. 8. A foil with chord length, c, thickness, h, and homogeneous turbulent inflow with mean shear, $dU_1/dy_2=$constant.

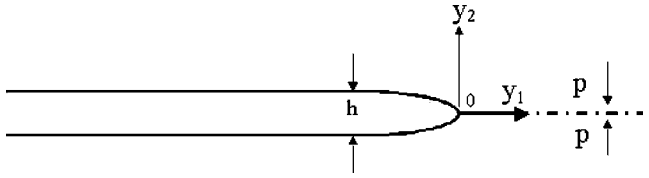


FIG. 9. The representation of the foil as a semi-infinite slab, with thickness, h, and rounded leading edge for estimation of the leading edge force due to the leading edge diffraction of the incident surface pressure, p.

The pressure, $\mathbf{p}$, and the fluctuating up-wash velocity, $\mathbf{u}_2'$, are in the time domain and are denoted by bold face type. The planar wave number spectrum of the wall pressure that is generated by the acoustic source in Eq. (47) becomes

$$p(k,\omega)=\frac{1}{(2\pi)^2}\int\int_V\int dV(y)2\rho_0\frac{\partial U_1(y)}{\partial y_2}\frac{\partial u_2'(y,\omega)}{\partial y_1}\times\int\int_S dx\frac{e^{-ik\cdot x+ik_0 r}}{r},\tag{48}$$

where the pressure, p, and up-wash velocity, u_2, have been Fourier transformed to the frequency domain and are no longer in bold face type.

The planar Fourier transform of the Green's Function is given by

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{e^{ik_0 r}}{r}e^{-ik_{1,3}\cdot x_{1,3}}d^2x_{1,3}=i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}e^{-ik_{1,3}\cdot y_{1,3}},\tag{49}$$

where

$$\gamma(k)=\sqrt{k_0^2-k_1^2-k_3^2},\tag{50}$$

and $k_0=\omega/c_0$, $k_{1,3}=(k_1,k_3)$, $y_{1,3}=(y_1,y_3)$, $x_{1,3}=(x_1,x_3)$, $d^2x_{1,3}=dx_1\,dx_3$.

Substituting Eq. (49) into Eq. (48),

$$p(k,\omega)=\frac{1}{(2\pi)^2}\int\int\int_V dy_2\,d^2y_{1,3}e^{-ik_{1,3}\cdot y_{1,3}}2\rho\times\frac{\partial U_1}{\partial y_2}\frac{\partial u_2'}{\partial y_1}(y_1,y_2,y_3,\omega)i\frac{e^{iy_2\gamma(k)}}{\gamma(k)},\tag{51}$$

and rearranging the terms of the integrand gives the following expression for the wall pressure wave number spectrum:

$$p(k,\omega)=\frac{1}{(2\pi)^2}\int_{y_2=0}^{\infty}dy_2\,2\rho\frac{\partial U_1}{\partial y_2}i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}\times\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}d^2y_{1,3}e^{-ik_{1,3}\cdot y_{1,3}}\frac{\partial u_2'}{\partial y_1}(y_1,y_2,y_3,\omega).\tag{52}$$

The Fourier transform of the fluctuating up-wash velocity with respect to the in plane coordinates is denoted by an overbar:

$$\overline{u_2'}(k_{1,3},y_2,\omega)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}d^2y_{1,3}\frac{1}{(2\pi)^2}\times e^{-ik_{1,3}\cdot y_{1,3}}u_2'(y_1,y_2,y_3,\omega),\tag{53}$$

and the Fourier transform of the derivative of the up-wash velocity is

$$ik_1\overline{u_2'}(k_{1,3},y_2,\omega)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}d^2y_{1,3}\frac{1}{(2\pi)^2}e^{-ik_{1,3}\cdot y_{1,3}}\times\frac{\partial u_2'(y_1,y_2,y_3,\omega)}{\partial y_1}.\tag{54}$$

Substituting Eq. (54) into Eq. (52) determines the wall pressure wave number spectrum,

$$p(k,\omega)=\int_{y_2=0}^{\infty}dy_2\,2\rho\frac{\partial U_1}{\partial y_2}i\frac{e^{iy_2\gamma(k)}}{\gamma(k)}ik_1\overline{u_2'}(k_1,y_2,k_3,\omega),\tag{55}$$

and the complex conjugate of the wall pressure spectrum becomes

$$p^*(k',\omega')=\int_{y_2=0}^{\infty}dy_2'\,2\rho\frac{\partial U_1'}{\partial y_2'}(-i)\frac{e^{-iy_2'\gamma^*(k')}}{\gamma^*(k')}\times(-ik_1)\overline{u_2'}(k_1',y_2',k_3',\omega').\tag{56}$$

The wall pressure power spectrum is then determined from the ensemble average of the pressure and its complex conjugate:

$$\Phi_{pp}(k,\omega)\,\delta(k-k')\,\delta(\omega-\omega')=\langle p(k,\omega)p^*(k',\omega')\rangle,\tag{57}$$

yielding the wall pressure wave number frequency power spectrum:

$$\Phi_{pp}(k_{1,3},\omega)=4\rho^2\int_0^{+\infty}dy_2\int_0^{+\infty}dy_2'\frac{\partial U_1'(y_2')}{\partial y_2'}\frac{\partial U_1(y_2)}{\partial y_2}\times k_1^2\Phi_{u_2u_2}(k_{1,3},y_2,y_2',\omega)]\frac{e^{iy_2\gamma(k)-iy_2'\gamma^*(k)}}{|\gamma(k)|^2}.\tag{58}$$

The following separable form of the up-wash velocity wave number power spectrum:

$$\Phi_{u_2u_2}(k_1,y_2,y_2',k_3,\omega)=\Phi(\omega)\phi_1(k_1)R_{22}(y_2,y_2')\phi_3(k_3),\tag{59}$$

facilitates the integration of Eq. (58). The wave number spectra are normalized,

$$\int_{-\infty}^{+\infty}dk_j\,\phi_j(k_j)=1,\tag{60}$$

and the streamwise (1) and spanwise (3) wave number spectra will use the following forms:

$$\phi_1(k_1)=\frac{l_1/\pi}{1+l_1^2(k_1-k_c)^2};\quad \phi_3(k_3)=\frac{l_3e^{-(k_3l_3/2)^2}}{2\sqrt{\pi}}.\tag{61}$$

The shear layer is considered to be much thicker than the airfoil. In this context, this model assumes that the surface normal integral scale, $l_2=1.4U_c/\omega$, is much smaller than the thickness of the shear layer. The surface normal correlation function is taken to be spatially homogeneous with integral scale l_2:

$$R_{22}(y_2,y_2',\omega)=\delta(y_2-y_2')l_2(\omega).\tag{62}$$

The major contribution to Eq. (58) is from small values of k_3. For $k_1>k_0$ and $k_1>k_3$, the exponential phase may be approximated as $\gamma(k)\sim ik_1$. Substituting Eqs. (59)–(62) into Eq. (58) and integrating with respect to y_2, and y_2' yields the following expression for the wall pressure frequency wave number power spectrum:

$$\Phi_{pp}(k_{1,3},\omega)=2\rho^2 k_1\left(\frac{\partial U_1(y_2)}{\partial y_2}\right)^2\Phi_{u_2u_2}(\omega)\phi_1(k_1)$$

$$\times\phi_3(k_3)l_2(\omega)\frac{1}{|\gamma(k)|^2}. \quad (63)$$

The wall pressure frequency spectrum is then obtained by integrating Eq. (63) with respect to k_1 and k_3. The frozen flow approximation is utilized to facilitate this integration by letting $\phi_1(k_1)\sim\delta(k_1-k_c)$. For $k_1^2+k_3^2>k_0^2$, the frequency–spanwise wave number power spectrum becomes

$$\Phi_{pp}(k_3,\omega)=2\rho^2 k_c\left(\frac{\partial U_1(y_2)}{\partial y_2}\right)^2\Phi_{u_2u_2}(\omega)\phi_3(k_3)$$

$$\times l_2(\omega)\frac{1}{k_c^2+k_3^2}. \quad (64)$$

Integrating Eq. (64) over k_3 yields the following expression for the MS-T wall pressure frequency power spectrum:

$$\Phi_{pp}(\omega)\cong 4\rho^2 U_1^2\Phi_{u_2u_2}(\omega)\left(\frac{1}{\omega}\frac{\partial U_1(y_2)}{\partial y_2}\right)^2. \quad (65)$$

The ratio of the point pressure spectrum from the MS-T source [Eq. (65)] to the source without mean shear (T-T) [Eq. (32)] is

$$\frac{\Phi_{pp}^{\text{MS-T}}(\omega)}{\Phi_{pp}^{\text{T-T}}(\omega)}\approx 0.6\left(\frac{\partial U_1}{\partial y_2}\frac{1}{\omega}\right)^2. \quad (66)$$

This defines a cut-off frequency below which the dipole sound from the MS-T source should dominate the dipole sound from the T-T source.

The leading edge dipole sound is determined by multiplying the ratio of the MS-T wall pressure spectrum to the T-T spectrum [Eq. (66)], by the dipole sound due to the T-T source [Eq. (40)]:

$$\Phi_{PP}^{\text{radMS-T}}(x,\omega)\approx\frac{|G_T(k_0 c\sin(\psi))|^2}{(1+M_{OR})^2}\frac{\omega L\sin^2(\theta/2)\sin\psi}{\pi^2 c_0|x|^2}$$

$$\times\frac{\pi}{k_c}e^{-\omega h/(2U_c)}\times 4\cdot\left(\frac{\partial U_1}{\partial y_2}\frac{1}{\omega}\right)^2\rho^2$$

$$\times U_1^2\Phi_{u_2u_2}(\omega)l_3(\omega)/(\pi). \quad (67)$$

V. CONCLUSIONS

The dipole sound due to the diffraction of homogeneous turbulence by the leading edge of a thick foil was estimated with a rigid surface Green's function that accounted for the foil thickness and the acoustic back scattering associated with the acoustically noncompact chord length. The incident surface stagnation enthalpy was calculated with a volume integral incorporating the rigid plane Green's function and an equivalent form of the lifting dipole component of Howe's (1975) acoustic source that was derived in terms of the mean free stream velocity and the fluctuating up-wash velocity of Sears' (1941) analysis.

An estimate of the dipole radiation was made using the foil geometry and measured grid-generated turbulence statistics from Paterson and Amiet (1976, 1977). The predictions show good agreement with the measured dipole sound. This establishes that the square of the magnitude of Sears' function, $S(k_1,h)$, for a foil of thickness, h, can be expressed as

$$|S(k_1,h)|^2\cong|S(k_1)|^2\exp[-k_1 h/2], \quad (68)$$

where $S(k_1)$ is Sears' function with $h=0$.

The modeling approach requires that the incident sources obey a ''rapid distortion'' approximation. The RMS turbulence fluctuations of the validation data of Paterson and Amiet (1976, 1977) were approximately 4.5% of the free stream. This is well within the limits (less than 10%) established by Grace (2001) to ensure that the sources do not deviate from the potential flow. Leading edge flow separation may also cause the trajectories of the incident sources to deviate sufficiently from that modeled by the Green's function. This modeling approach also required the use of the frozen flow approximation that is strictly applicable to constant velocity flows in the vicinity of the diffraction zones. The success of the dipole sound predictions suggests that the frozen flow approximation may be applied to steady flows whose potential flow streamlines exhibit curvature and the consequential acceleration of the sources around an edge. Extensions of the modeling approach may be made to predict the affect of foil camber on the dipole sound by incorporating this effect into the velocity potential of the Green's function via a conformal mapping, as demonstrated by Martinez and Rudzynsky (1997) for the incompressible problem.

The dipole sound from a foil cutting through a mean shear layer was obtained using the acoustic source that was tailored for mean shear sources. Even though the shear layer is the by-product of a viscous flow, the acoustic analogy enables the sound associated with the diffraction of the mean shear source by a rigid surface to be predicted as if the distributed source was carried by a potential flow. The ratio of the dipole sound of the mean shear source to the source without mean shear was determined to be proportional to the square of the ratio of the mean shear to the frequency of the sources. This establishes a critical frequency below, which the dipole sound from the mean shear should dominate and above which the dipole sound from the source without mean shear should dominate. This result has been reported for the frequency dependence of the mean shear and non-mean shear quadrupole sources associated with jet noise (e.g., Moon and Zelazny, (1975)).

ACKNOWLEDGMENTS

This work was funded by Code 73R/93R, Naval Sea Systems Command.

LIST OF SYMBOLS

B	Stagnation enthalpy; see Eq. (2)
B_{rad}	Far field radiated stagnation enthalpy; see Eq. (33)
c	Foil chord length
c_0	Speed of sound
$dx_{1,3}$	$dx_1\,dx_3$
G	Green's function
G_T	Normalized Green's function; see Eqs. (41), (42), and (43)

h	Foil maximum thickness
k_c	Convection wave number, ω/U_c
k_0	Acoustic wave number
$k_{1,3}$	(k_1,k_3)
$k_{1,2,3}$	(k_1,k_2,k_3)
L	Wetted span length of foil; see Fig. 3
l_j	Correlation scale in direction j; see Eqs. (22) and (23)
M_c	Convection Mach number, U_c/c_0
p	Acoustic pressure; see Eq. (47)
p_{rad}	Far field acoustic pressure, see Eqs. (34) and (35)
r	See Fig. 3
R_{22}	Up-wash velocity correlation coefficient; see Eqs. (59) and (62)
$S(\ )$	Sears' function
t	Time
$-U_1$	Mean free stream flow velocity
U_c	Convection velocity
u_j	Fluctuating velocity in direction j
$x_{1,3}$	(x_1,x_3)
x	(x_1,x_2,x_3)
y	(y_1,y_2,y_3)
$(\ )^*$	Complex conjugate of $(\ )$
$\delta(\ )$	Dirac delta function
$\delta^n(\ )$	nth order derivative of Dirac delta function
$\phi_j(k_j)$	jth component of incident wave number spectrum; see Eqs. (22) and (23)
Φ_{BB}	Stagnation enthalpy spectrum; see Eq. (17)
Φ_{pp}^{rad}	Far field acoustic pressure spectrum
Φ_{pp}^{surf}	Incident blocked surface pressure spectrum
$\Phi_{u_2u_2}$	Up-wash velocity spectrum; see Eq. (19)
$\Phi_{pp}^{\text{MS-T}}$	Mean shear-turbulence surface pressure spectrum; see Eq. (66)
$\Phi_{pp}^{\text{T-T}}$	Turbulence–turbulence surface pressure spectrum; see Eq. (66)
$\Phi_{pp}^{\text{rad,MS-T}}$	Radiated sound from mean shear-turbulence source; see Eq. (67)
$\gamma(k)$	See Eq. (6)
λ	Acoustic wave length
λ_f	Isotropic turbulence integral scale
ρ	Fluid density
ψ	See Fig. 3
θ	See Fig. 3
Ω	Vorticity
ω	Frequency (radians)

List of Abbreviations

BVI	Blade vortex interaction
MS-T	Mean shear-turbulent source
TBL	Turbulent boundary layer
T-T	Source without mean shear

Adamczyk, J. J. (**1974**). ''The passage of an infinite swept airfoil through an oblique gust,'' NASA CR-2395.

Amiet, R. K. (**1975**). ''Acoustic radiation from an airfoil in a turbulent stream,'' J. Sound Vib. **41**, 407–420.

Blake, W. K. (**1986**). Mechanics of Flow-Induced Sound and Vibration (Academic, New York), Vols. I and II.

Blake, W. K. (**1971**). ''Turbulent velocity and pressure fields in boundary-layer flows over rough surfaces,'' *Proceedings of the Symposium on Turbulence in Liquids*, University of Missouri—Rolla.

Chase, D. M. (**1980**). ''Modeling of wavevector-frequency spectrum of turbulent boundary layer wall pressure,'' J. Sound Vib. **70**, 29–67.

Corcos, G. M. (**1963**). ''The resolution of pressure in turbulence,'' J. Acoust. Soc. Am. **35**, 192–199.

Ffowcs-Williams, J. E. (**1965**). ''Surface pressure fluctuations induced by boundary-layer flow at finite Mach number,'' J. Fluid Mech. **22**, 507–519.

Grace, S. M. (**2001**). ''Unsteady blade response: the BVI model versus the gust model,'' AIAA 2001-2209, 7th AIAA/CEAS Aeroacoustics Conference, 28–30 May, 2001/Maastricht, The Netherlands (2001).

Howe, M. S. (**1975**). ''Contributions to the theory of aerodynamic sound, with application to excess jet noise and the theory of the flute,'' J. Fluid Mech. **126**, 625–673.

Howe, M. S. (**1988**). ''The influence of surface rounding on trailing edge noise,'' J. Sound Vib. **126**, 503–523.

Howe, M. S. (**1989**). ''Structural and acoustic noise generated by a large-eddy break-up device,'' Proc. R. Soc. London, Ser. A **24**, 461–486.

Howe, M. S. (**1998a**). *Acoustics of Fluid-Structure Interactions*, 1st ed. (Cambridge University Press, Cambridge).

Howe, M. (**1999**). ''Trailing edge noise at low mach numbers,'' J. Sound Vib. **225**, 211–238.

Howe, M. (**2000**). ''Trailing edge noise at low mach numbers, Part 2: Attached and separated edge flows,'' J. Sound Vib. **234**, 761–775.

Howe, M. S. (**2001a**). ''Edge-source acoustic Green's function for an airfoil of arbitrary chord, with application to trailing-edge noise,'' Q. J. Mech. Appl. Math. **54**, 139–155.

Howe, M. S. (**2001b**). ''Unsteady lift and sound produced by an airfoil in a turbulent boundary layer,'' J Fluids Struct. **15**, 207–225.

Kraichnan, R. H. (**1956**). ''Pressure fluctuations in turbulent flow over a flat plate,'' J. Acoust. Soc. Am. **28**, 378–390.

Lighthill, M. J. (**1958**). *Introduction to Fourier Analysis and Generalized Functions* (Cambridge University Press, Cambridge).

Martinez, R., and Widnall, S. E. (**1980**). ''Unified aerodynamic-acoustic theory for a thin rectangular wing encountering a gust,'' AIAA J. **18**, 636–645.

Martinez, R., and Rudzinsky, J. (**1997**). ''Analytic evaluation of shape effects on blade-vortex interaction,'' CAA Report U-2466-402.14.

Meecham, W. C. (**1965**). ''Surface and volume sound from boundary layers,'' J. Acoust. Soc. Am. **37**, 516–522.

Moon, L. F., and Zelazny, S. W. (**1975**). ''Experimental and analytical study of jet noise modeling,'' AIAA J. **13**, 387–393.

Olsen, W., and Wagner, J. (**1982**). ''Effect of thickness on airfoil surface noise,'' AIAA 82-4068.

Paterson, R. W., and Amiet, R. K. (**1976**). ''Noise and surface pressure response of an airfoil to incident turbulence,'' NASA CR-2733.

Paterson, R. W., and Amiet, R. K. (**1977**). ''Noise and surface pressure response of an airfoil to incident turbulence,'' J. Aircr. **14**, 729–736.

Powell, A. (**1960**). ''Aerodynamic noise and the plane boundary,'' J. Acoust. Soc. Am. **32**, 982–990.

Roger, M, Moreau, S., and Wang, M. (**2002**). ''An analytical model for predicting airfoil self-noise using wall pressure statistics,'' Center for Turbulence Research Annual Briefs, NASA Ames and Stanford University, pp. 405–414.

Sears, W. R. (**1941**). ''Some aspects of non-stationary airfoil theory and its practical applications,'' J. Aeronaut. Sci. **8**, 104–108.

Method of evaluating dipole sound source in a finite computational domain

T. Takaishi[a] and M. Ikeda
Railway Technical Research Institute, Tokyo, 185-8540, Japan

C. Kato
Institute of Industrial Science, University of Tokyo, Tokyo, 153-8505, Japan

(Received 8 January 2004; revised 11 May 2004; accepted 18 May 2004)

Numerical prediction of dipole sound based on Lighthill–Curle's equation gives little information on the structure of sound sources. On the other hand, a hybrid method that combines the large eddy simulation (LES) and the compact Green's function proposed by Howe provides detailed information on the vortices in the flow that most contribute to the generation of sound. However, when the dipole sound is evaluated from the momentum change in fluid inside a finite computational domain, the result does not in general agree with the sound evaluated from the fluctuating pressure on the body surface because contribution from vortices outside the computational domain is not taken into account. In this study, the balance of momentum in a finite computational domain is considered strictly, and the effect of outer vortices is replaced with contribution from inner properties by using an imaginary velocity potential φ_i. This process avoids sudden termination of Lighthill's stress tensor at the outer boundary and extracts the net contribution from dipole sound sources. With this new method, this paper simulates dipole sound sources around pantograph horns numerically, and clarifies how strong dipole sound sources are formed in the shear layer close to the model surface and how the flow through holes plays an important role in cancellation of sound sources. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1776189]

PACS numbers: 43.28.Ra, 43.50.Nm, 43.50.Lj [MSH] Pages: 1427–1435

I. INTRODUCTION

When aiming at harmony between the environment and technology it is a crucial issue to reduce the noise generated by the unsteady motion of fluid. Noise generated aerodynamically has become a dominant noise source for high-speed trains that are currently running at a maximum speed of 300 km/h (190 mile/h) in Japan. This is because (1) aerodynamic noise, in general, increases drastically with the increasing wind (vehicle) speed, and (2) other noises, such as a rolling noise from wheels and rails, and spark noise due to contact breaks between overhead line equipment and pantograph, have been substantially reduced during the past decades.

Although experiments in a low-noise wind tunnel are an effective way for examining countermeasures, various numerical techniques for the prediction of aeroacoustic noise have been proposed for understanding and reducing aeroacoustic sources as a result of the remarkable improvement in a computer performance in recent years. The direct Navier–Stokes simulation (DNS) with high-order accuracy including compressibility can capture a flow field and a sound field simultaneously, and will be a promising technique in the future.[1] Hardin *et al.*[2] presented another method which obtains the time-dependent incompressible flow first, and then solves the compressible Navier–Stokes equation for propagation of acoustic waves.

Since the Mach number of flow around trains and cars is low enough, many of the former studies[3,4] (especially in Japan) calculated the pressure fluctuations at the body surface and then estimated the acoustic pressure fluctuations in the far field by Curle's equation.[5] However, they could not identify the structure of sound sources in the flow directly, and the relation between flow and radiated sound is not clear. Some of the studies tried to evaluate the distribution of the Lighthill stress tensor in the flow as sound sources, but this is not a suitable way to estimate the dipole sound sources because it does not include the effect of acoustical scattering at the body surface.

On the other hand, Hardin *et al.*[6] calculated the sound radiation from a cylinder with the compact Green's function proposed by Howe.[7,8] The present authors[9] also clarified the distribution of dipole sound sources at the bogie section of high-speed trains by coupling the instantaneous flow properties obtained from the large eddy simulation (LES) technique with the compact Green's function which is determined numerically. In a sufficiently large domain containing all vortices, dipole sound obtained by this method coincides mathematically with the surface integral of Curle's equation. However, when the dipole sound is evaluated from the momentum change in fluid inside a finite computational domain, the result does not in general agree with Curle's equation. Hardin[10] pointed out that the Lighthill stress terms are still significant at the outflow boundary, and their sudden termination is known to cause strong, spurious acoustic sources. Some correction procedures have been proposed,[11,12] but most of them are concerned merely with the fluid dynamics and do not consider acoustical aspects of the problem.

In this paper, a new method for evaluating dipole sound sources in a finite computational domain is proposed. In Sec.

[a]Electronic mail: takaishi@rtri.or.jp

II, formulas from Howe's vortex theory are summarized first, and then the balance of momentum in a finite computational domain is considered strictly. Though this process is based on fluid dynamics, an interpretation from an acoustical viewpoint is also presented. In Sec. III, the noise reduction mechanism of pantograph horns is clarified numerically in order to verify the superiority of the newly presented method to the former ones.

II. FORMULAS FOR EVALUATION

A. Vortex sound theory

The theory of aerodynamic sound was first developed by Lighthill,[13] who reformulated the Navier–Stokes equation into an exact, inhomogeneous wave equation. Howe[7,8,14] simplified Lighthill's equation for high Reynolds number flow at low Mach number,

$$\frac{1}{c_0^2}\frac{\partial^2 B}{\partial t^2}-\nabla^2 B=\mathrm{div}(\boldsymbol{\omega}\times\mathbf{u}), \tag{1}$$

where B is the total enthalpy,

$$B\equiv\frac{p}{\rho_0}+\frac{1}{2}|\mathbf{u}|^2, \tag{2}$$

and $\boldsymbol{\omega}=\nabla\times\mathbf{u}$ is the vorticity. Consider the fluid domain Ω_{all} whose inner boundary is the body surface S and outer boundary is the surface Σ_{inf} at infinity, which is large enough to contain all of vorticity and the flow is uniform there. In this study, the body surface is assumed to be fixed. Then the velocity at S should vanish. If the observer $\mathbf{x}$ is far enough from the source region, Green's theorem yields the sound pressure at $\mathbf{x}$:

$$p_a(\mathbf{x},t)=-\rho_0\int_{\Omega_{all}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla G(\mathbf{x},\mathbf{y},t-\tau)d\mathbf{y}d\tau, \tag{3}$$

where the Green's function $G(\mathbf{x},\mathbf{y},t-\tau)$ is chosen to satisfy

$$\frac{1}{c_0^2}\frac{\partial^2 G}{\partial t^2}-\nabla^2 G=\delta(\mathbf{x}-\mathbf{y})\,\delta(t-\tau) \tag{4}$$

in the fluid and its normal derivative is 0 on S, namely,

$$\frac{\partial G}{\partial n}=0 \quad on\ \ S. \tag{5}$$

If the characteristic length of the source region is small compared to the wavelengths of the sound waves, differences of retarded time in the source region are negligible, and the source is said to be acoustically compact. Then the three-dimensional Green's function may be expanded around the source region in the form[7,8]

$$G_c\cong\frac{1}{4\pi|\mathbf{x}|}\,\delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right)+\frac{\mathbf{x}\cdot\mathbf{Y}}{4\pi c_0|\mathbf{x}|^2}\,\frac{\partial}{\partial t}\,\delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right)+\frac{1}{|\mathbf{x}|}O\left[\left(\frac{2\pi l}{\lambda}\right)^2\right]. \tag{6}$$

In this equation, Y_i denotes the velocity potential of an imaginary flow around the body that has unit speed in the i direction at large distances from the body. This velocity potential must satisfy Laplace's equation

$$\nabla^2 Y_i=0 \tag{7}$$

in the fluid and also

$$\frac{\partial Y_i(\mathbf{y})}{\partial n}=0 \tag{8}$$

on surfaces. Note that the Green's function governs the acoustical properties of the system: the velocity potential Y_i depends on the body shape, but does not represent a real flow. Substituting Eq. (6) into (3), we obtain

$$p_a(\mathbf{x},t)=-\frac{\rho_0 x_i}{4\pi c_0|\mathbf{x}|^2}\times\int_{\Omega_{all}}\frac{\partial}{\partial t}(\boldsymbol{\omega}\times\mathbf{u})(\mathbf{y},t-|\mathbf{x}|/c_0)\cdot\nabla Y_i\,d\mathbf{y}. \tag{9}$$

The force F_i exerted on an incompressible inviscid (or high Reynolds number) fluid by a body is generally given as

$$F_i=-\rho_0\int_{\Omega_{all}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla Y_i\,d\mathbf{y}=\int_S p n_i\,dS, \tag{10}$$

where the unit normal vector n_i on S is directed into the fluid domain Ω_{all}. By using the relation (10), Eq. (9) is reformulated into the form

$$p_a(\mathbf{x},t)=\frac{x_i}{4\pi c_0|\mathbf{x}|^2}\int_S\frac{\partial}{\partial t}p(\mathbf{y},t-|\mathbf{x}|/c_0)n_i\,dS, \tag{11}$$

which is identical to the dipole sound term in Curle's equation.[5]

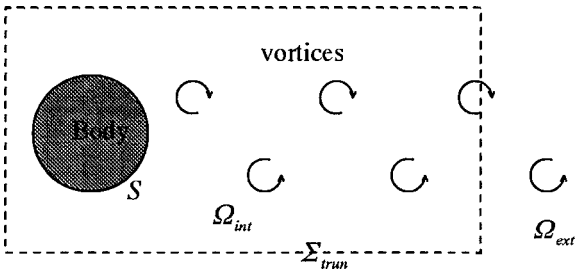


FIG. 1. Schematic diagram of a finite computational domain.

B. Evaluation in a finite computational domain

Formulas in the previous section hold strictly under assumptions that are imposed during the derivation. If the fluid velocity $\mathbf{u}$ and vorticity $\boldsymbol{\omega}$ are given by a numerical simulation, however, the computational domain is inevitably finite and the flow is not necessarily uniform at the outer boundary.

Consider a domain as shown in Fig. 1. A finite computational domain Ω_{int} is surrounded by the artificial outer boundary Σ_{trun}. The fluid domain outside Σ_{trun} is named Ω_{ext}. The relationship among fluid domains is

$$\Omega_{all}\equiv\Omega_{int}\cup\Omega_{ext}. \tag{12}$$

In the previous section, the velocity potential Y_i which satisfies Eqs. (7) and (8) is adopted. If we define another

velocity potential φ_i that would be produced by translational motion of S as a rigid body at unit speed in the i direction, the relationship between Y_i and φ_i is

$$Y_i = y_i - \varphi_i, \tag{13}$$

where y_i is the coordinate with respect to the sound source. Since the Green's function governs the acoustical properties of the system, the velocity potential Y_i and φ_i depend on the body shape, but they do not represent a real flow around the body, as mentioned before.

Though the real turbulent flow is defined to be uniform at Σ_{inf}, it is not necessarily satisfied at Σ_{trun}. This implies that Ω_{int} does not contain all vortices, and Ω_{ext} also contains some. In spite of this, Σ_{trun} must be far enough from the body so that φ_i and $\nabla\varphi_i$ can be regarded as 0 on it.

Equation (10) yields

$$\begin{aligned} F_i &= \int_S p n_i \, dS = -\rho_0 \int_{\Omega_{int}\cup\Omega_{ext}} \nabla Y_i \cdot (\boldsymbol{\omega}\times\mathbf{u}) d\mathbf{y} \\ &= -\rho_0 \int_{\Omega_{int}} \nabla Y_i \cdot (\boldsymbol{\omega}\times\mathbf{u}) d\mathbf{y} - \rho_0 \int_{\Omega_{ext}} \nabla Y_i \cdot (\boldsymbol{\omega}\times\mathbf{u}) d\mathbf{y}. \end{aligned} \tag{14}$$

This implies that the second term of Eq. (14) must be considered if Eq. (9) is evaluated in the finite domain. This additional term is denoted by δF_i hereafter. Since the Mach number of the flow is small enough to neglect the compressibility, we can write

$$(\boldsymbol{\omega}\times\mathbf{u})_i = \frac{\partial}{\partial y_j}(u_i u_j) - \nabla_i\left(\frac{1}{2}u^2\right). \tag{15}$$

The divergence theorem in Ω_{all} supplies

$$\begin{aligned} \int_{\Omega_{all}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} &= -\int_S \left(\frac{1}{2}u^2\delta_{ij} - u_i u_j\right) n_j \, dS \\ &\quad - \int_{\Sigma_{inf}} \left(\frac{1}{2}u^2\delta_{ij} - u_i u_j\right) n_j \, dS. \end{aligned} \tag{16}$$

By definition, the velocity at S should be 0 and the flow at Σ_{inf} is uniform, so that each term of the right-hand side is 0. As a result, Eq. (16) becomes

$$\begin{aligned} 0 &= \int_{\Omega_{all}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} \\ &= \int_{\Omega_{int}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} + \int_{\Omega_{ext}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} \\ \therefore \int_{\Omega_{ext}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} &= -\int_{\Omega_{int}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} \end{aligned} \tag{17}$$

and

$$\delta F_i \cong -\rho_0 \int_{\Omega_{ext}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} = \rho_0 \int_{\Omega_{int}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y}. \tag{18}$$

Then Eq. (14) is reformulated into

$$\begin{aligned} F_i &\cong -\rho_0 \int_{\Omega_{int}} \nabla Y_i \cdot (\boldsymbol{\omega}\times\mathbf{u}) d\mathbf{y} + \rho_0 \int_{\Omega_{int}} (\boldsymbol{\omega}\times\mathbf{u})_i \, d\mathbf{y} \\ &= \rho_0 \int_{\Omega_{int}} \nabla\varphi_i \cdot (\boldsymbol{\omega}\times\mathbf{u}) d\mathbf{y}. \end{aligned} \tag{19}$$

Finally, the dipole sound is evaluated in the form

$$\begin{aligned} p_a(\mathbf{x},t) &= \frac{\rho_0 x_i}{4\pi c_0 |\mathbf{x}|^2} \\ &\quad \times \int_{\Omega_{int}} \frac{\partial}{\partial t} (\boldsymbol{\omega}\times\mathbf{u})(\mathbf{y}, t - |\mathbf{x}|/c_0) \cdot \nabla\varphi_i \, d\mathbf{y}. \end{aligned} \tag{20}$$

C. Interpretation from an acoustical viewpoint

In the previous section, the effect of outer vortices is replaced with contribution from properties inside the finite computational domain. Although the derivation of Eq. (20) is based on fluid-dynamical process, interpretation from an acoustical viewpoint will be presented in this section.

A free-space Green's function $G_0(\mathbf{x},\mathbf{y},t-\tau)$ is expressed as

$$G_0(\mathbf{x},\mathbf{y},t-\tau) = \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\,\delta\left(t-\tau-\frac{|\mathbf{x}-\mathbf{y}|}{c_0}\right). \tag{21}$$

If the observer $\mathbf{x}$ is far enough from the source region, Eq. (21) is expanded

$$\begin{aligned} G_0(\mathbf{x},\mathbf{y},t-\tau) &\cong \frac{1}{4\pi|\mathbf{x}|}\,\delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right) \\ &\quad + \frac{\mathbf{x}\cdot\mathbf{y}}{4\pi c_0|\mathbf{x}|^2}\,\frac{\partial}{\partial t}\,\delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right). \end{aligned} \tag{22}$$

From Eqs. (6) and (13), the compact Green's function is rewritten as

$$\begin{aligned} G_c(\mathbf{x},\mathbf{y},t-\tau) &\cong \frac{1}{4\pi|\mathbf{x}|}\,\delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right) + \frac{\mathbf{x}\cdot\mathbf{y}}{4\pi c_0|\mathbf{x}|^2}\,\frac{\partial}{\partial t} \\ &\quad \times \delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right) - \frac{\mathbf{x}\cdot\varphi}{4\pi c_0|\mathbf{x}|^2}\,\frac{\partial}{\partial t} \\ &\quad \times \delta\left(t-\tau-\frac{|\mathbf{x}|}{c_0}\right). \end{aligned} \tag{23}$$

Compared with Eq. (22), the first and second terms of Eq. (23) denote the direct sound from the turbulent flow, and the third term denotes the effect of scattering at the body surface. Substituting Eq. (23) into Eq. (3), we obtain

$$\begin{aligned} p_a(\mathbf{x},t) &= -\frac{\rho_0 x_i}{4\pi c_0|\mathbf{x}|^2} \int_{\Omega_{all}} \frac{\partial}{\partial t} (\boldsymbol{\omega}\times\mathbf{u})_i (\mathbf{y}, t - |\mathbf{x}|/c_0) d\mathbf{y} \\ &\quad + \frac{\rho_0 x_i}{4\pi c_0|\mathbf{x}|^2} \int_{\Omega_{all}} \frac{\partial}{\partial t} (\boldsymbol{\omega}\times\mathbf{u}) \\ &\quad \times (\mathbf{y}, t - |\mathbf{x}|/c_0) \cdot \nabla\varphi_i \, d\mathbf{y}. \end{aligned} \tag{24}$$

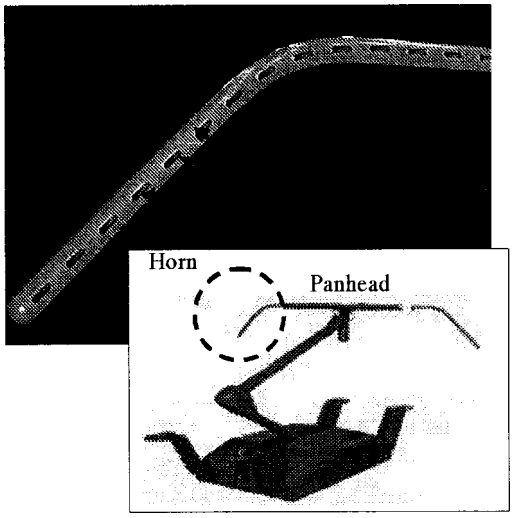


FIG. 2. Low noise pantograph with horns. Horns are installed at both side-ends of a panhead, and holes are drilled on them periodically.

In the case of aeroacoustic noise from the low Mach number flow, the direct sound from the turbulence flow [the first term of Eq. (24)] is quadrupole, and the scattered sound (the second term) is dipole. From Eq. (17), the first term of Eq. (24) becomes order 0 compared to the second term, which coincides with the well-known fact that the acoustical efficiency of quadrupole is less than that of dipole in the case of low Mach number flow. Since $\nabla\varphi_i$ can be regarded as 0 outside Σ_{trun}, the second term of Eq. (24) finally coincides with Eq. (20).

This acoustical interpretation leads to the conclusion that the newly proposed method will succeed in avoiding sudden termination of Lighthill's stress tensor at the outer boundary and extracting the net contribution from dipole sound sources. However, accurate calculation of the boundary layer and bounded vortices is indispensable in order to evaluate Eq. (20) properly. In the following sections, these will be confirmed numerically.

III. EXAMPLE OF APPLICATION

In this section, the noise reduction mechanism of pantograph horns with periodic holes is clarified numerically by using the method proposed in Sec. II. In order to validate the numerical results, an experiment was performed in a small-scale low-noise wind tunnel at the Railway Technical Research Institute (RTRI).[15]

A. Summary of calculation

Since the Environmental Quality Standards for high-speed trains established by the Japanese government require that the peak noise level must not exceed 75 dB(A), it is a crucial issue to reduce the aerodynamic noise that is radiated from electrical devices on the roof of the train, such as pantographs and cable heads. In this paper, we direct attention to the sound from pantograph horns shown in Fig. 2, that are installed at both side-ends of a panhead in order to ensure movement onto the contact strip of out-of-running wires. Periodic holes in the cylindrical horns are already known to reduce the noise level, but the detailed mechanism of noise reduction is not clear yet.

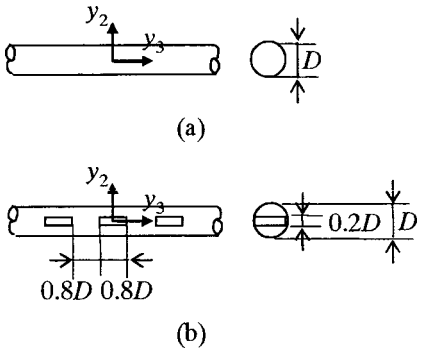


FIG. 3. Design of horn models: (a) simple cylinder and (b) cylinder with periodic holes.

Though actual horns are bent at the middle as shown in Fig. 2, straight models are used here in order to simplify the phenomenon. Two types of horn model, a simple cylinder and a cylinder with periodic holes, are designed as shown in Fig. 3. Figure 4 shows a schematic diagram of the computational domain, that should be defined as Ω_{int}. The center of each model is set as the origin of the coordinates, main flow direction as the y_1 axis, side direction as the y_2 axis, and spanwise direction as the y_3 axis. In case of the cylinder with periodic holes, the plane $y_3=0$ cuts across the center of a hole. The Reynolds number based on the diameter of models D and the uniform flow velocity U_∞ is 1.1×10^4. The unsteady flow around the model is obtained by using the large eddy simulation (LES), which solves the spatially filtered equations of continuity and momentum for incompressible flow. In this study, the governing equations for energy-containing eddies are converted to a system of algebraic equations by the finite element method with structured cells, and both time and space derivatives are discretized using the second-order upwind methods.[3] The dynamic Smagorinsky model is adopted for the subgrid-scale eddy viscosity.[16,17]

A uniform flow U_∞ without turbulent disturbance is assumed for the inlet boundary. At the upper and lower boundaries, slip (symmetry) condition is applied, and the no-slip condition for velocity is applied at the model surface. At the

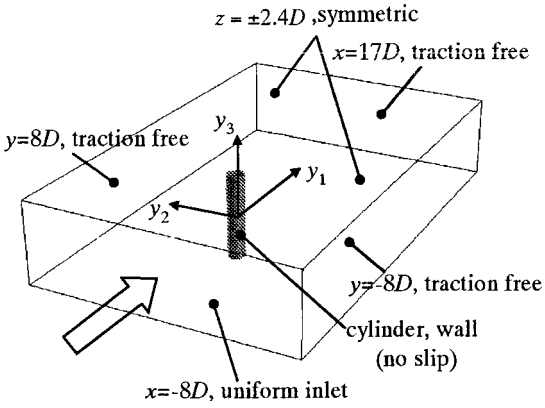


FIG. 4. Schematic diagram of the computational domain around a horn model. This domain corresponds to Ω_{int} in Fig. 1 and outer boundary also corresponds to Σ_{trun}.

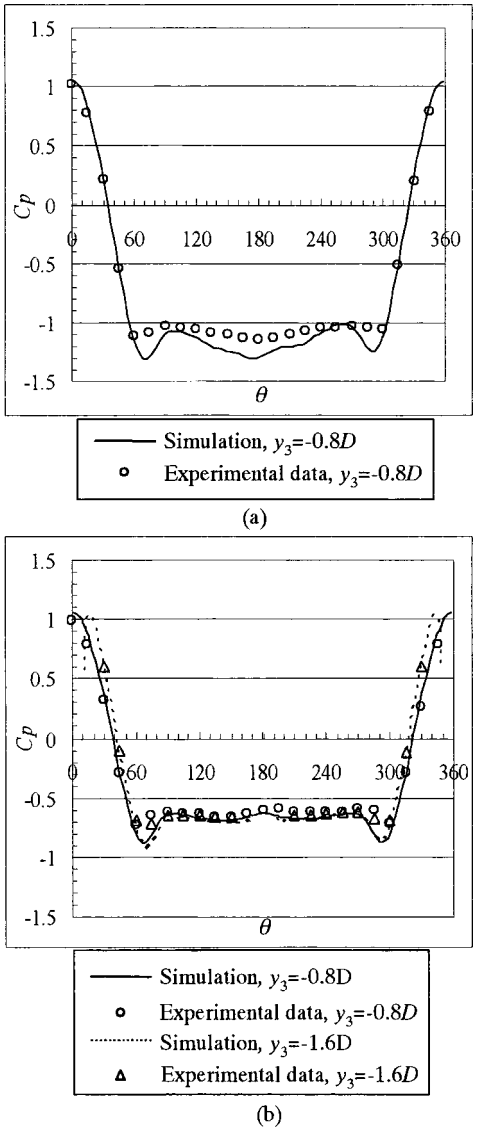


FIG. 5. Profile of time-averaged pressure at the model surface: (a) simple cylinder and (b) cylinder with periodic holes. Point marks in each graph show the experimental data (Ref. 15).

remainder, the traction-free condition defined as

$$n_j\left(-\delta_{ij}\frac{\overline{p}}{\rho}+(\nu+\nu_{SGS})\left(\frac{\partial\overline{u_i}}{\partial y_j}+\frac{\partial\overline{u_j}}{\partial y_i}\right)\right)=0 \tag{25}$$

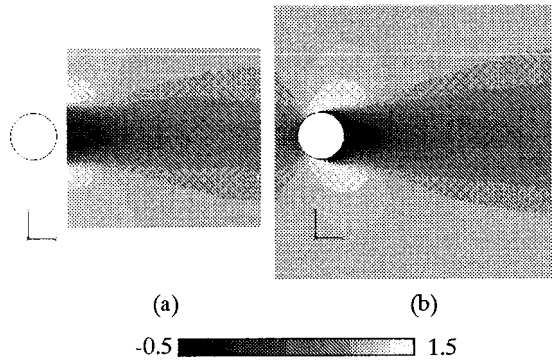


FIG. 6. Distribution of mean velocity in the y_1 direction ($\overline{u_1}/U_\infty$) of flow behind the simple cylinder at $y_3=0$: (a) experimental data (Ref. 15) measured by PIV and (b) present numerical result.

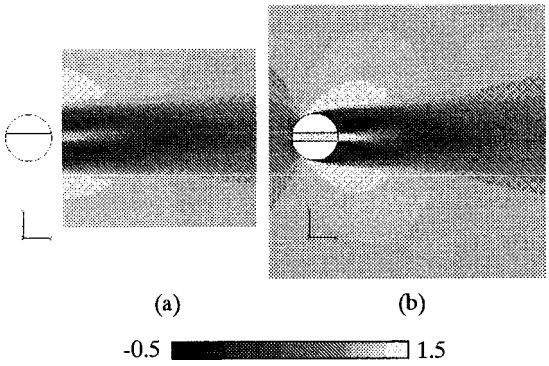


FIG. 7. Distribution of mean velocity in the y_1 direction ($\overline{u_1}/U_\infty$) of flow behind the cylinder with periodic holes at $y_3=0$: (a) experimental data (Ref. 15) measured by PIV and (b) present numerical result.

is applied. The grid size in the radius direction close to the model surface is set $0.005D$, which should be small enough to capture surface vorticity.

The imaginary velocity potential φ_i is also obtained numerically by solving Laplace's equation (7) for Y_i with boundary condition (8) first and then using the relation (13). At the outer boundary Σ_{trun}, the following condition is applied

$$\begin{aligned} &Y_i=y_i,\\ &\varphi_i=0. \end{aligned} \tag{26}$$

The accuracy of the numerical calculation of φ_i subject to the artificial boundary condition (26) is carefully examined. In the case of a long cylinder, there exists the analytical solution

$$\begin{aligned} \varphi_1&=-\frac{y_1}{y_1^2+y_2^2}\left(\frac{D}{2}\right)^2,\\ \varphi_2&=-\frac{y_2}{y_1^2+y_2^2}\left(\frac{D}{2}\right)^2,\\ \varphi_3&=0. \end{aligned} \tag{27}$$

Equation (20) is evaluated by using both numerical and analytical solution of φ_i, and the error is confirmed to be negligible.

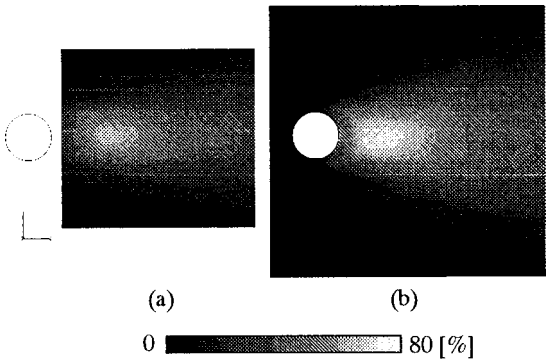


FIG. 8. Distribution of turbulence level in the y_2 direction ($\sqrt{\overline{u_2^2}}/U_\infty$) of flow behind the simple cylinder at $y_3=0$: (a) experimental data (Ref. 15) measured by PIV and (b) present numerical result.

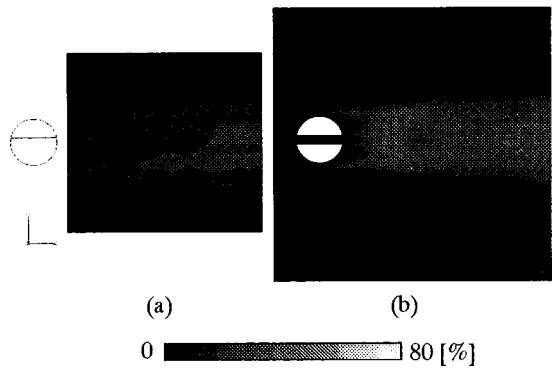


FIG. 9. Distribution of turbulence level in the y_2 direction ($\sqrt{u_2^2}/U_\infty$) of flow behind the cylinder with periodic holes at $y_3=0$: (a) experimental data (Ref. 15) measured by PIV and (b) present numerical result.

B. Results of flow simulation

Figure 5 shows the profile of time-averaged pressure at the model surface. Point marks in each graph show the experimental data.[15] The vertical axis denotes a pressure coefficient C_p which is divided by the dynamic pressure of the uniform flow, and the horizontal axis denotes an anticlockwise angle from the x axis (or the front stagnation point). The base pressure of the cylinder with periodic holes takes higher values than that of the simple cylinder. This means that the separation velocity of the simple cylinder is higher, which will release stronger vortices into the wake. The pressure profile of the cylinder with periodic holes at the cross section which contains a hole ($y_3=-1.6D$) is almost the same as that at the cross section without a hole ($y_3=-0.8D$). This implies that spanwise distribution of time-averaged surface pressure is not sensitive in spite of intermittent holes.

Figure 6 shows the mean velocity field in the y_1 direction of flow behind the simple cylinder. The left one shows the experimental data[15] measured by the particle image velocimetry (PIV) and the right shows the present numerical result. Figure 7 also shows the mean velocity field of flow behind the cylinder with periodic holes at the cross section which contains a hole ($y_3=0$). In both cases, the numerical result proves to agree well with the experimental results.

Figures 8 and 9 show the distribution of turbulence level in the y_2-direction at the same cross section as Figs. 6 and 7. The turbulence level behind the simple cylinder takes the highest value at around $y_1=2.5D$ along the y_2 axis. Vortices of opposite sign are released into the wake alternately from each side of the cylinder, and passage of these vortices changes the sign of the y_2 component of the velocity vector around $y_1=2.5D$ periodically. On the other hand, a flow through the hole is injected into the dead water region behind the cylinder with periodic holes and seems to suppress the formation of strong vortices. As a result, there are no regions of high turbulence level.

From these comparisons, flow simulation is proved to have good agreement with the experimental results. Though it is difficult to ensure that unsteady motion of the thin boundary layer and bounded vortices are evaluated with accuracy because there is no way to measure them directly in experiments, the following facts should be noted;

(i) There are six grid points in the boundary layer at 80° where the flow separates, and they must be enough to capture the bound vortices.

(ii) Numerical results of time-averaged pressure at the model surface, mean velocity field and turbulence intensity at the wake agree well with the experimental results. These imply that the Navier–Stokes equations are solved properly in the whole fluid region including the boundary layer.

C. Estimation of dipole sound source

In the previous sections, the instantaneous flow properties around models are solved by the LES technique, and φ_i is also obtained by solving Laplace's equation numerically. Hence, in this section, the distribution of dipole sound sources is predicted numerically.

Figure 10 shows the time histories of the nondimensional force exerted on a fluid by each model. Since these are evaluated in the finite computational domain Ω_{int} shown in

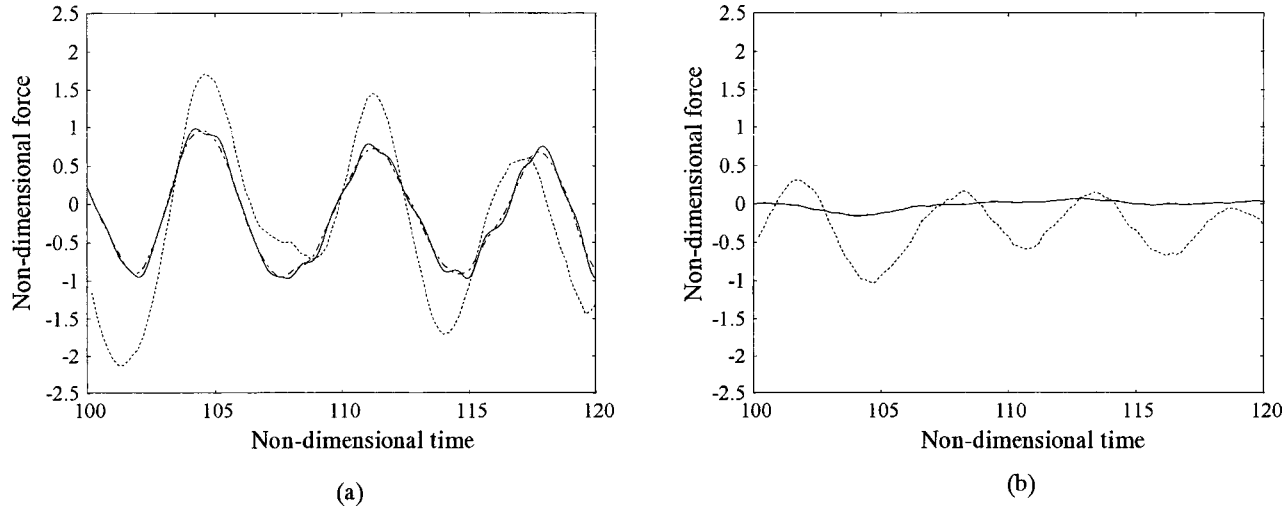


FIG. 10. Time histories of the nondimensional force exerted on a fluid by each model: (a) simple cylinder and (b) cylinder with periodic holes. Solid line: $\int_s pn_2\, dS/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$, dotted line: $-\rho_0\int_{\Omega_{int}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla Y_2\, d\mathbf{y}/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$ and dashdot line: $\rho_0\int_{\Omega_{int}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla\varphi_2\, d\mathbf{y}/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$.

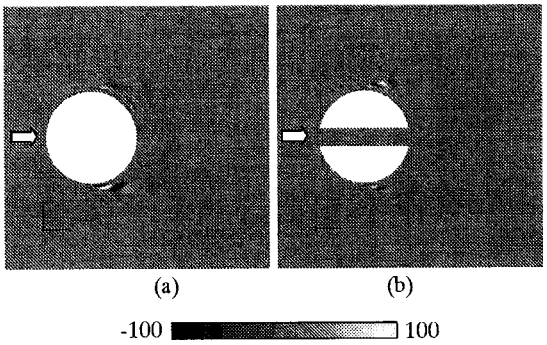


FIG. 11. Distribution of dipole sound sources [expressed as the integrand of Eq. (20)] at $y_3=0$: (a) simple cylinder and (b) cylinder with periodic holes.

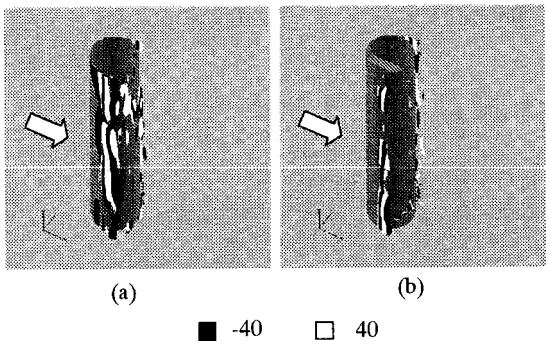


FIG. 13. Three-dimensional distribution of dipole sound sources: (a) simple cylinder and (b) cylinder with periodic holes.

Fig. 4, the volume integral with Y_2 does not agree with the surface integral of the pressure, because the flow is not uniform at the outer boundary. To make matters worse, artificial oscillation occurs in the case of the cylinder with periodic holes. This would bring serious misunderstanding on the phenomenon. On the other hand, the volume integral with φ_2 is confirmed to agree well with the surface integral at both cases.

Figure 11 shows the distribution of the time derivative of the instantaneous force at the same cross section as Figs. 6–9. Since the volume integral of these parts, namely Eq. (20), yields the acoustic pressure fluctuation of the dipole sound in the far field, the integrand in (20) can be regarded as a dipole sound source. The results show that strong dipole sound sources are formed in the shear layer close to the model surface. It is a remarkable result that there are no strong dipole sound sources in the wake of each model, where turbulence level takes the highest value as shown in Figs. 8 and 9.

In order to estimate contribution of bound vortices to the force, Eq. (19) is integrated over the thin boundary region satisfying $\sqrt{y_1^2+y_2^2}\leqslant 0.51D$. Figure 12 shows that this plays a prominent role in the force calculation. Howe *et al.*[18] modeled vortex shedding from a sphere as a succession of vortex rings with the ring plane inclined at a random angle to the mean flow direction. Each ring is gradually shed from the body surface and the force exhibits great fluctuation during this process. Their result also proved that even though there is a strong region of vorticity in the wake within a diameter or so from the body surface, the influence of that vortex on the force is practically negligible, that also happens in Figs. 11 and 12 in the present study. Though the time derivative of $\boldsymbol{\omega}\times\mathbf{u}$ also takes large values there, the imaginary velocity potential φ_2 and its gradient decrease rapidly as the distance from the model gets large. This can be explained easily from the acoustical discussion in Sec. II C. Since φ_i represents the effect of scattering at the body surface, it is natural that the unsteady motion of the flow near the surface plays an important role. When using φ_i instead of Y_i in the evaluation of the force, it is important to be able to resolve the flow within the boundary layer exactly, because the force is generated in the process the bound vortices due to the viscosity shed into the fluid from the body surface. On the other hand, the use of Y_i suppresses the contribution from the surface because Eq. (8) holds there. The use of Y_i avoids errors due to possible uncertainty in the calculation of the boundary layer, and make it easier to model the flow by vortices in the wake. But this also causes the crude error due to the sudden truncation of the vortices at the finite outer boundary Σ_{trun}.

Figure 13 shows three-dimensional distributions of di-

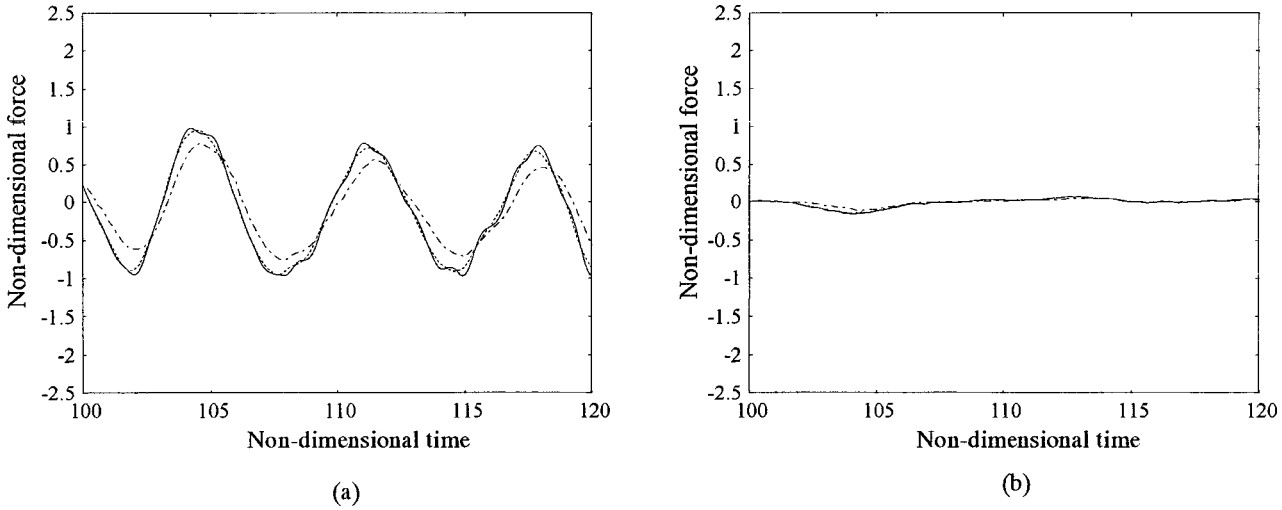
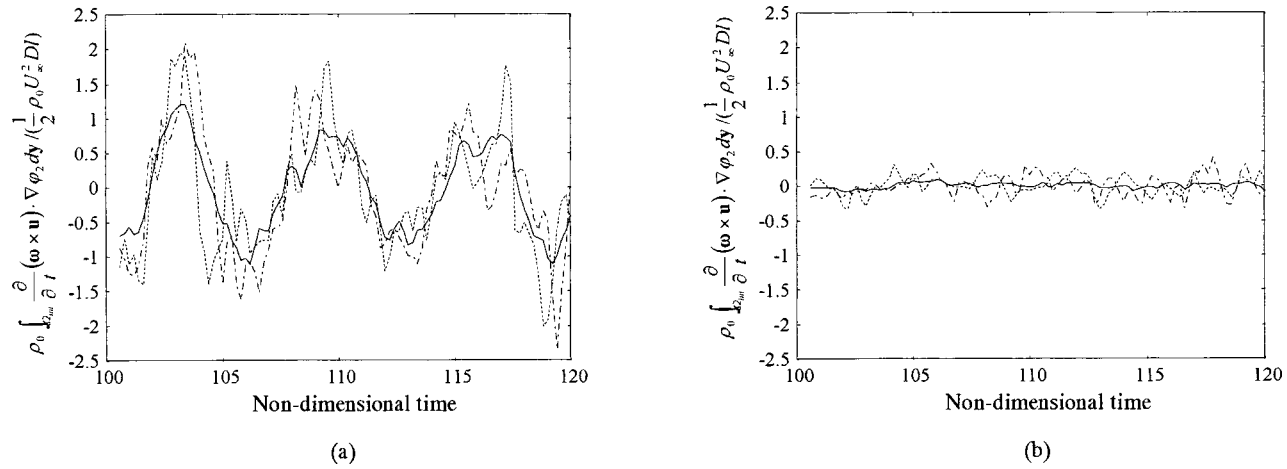


FIG. 12. Contribution of the bound vortices to the nondimensional force: (a) simple cylinder and (b) cylinder with periodic holes. Solid line: $\int_s pn_2\,dS/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$, dotted line: $\rho_0\int_{\Omega_{int}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla\varphi_2\,d\mathbf{y}/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$ and dashdot line: $\rho_0\int_{\Omega_{\sqrt{y_1^2+y_2^2}\leqslant 0.51D}}(\boldsymbol{\omega}\times\mathbf{u})\cdot\nabla\varphi_2\,d\mathbf{y}/(\frac{1}{2}\rho_0 U_\infty^2 4.8D^2)$.



(a) (b)

FIG. 14. Effect of integral in the spanwise direction: (a) simple cylinder and (b) cylinder with periodic holes. The vertical axis denotes $\rho_0 \int_{\Omega_{int}} (\partial/\partial t) \times (\boldsymbol{\omega} \times \mathbf{u}) \cdot \nabla \varphi_2 \, d\mathbf{y} / (\frac{1}{2} \rho_0 U_\infty^2 Dl)$, where l is the spanwise length of each integral volume for normalization. Solid line: $-2.4D \leqslant y_3 \leqslant 2.4D$, dotted line: $-0.1D \leqslant y_3 \leqslant 0.1D$ and dashdot line: $0.7D \leqslant y_3 \leqslant 0.9D$.

pole sound sources around each model. Dipole sound sources around the simple cylinder have large two-dimensional structure in the spanwise direction, but the intermittent jet flow through holes seems to collapse it in case of the cylinder with periodic holes. Figure 14 shows the effect of integral in the spanwise direction. Since dipole sound sources around the simple cylinder have large two-dimensional structure in the spanwise direction, the integral at each cross section will be summed. It is interesting that the volume integral at each cross section of the cylinder with periodic holes already takes small value, and cancellation in the spanwise direction is not necessarily dominant. This means that dipole sound sources are mainly suppressed due to momentum injection through holes and the effect of periodicity is not important, and also implies that dipole sound sources in Fig. 11(b) would be canceled when integrated, though it is hard to distinguish from Fig. 11(a) those that would not be canceled.

IV. CONCLUSION

In this paper, a new method to evaluate dipole sound sources in a finite computational domain is proposed. As a result, the following major conclusions are obtained.

(1) The balance of momentum in a finite computational domain is considered strictly, and the effect of outer vortices is replaced with contribution from inner properties by using an imaginary velocity potential φ_i. From an acoustical viewpoint, this process avoids sudden termination of Lighthill's stress tensor at the outer boundary and extracts the net contribution from dipole sound sources.

(2) When using φ_i instead of Y_i, it is important to be able to resolve the flow within the boundary layer exactly, because the principal component of the force is generated by the bound vorticity produced in response to the shedding of vorticity into the fluid from the body surface. On the other hand, the use of Y_i suppresses the contribution from the surface. This avoids possible uncertainties in the calculation of the boundary layer, and makes it easier to model the flow by vortices in the wake. But this also causes the crude error due to the sudden truncation of the vortices at the finite outer boundary.

(3) The noise reduction mechanism of pantograph horns with periodic holes is clarified by the numerical method. The unsteady flow around the model is solved by using the large eddy simulation, and is proved to have a good agreement with the experimental results.

(4) Strong dipole sound sources are formed in the shear layer close to the model surface, but there are no strong dipole sound sources in the wake of each model, where turbulence level takes the highest value. Since φ_i represents the effect of scattering at the body surface, it is natural that the unsteady motion of the flow near the surface plays an important role.

(5) The volume integral at each cross section of the cylinder with periodic holes already takes small value, and cancellation in the spanwise direction is not necessarily dominant. This means that dipole sound sources are mainly suppressed due to momentum injection through holes and the effect of periodicity is not important.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the benefit of detailed advice from Associate Editor Professor M. S. Howe, especially on the physical interpretation discussed in Sec. III C.

[1] C. K. W. Tam and K. A. Kurbatskii, "A wavenumber based extrapolation and interpolation method for use in conjunction with high-order finite difference schemes," J. Comput. Phys. **157**, 588–617 (2000).
[2] J. C. Hardin and D. S. Pope, "Sound generation by flow over a two-dimensional cavity," AIAA J. **33-3**, 407–412 (1995).
[3] C. Kato, M. Kaiho, and A. Manabe, "An overset finite-element large-eddy simulation method with applications to turbomachinery and aeroacoustics," J. Appl. Mech. **70**, 32–43 (2003).
[4] B. S. Holmes, J. B. Dias, B. A. Jaroux, T. Sassa, and Y. Ban, "Predicting the wind noise from the pantograph cover of a train," Int. J. Numer. Methods Fluids **24**, 1307–1319 (1997).
[5] N. Curle, "The influence of solid boundaries on aerodynamic sound," Proc. R. Soc. London, Ser. A **231**(1187), 505–514 (1955).

[6] J. C. Hardin and S. L. Lamkin, ''Aeroacoustic computation of cylinder wake flow,'' AIAA J. **22-1**, 51–57 (1984).
[7] M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge U.P., Cambridge, 1998).
[8] M. S. Howe, *Theory of Vortex Sound* (Cambridge U.P., Cambridge, 2003).
[9] T. Takaishi, A. Sagawa, K. Nagakura, and T. Maeda, ''Numerical analysis of dipole sound source around high speed trains,'' J. Acoust. Soc. Am. **111**, 2601–2608 (2002).
[10] J. C. Hardin and M. Y. Hussaini, *Computational Aeroacoustics* (Springer-Verlag, New York, 1993).
[11] B. E. Mitchell, S. K. Lele, and P. Moin, ''Direct computation of the sound generated by vortex pairing in an axisymmetric jet,'' AIAA-95-0504 (1995).
[12] M. Wang, ''Aerodynamic sound of flow past an airfoil,'' Annual Research Briefs 1995, Center for Turbulence Research, Stanford University/NASA Ames (1995), pp. 257–271.
[13] M. J. Lighthill, ''On sound generated aerodynamically,'' Proc. R. Soc. London, Ser. A **211**(1107), 564–587 (1952).
[14] M. S. Howe, ''Contributions to the theory of aerodynamic sound, with application to excess jet noise and the theory of the flute,'' J. Fluid Mech. **71**, 625–673 (1975).
[15] T. Takaishi, M. Ikeda, and C. Kato, ''Effects of periodic holes on the suppression of aeroacoustic noise from a pantograph horn,'' Proceedings of 4th ASME/JSME Joint Fluids Engineering Conference (on CD-ROM) (2003).
[16] J. Smagorinsly, ''General circulation experiments with primitive equations,'' Mon. Weather Rev. **91**(3), 99–164 (1963).
[17] M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, ''A dynamic subgrid-scale eddy viscosity model,'' Phys. Fluids **3**, 1760–1765 (1991).
[18] M. S. Howe, G. C. Lauchle, and J. Wang, ''Aerodynamic lift and drag fluctuations of a sphere,'' J. Fluid Mech. **436**, 41–57 (2001).

Fredholm integral equation for shape reconstruction of an underwater object using backscattered ramp response signatures

Wei Li[a)]
Department of Naval and Ocean Engineering, Hauzhoung University of Science and Technology, P.R. China, 430074.

G. R. Liu[b)]
Singapore-MIT Alliance, Centre for Advanced Computations in Engineering Science (ACES), Department of Mechanical Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

V. K. Varadan
Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 18 April 2003; revised 18 May 2004; accepted 4 June 2004)

Inverse scattering problems for identifying underwater objects are of importance for all kinds of applications in seas that occupy more than two-thirds of the surface of the earth. The physical optics approximation has been applied to the acoustic scattering integral representation, yielding an inverse scattering equation which relates the profile function of a submerged object to the Fourier transform of the complex backscattered acoustic field. This technique is the so-called ramp response technique and usually requires the full scattering information for all frequencies to obtain accurate profile function of an object. In order to retrieve the shape information of an object from the band-limited backscattered frequency data, a Fredholm integral equation is formulated in this paper for calculating the ramp response of the submerged object. This Fredholm integral equation is for the unknown ramp response of an object in terms of the known incomplete scattering information. The kernel of this integral equation is the inverse Fourier transform of the known characteristic function of the scattering information aperture. In order to simulate the real experiment environment, a computer-generated Gauss noise is added directly to the simulated impulse response. This noise-contaminated impulse response is then used for the inverse identification of the underwater object. It is found that adding noise would lead to the ill-posedness of this inverse scattering problem. To deal with the ill-posedness, the regularization method by filtering is employed. Numerical examples have confirmed that the method presented in this paper can successfully reconstruct the shape of an underwater object using an incomplete set of backscattered frequency data with or without noise contamination. Meanwhile, the effect of bandwidth to the ramp response is also discussed in the numerical examples. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1776190]

PACS numbers: 43.30.Gv, 43.20.Fn [RAS] Pages: 1436–1445

I. INTRODUCTION

Underwater acoustical imaging techniques and the inverse analysis of acoustic scattering problems have now found many important engineering applications. In general, an underwater acoustical imaging system is expected to indicate not only WHERE a scatter is located, but also WHAT it looks like. Solving a properly formulated inverse scattering problem can well determine the size, shape, and orientation of a scatterer using the scattered field without *a priori* knowledge of a scatterer. Solving an inverse problem requires a good solution for the corresponding forward problem involving the determination of the amplitude and phase of the scattered field of a known object submerged in fluid. Many computational methods have been developed for the forward scattering problems. Normally, the inverse scattering problems are much more difficult to solve than the forward scattering problems, as they are often ill-posed. Therefore, care must be taken in both proper formulation of inverse problem and the measures of dealing with the ill-posedness.[1]

The concept of the impulse response was first introduced by Kennaugh and Cosgriff[2,3] and applied to the electromagnetic scattering problems. Based on the physical optics approximation, the far-field impulse response was interpreted as the second derivative of the profile function of a scatterer, which is the cross sectional area versus distance along the line of sight. However, the impulse response was not convenient for retrieving the shape information of a scatterer directly, as this technique involved the differentiation operation on the profile function. Consequently, by integrating twice the impulse response, it was found that the ramp response of

[a)]Electronic mail: engp0341@nus.edu.sg
[b)]http://www.nus.edu.sg/ACES/

 © 2004 Acoustical Society of America

a scatterer is proportional to its profile function. The ramp response technique has been widely utilized in the electromagnetic inverse scattering for profile reconstruction.[4–8] Similar to the electromagnetic wave field, the ramp response technique also has been successfully applied in the acoustic wave field[9–11] and the elastic wave field[12] for solving the inverse scattering problems, such as underwater acoustical imaging and nondestructive evaluation. In order to generate a 3D image of an object from its ramp response, Young[8] presented a new image synthesis technique known as ''approximate limiting surface'' technique. This image synthesis technique requires at least three sets of the backscattered frequency data in different directions to calculate the corresponding profile functions. Although three look-angle profile functions are not sufficient to uniquely determine the object, the final image generated by this technique is still acceptable. Normally, the ramp response technique requires full scattering information at all frequencies.[9,12] In practice, however, the backscattered frequency data are only measured in a frequency band, such as in the high-frequency band. When the data were only available in the high-frequency band, the ramp response obtained by the conventional method did not relate well to the profile function.[9,11]

To solve the inverse scattering problems with incomplete frequency information, Bojarski[13] presented an integral equation for a physical optics inverse scattering identity that related the characteristic function of a scatterer to the three-dimensional spatial Fourier transform of the augmented far field backscattering amplitude. This integral equation is a triple integral equation that is notoriously difficult to solve. Besides this difficulty, Bojarski's method required scattering information for all aspect angles in order to reconstruct the shape of an object. This requirement is really hard to satisfy in practice.

There are also many other inverse scattering methods, such as seismic migration and inversion in geophysical exploration. Migration is an inversion process, which derives two-dimensional (or three-dimensional) maps of local reflectivity from seismic data. Furthermore, migration-inversion is to identify what physical properties changed across that boundary. Stolt and Weglein[14] introduced the concept of migration-inversion and described a wave equation linearized inverse for the three elastic parameters in a vertically changing earth. They also considered the practical difficulties, such as the limited data problem. In 1997, Weglein *et al.*[15] presented a new multidimensional multiple-attenuation method, which could be used to eliminate free surface multiples and attenuate internal multiples while preserving primaries at all offsets, and neither requires any subsurface information for either surface or internal multiples. The inverse-scattering series method has been tested with good results on band-limited synthetic data and could be applied to geophysical exploration, e.g., in deep water with a variable water bottom or in subsalt plays.[16]

In this paper, an integral equation for incomplete scattering information is proposed to solve the inverse scattering problems based on the ramp response technique. This integral equation is a Fredholm integral equation of the first kind, in terms of the unknown ramp response of a scatterer and the known, but incomplete, scattering information. The kernel of the integral equation is the inverse Fourier transform of the scattering information aperture function. To verify the validity of the present method, a finite number of numerical examples are performed for a rigid prolate ellipsoid. Moreover, in order to simulate the real experiment environment, a Gauss noise generated by computer is added on the simulated impulse response of an object directly, and it results in the ill-posedness of this inverse scattering problem. Regularization by filtering is employed to deal with the ill-posedness. The numerical results show that the new method presented in this paper is more effective for retrieving the shape information of an underwater object than the conventional ramp response technique.

II. RAMP RESPONSE TECHNIQUE

For an underwater rigid scatterer, based on the physical optics approximation, the impulse response in the acoustic far field is given by[9]

$$P_I(t)=\frac{1}{2\pi c}\frac{d^2A(ct/2)}{dt^2} \tag{1}$$

and the ramp response is

$$P_R(t)=\int_{-\infty}^{t}\int_{-\infty}^{t'}P_I(t'')dt''\,dt'. \tag{2}$$

In Eq. (1), function $A(ct/2)$ is known as the profile function, which is the cross-sectional area of the object perpendicular to the line of sight, where c is the sound speed in water. Substituting Eq. (1) into Eq. (2), we obtain

$$P_R(t)=\frac{1}{2\pi c}A\left(\frac{ct}{2}\right). \tag{3}$$

For an acoustically rigid object, Eq. (3) shows that the ramp response of an object is proportional to its profile function. Therefore, if we know the ramp response of an object, then the profile function could be obtained easily. Moreover, the 3D image of the object could be generated by the approximate limiting surface method,[8] which requires at least three directions' ramp responses of the object.

On the other hand, the impulse response of an underwater object can be obtained by taking the inverse Fourier transform of its backscattered frequency spectrum $S(\omega)$,[9] and it is shown as follows:

$$P_I(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}(S(\omega)+S^*(-\omega))e^{-i\omega t}\,dt, \tag{4}$$

where the symbol ''*'' denotes complex conjugate. In this paper, the backscattered frequency spectrum $S(\omega)$ is computed by using the well-known T-matrix method,[17–23] which is widely used for solving the acoustical scattering problems of arbitrarily shaped objects in water. Substituting Eq. (4) into Eq. (2) yields

$$P_R(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Gamma(\omega)e^{-i\omega t}\,d\omega, \tag{5}$$

where $\Gamma(\omega)$ is the normalized complex backscattering amplitude and is defined as

$$\Gamma(\omega)=\frac{S(\omega)+S^*(-\omega)}{(-i\omega)^2}. \tag{6}$$

From Eq. (5), it is clear that the relationship between the ramp response and the normalized complex backscattering amplitude is a Fourier transform pair. It should be noted that the normalized complex backscattering amplitude Γ is specified completely by the scattering information of an object, while the ramp response P_R encodes the shape, size, and orientation of an object. In general, Eq. (5) is used to calculate the ramp response in the numerical computation, as the ramp response can be computed directly by performing the fast Fourier transform (FFT) of the normalized complex backscattering amplitude. In the experimental process, however, the impulse response of an underwater object could be measured directly using a hydrophone. Consequently, the ramp response is computed by taking twice the integral of the measured impulse response [see Eq. (2)].

III. INTEGRAL EQUATION FOR INCOMPLETE SCATTERING INFORMATION

When Eq. (5) is employed to compute the ramp response of an object, the backscattered data with complete scattering frequency information are necessary to obtain a perfect ramp response. However, in most practical situations, the scattering information of an object is only measurable in a certain frequency band. In this section, we will discuss how to obtain an excellent ramp response of an object when the backscattered frequency data are only available in a finite frequency band.

First, the scattering information aperture function $H(\omega)$ is defined as unity or zero for the whole frequency domain for which the normalized complex backscattering amplitude Γ is known or unknown, respectively,

$$H(\omega)=\begin{cases}1, & \text{when } \Gamma(\omega) \text{ is known,}\\ 0, & \text{when } \Gamma(\omega) \text{ is unknown.}\end{cases} \tag{7}$$

Furthermore, let $h(t)$ be the inverse Fourier transform of this scattering information aperture function $H(\omega)$,

$$h(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}H(\omega)e^{-i\omega t}\,d\omega. \tag{8}$$

According to the Fourier convolution theorem and by defining $\overline{P}_R(t)=(1/2\pi)\int_{-\infty}^{+\infty}\Gamma(\omega)H(\omega)e^{i\omega t}\,d\omega$, Eq. (5) becomes

$$\int_{-\infty}^{+\infty}h(t-\tau)P_R(\tau)d\tau=\overline{P}_R(t). \tag{9}$$

Obviously, the left-hand side of Eq. (9) is the convolution integral of $h(t)$ and $P_R(t)$. The function $\overline{P}_R(t)$ is the conventional ramp response obtained by the conventional method, which does not consider the effect of the incomplete scattering frequency information. Clearly, it is always possible to calculate the function $\overline{P}_R(t)$ for all possible scattering information aperture functions. Therefore, Eq. (9) is the Fredholm integration of the first kind for the unknown ramp response P_R of an object with the known kernel h and function $\overline{P}_R(t)$. The solution of this Fredholm integration of the first kind is the ramp response, which we want to obtain.

If the whole frequency domain information is known, then the scattering information aperture function $H(\omega)$ will be equal to 1 for all ω, and it yields $h(t)=\delta(t)$ according to the property of the Fourier transform. Therefore Eq. (9) reduces to

$$\overline{P}_R(t)=\int_{-\infty}^{+\infty}\delta(t-\tau)P_R(\tau)d\tau=P_R(t). \tag{10}$$

It is quite apparent that the ramp response obtained by solving the integral equation (9) will be identical to that obtained by employing the conventional method, if the scattering frequency information is complete. However, we cannot get the normalized complex backscattered frequency amplitude Γ at all frequencies in practice. Normally, the backscattering data are only available in a certain frequency range, such as $[f_{\min},f_{\max}]$, and $f_{\max}$ is the Nyquist frequency and cannot be infinite. If $f_{\min}=0$, then the kernel function h will be zero except at $t=0$, and $h(t=0)=2f_{\max}$, when the FFT algorithm is used to calculate the kernel function h. Also, when using the numerical method to solve Eq. (9), the time interval is equal to $1/2f_{\max}$. (The detail procedure of solving the Fredholm integration will be introduced in Section IV A.) Therefore, we could also get the numerical solution of the Fredholm integration $P_R(t)=\overline{P}_R(t)$. It indicates that when the frequency range of the backscattered frequency data starts from zero, then we could consider that the scattering frequency information is complete. Otherwise, the scattering frequency information is incomplete.

IV. NUMERICAL EXAMPLES

In this paper, a submerged rigid prolate ellipsoid is considered as our computational model. The ellipsoid is described below:

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, \tag{11}$$

where a, b, c are the three semiaxes of an ellipsoid. In this paper, we choose $a=b=0.5$ m and $c=1.0$ m. Then the ellipsoid is a rotationally symmetric body. The sound speed in water is chosen as 1500 m/s, and the propagation direction of the incident wave is assumed to be parallel to the z axis. The backscattered frequency data are computed by using the T-matrix method and the FFT algorithm is used to obtain the function $\overline{P}_R$ and the kernel function $h(t)$. For the purpose of convenience, nondimensional wavenumber ka is used in our computation.

A. Analysis using noise-contaminated data

In practical measurement, noise always comes with the useful backscattered data. When we measure the backscattered data of an underwater object, system noise, environment noise and so on will compound with the useful data. When the backscattered data are contaminated by noise, the inverse scattering problems become more complicated to

solve. In this paper, a computer-generated Gauss noise is added directly on the simulated impulse response to simulate the noise-contamination in practice, as the measured data in the experiment is the impulse response. A vector of pseudo-random number is generated from a Gauss distribution with mean ε and standard deviation σ. In this work, the mean ε is set to zero, and the standard deviation σ is defined as

$$\sigma = \frac{\sqrt{2}}{2} |P_I|_{\max} \times \gamma \times 100\%, \tag{12}$$

where P_I is the impulse response data, and γ is a positive constant, which represents the ratio of the Gauss noise to the impulse response (measured data). It implies that γ Gauss noise is added on the measured data.

In this paper, the argument of P_R and $\bar{P}_R$ is set as $ct/2$. Then the dimension of the x axis in the ramp response figure will be length in our computation. The backscattered frequency spectrum $S(\omega)$ and the normalized complex backscattering amplitude Γ are discrete data in the frequency domain. Consequently the functions $P_R(t)$ and $h(t)$ are also discrete data, as they are obtained using the FFT algorithm. Then the Fredholm integral equation [Eq. (9)] becomes a system of linear equations. Suppose 1024 frequency points are used in computing $S(\omega)$; the number of the linear equations will also be 1024, and hence the length of the computed ramp response will be tens or hundreds of times as long as the size of an object. It is really unnecessary and impractical. In practice, the length of the impulse response as well as the ramp response obtained by measurement is only several times longer than the size of an object. In this paper, therefore, we assume that the length of the ramp response is two times as long as the size of an object. According to this assumption, the length of the functions h and $\bar{P}_R$ should be truncated properly. Consequently, the number of the linear equations will be determined by the truncation length. After the truncation processing, normally, the number of the linear equations is reduced to 10–20, and hence the cost for solving the linear equations is largely reduced.

First, $\gamma=2\%$ is considered in the computation, and it is a very good measurement in practice. The scatterer is the ellipsoid described in Eq. (11) and the range of computational frequency is set as $1 \leqslant ka \leqslant 5$. The numerical results of the functions $P_R(t)$ and $\bar{P}_R(t)$ are shown in Fig. 1 (solid lines). For the purpose of comparison, the numerical results obtained using the pure backscattered data (without noise), and the exact profile function of the ellipsoid are also shown in Fig. 1. It was quite apparent that the conventional ramp response $P_R(t)$ of the ellipsoid was totally different with its profile function [see Fig. 1(b)]. It suggests that the conventional ramp response technique cannot retrieve the shape information of an object along the incident direction while the backscattered frequency data are only available in a certain frequency band. However, the ramp response P_R obtained using the present method was nearly proportional to the profile function of the ellipsoid [see Fig. 1(a)], if the band-limited backscattered frequency data were noise free. It shows that the ramp response obtained by solving the Fredholm integral equation could be successfully used to identify the shape of an underwater object even when the pure backscattered data are band limited.

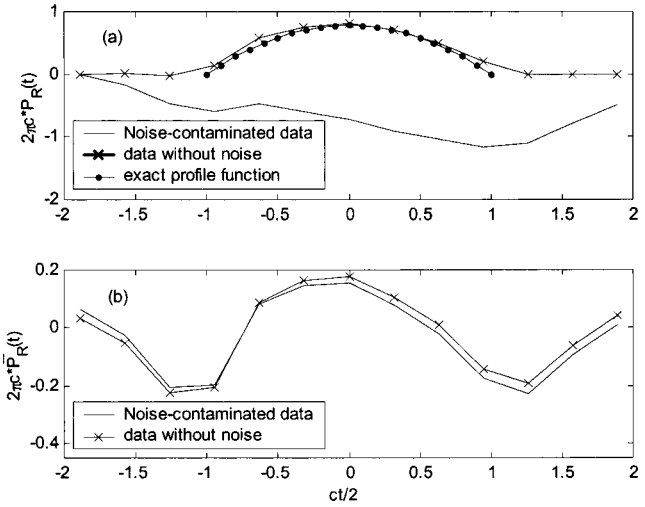


FIG. 1. Comparison between the pure ramp responses ($\gamma=0\%$) of the ellipsoid in water and the noise-contaminated ramp responses ($\gamma=2\%$): (a) ramp responses P_R obtained by the present method and (b) ramp responses $\bar{P}_R$ obtained by the conventional method.

When the noise existed, we found that the function $\bar{P}_R$ obtained using the noise-contaminated data was very close to that obtained using the data without noise [see Fig. 1(b)], as γ was very small. However, when the noise-contaminated data $\bar{P}_R$ were used as the input data of Eq. (9), it was found that the solution P_R [solid line in Fig. 1(a)] was totally different with the solution [solid line with "×" marks in Fig. 1(a)] obtained by using the noise-free backscattered data. It implies that this inverse scattering problem becomes an ill-posed problem, when the band-limited backscattered data are the noise-contaminated data. Although the relative error in the input data is very small, the relative error in the solution becomes very large. The main reason is that the condition number of the coefficient matrix cond(C) in Eq. (9) is significant large. Suppose the relative error in the input data (function $\bar{P}_R$) is E_X, and then the upper limit of the relative error in the solution will be $\mathrm{cond}(C)\cdot E_X$.[24] It indicates that the relative error in the solution, i.e., ramp response P_R, is probably amplified, when the value of cond(C) is quite large.

In order to deal with the ill-posedness, many regularization methods have been developed,[1] such as Tikhonov regularization[25] that needs to regularize the elements of the coefficient matrix in the inverse problem to get a stable solution. However, the elements of the coefficient matrix in our problem are generated exactly using Eq. (8). If we do some regularization on the coefficient matrix, it may change the characterization of the coefficient matrix. In this paper, we do not do any regularization on the coefficient matrix. Instead, we use filtering regularization to solve this ill-posed problem.

As the useful backscattered data are only available in a certain frequency band, such as $1 \leqslant ka \leqslant 5$, it means that the useful impulse response of an underwater object as well as its ramp response are the band-limited frequency data. On

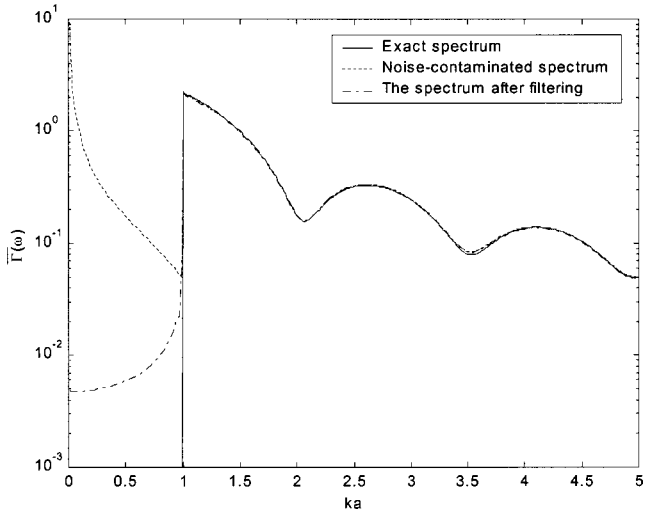


FIG. 2. Comparison of the exact spectrum, the noise-contaminated spectrum, and the spectrum after filtering for the underwater ellipsoid when the backscattered data are available in $1 \leqslant ka \leqslant 5$ and $\gamma = 2\%$.

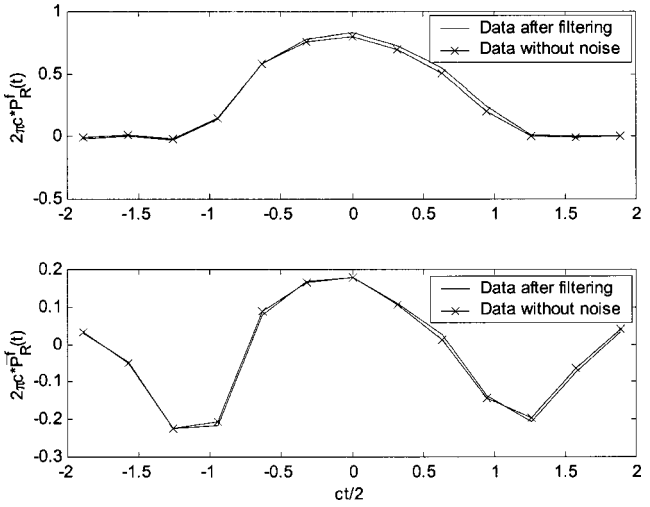


FIG. 3. Comparison between the ramp responses after filtering ($\gamma = 2\%$) and the pure ramp responses ($\gamma = 0\%$) for the submerged ellipsoid ($1 \leqslant ka \leqslant 5$): (a) ramp responses P_R and P_R^f obtained by the present method and (b) ramp responses $\overline{P}_R$ and $\overline{P}_R^f$ obtained by the conventional method.

the other hand, the frequency components of the computer-generated Gauss noise are roughly equally distributed on the whole frequency domain. As a result, the frequency components of the noise-contaminated impulse response also cover the whole frequency domain. Furthermore, after taking twice integration of the noise-contaminated impulse response to yield the conventional ramp response, it could be expected that the high-frequency components in the noise-contaminated ramp response would decline rapidly, while the low-frequency components would decline slowly. Thus it is necessary to input the noise-contaminated ramp response into a proper high-pass filter and minimize the noise components. In this paper, a digital elliptic high-pass filter is designed for this purpose. In order to eliminate the nonlinear phase distortion of the elliptic filter, a zero-phase filtering approach[26] is employed.

Figure 2 shows the frequency spectrums of the conventional ramp responses $\overline{P}_R$, i.e., $\overline{\Gamma}(\omega)$. The computational parameters used in Fig. 2 are the same as those used in Fig. 1. For the purpose of comparison, the exact spectrum (solid line), the noise-contaminated spectrum (dash line), and the spectrum after filtering (dash-dot line) are all shown in Fig. 2. In the $1 \leqslant ka \leqslant 5$ range, these three curves were matched very well as the ratio of the noise components to the useful data was very low ($\gamma = 2\%$). It was also due to the fact that the noise components in the ramp response declined rapidly with increase of frequency. However, in the $0 \leqslant ka \leqslant 1$ range, the amplitude of the noise-contaminated spectrum was somewhat big, especially when ka approached zero. But by the processing of inputting the noise-contaminated data into the high-pass elliptic filter with the cutoff frequency $ka = 1$, the filtering results (dash-dot line in Fig. 2) were very good. The amplitude of the filtering spectrum declined rapidly in the $0 \leqslant ka \leqslant 1$ range with decrease of frequency. It indicates that the elliptic filter works very well for our case, and most of the noise components could be filtered out by taking the filtering regularization. Here $\overline{P}_R^f(t)$ is used to denote the conventional ramp response after filtering regularization. Substituting the filtering result $\overline{P}_R^f(t)$ into Eq. (9), then it yields the ramp response $P_R^f(t)$ through solving the Fredholm integral equation. The calculation results are shown in Fig. 3. After the filtering regularization, the numerical result in the $\gamma = 2\%$ case was in excellent agreement with the pure ramp response (data without noise). It suggests that this filtering regularization method is valid to solve this ill-posed inverse scattering problem.

To verify again the efficiency of the filtering regularization method, we increase the noise components in the input data. In other words, we increase the value of γ, and choose $\gamma = 5\%$, 10%, 15%, 20%, respectively. The numerical results of P_R^f and $\overline{P}_R^f$ for the underwater ellipsoid are all shown in Fig. 4. In Fig. 4, it was found that when γ increased, the differences between the pure ramp response ($\gamma = 0\%$) and the noise-contaminated ramp response also increased. But the differences were not very large, even for the $\gamma = 20\%$ case.

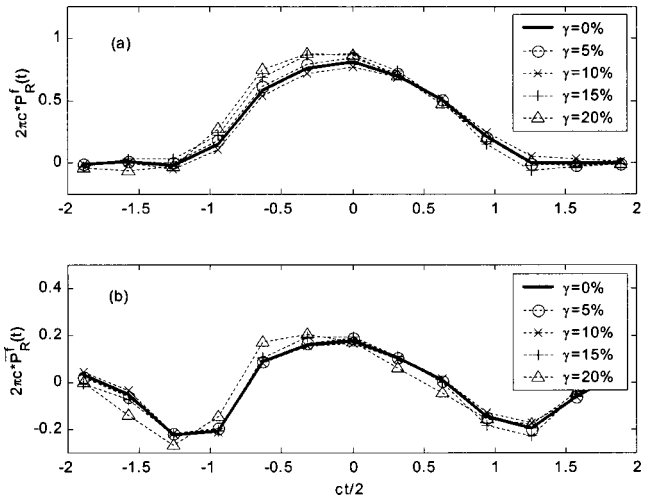


FIG. 4. Comparison between the ramp responses of the submerged ellipsoid after filtering regularization with different γ ($1 \leqslant ka \leqslant 5$): (a) ramp responses P_R^f obtained by the present method and (b) ramp responses $\overline{P}_R^f$ obtained by the conventional method.

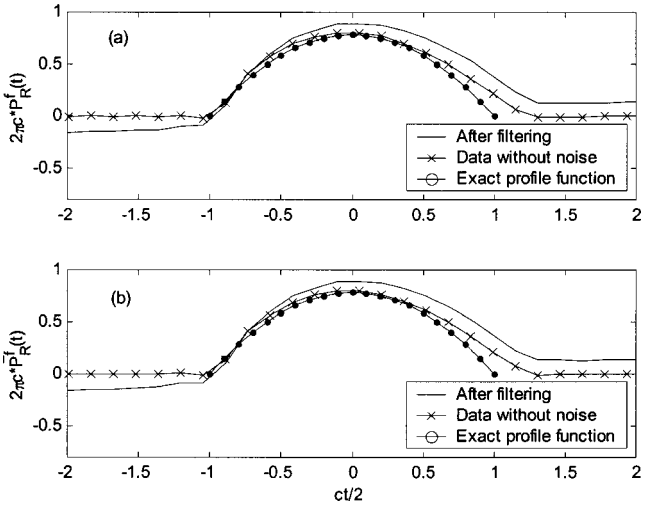


FIG. 5. Comparison between the computed ramp responses of the submerged ellipsoid using the complete noise-contaminated scattering information ($0 \leqslant ka \leqslant 10$) and the corresponding pure ramp responses: (a) ramp responses P_R and P_R^f obtained by the present method and (b) ramp responses $\overline{P}_R$ and $\overline{P}_R^f$ obtained by the conventional method.

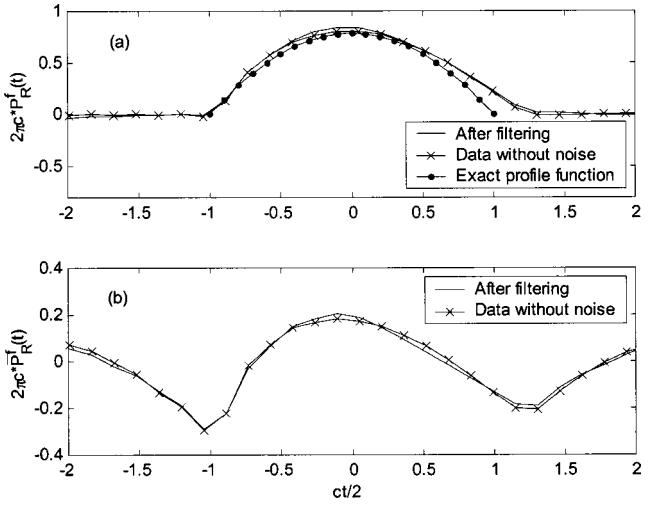


FIG. 6. Comparison between the computed ramp responses of the submerged ellipsoid using the noise-contaminated band limiting backscattered frequency data ($1 \leqslant ka \leqslant 10$) and the corresponding pure ramp responses: (a) ramp responses P_R and P_R^f obtained by the present method and (b) ramp responses $\overline{P}_R$ and $\overline{P}_R^f$ obtained by the conventional method.

The ramp response signatures obtained by the present method were good enough for us to retrieve the shape information of the underwater ellipsoid. It implies that after the filtering regularization, the most noise could be effectively filtered out, and the numerical results become less sensitive on noise. Therefore, this method is effective to calculate the ramp response even when the input band-limited data are noise contaminated.

B. Effect of bandwidth on ramp response

In Sec. IV A, we presented how to obtain the ramp response of an underwater object by using the present method when the frequency data were band limited, and pointed out that the noise mixed in the useful data would lead to seriously ill-posed problem. A filtering regularization method was addressed to solve the ill-posed problem. In this section, we will discuss the effect of bandwidth on calculating the ramp response by using the present method. $\gamma = 10\%$ is considered in all the following computations. Figures 5–8 show the computational results of the functions P_R^f and $\overline{P}_R^f$ for the prolate ellipsoid in the different frequency bands using the noise-contaminated data, and solid lines are used to denote them. For the purpose of comparison, the ramp responses without noise (solid line with "×" marks) and the exact profile functions (solid line with dots) are also given in these figures.

In Fig. 5, the nondimensional wavenumber ka is set from 0 to 10. According to the analysis in Sec. III, it means that the backscattered frequency information is complete, and the ramp response P_R obtained by the present method should be equal to $\overline{P}_R$. The numerical results shown in Fig. 5 also proved this point. Whether the backscattered data were contaminated by noise or not, the calculated ramp responses P_R and P_R^f of the ellipsoid were identical with the corresponding functions $\overline{P}_R$ and $\overline{P}_R^f$. When the backscattered data did not contain noise, the ramp response data multiplied by $2\pi c$ in Fig. 5 were nearly the same as the exact profile function of the ellipsoid. However, if the backscattered data were contaminated by noise, the difference between the ramp response after filtering regularization P_R^f and the ramp response without noise was significant. In this situation, the high-pass filter became useless due to $ka_{\min} = 0$. It suggests that when the backscattered frequency information is complete and noise free, the present method could be used to identify the size and shape of an underwater scatterer. However, for the noise-contaminated data, this method may not generate a perfect profile function.

In Figs. 6 and 7, the maximum nondimensional wavenumber $ka_{\max}$ is equal to 10 and the minimum nondimensional wavenumber $ka_{\min}$ is equal to 1 and 2, respectively. We found that if the band-limited scattering data ($1 \leqslant ka$

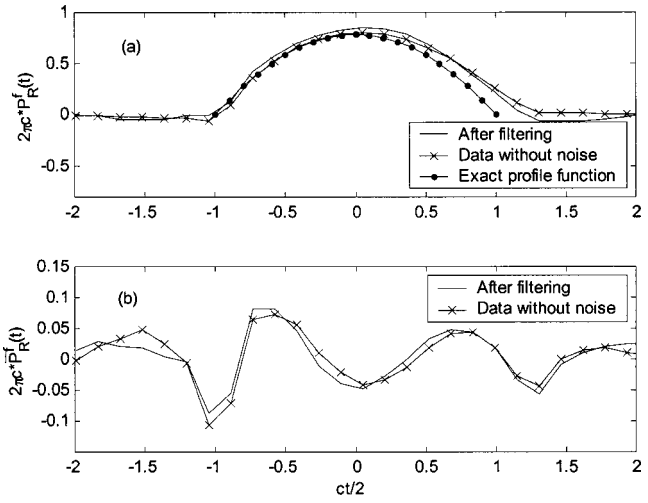


FIG. 7. Comparison between the computed ramp responses of the submerged ellipsoid using the noise-contaminated band limiting backscattered frequency data ($2 \leqslant ka \leqslant 10$) and the corresponding pure ramp responses: (a) ramp responses P_R and P_R^f obtained by the present method and (b) ramp responses $\overline{P}_R$ and $\overline{P}_R^f$ obtained by the conventional method.

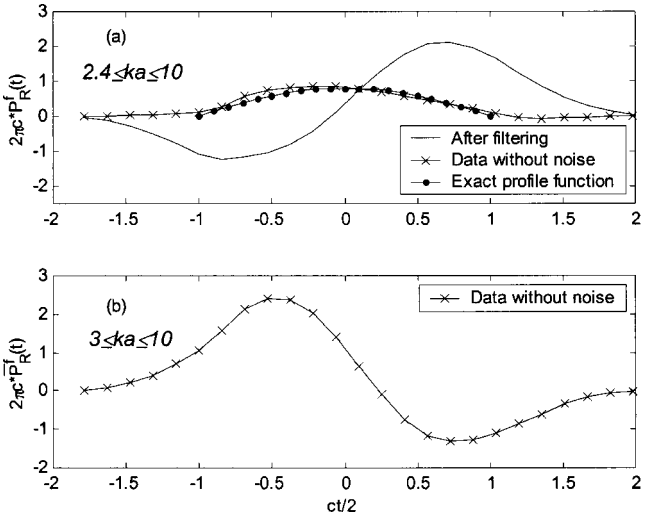


FIG. 8. Ramp responses of the submerged ellipsoid using the band limiting backscattered frequency data, which are obtained by the present method: (a) $2.4 \leqslant ka \leqslant 10$ and (b) $3 \leqslant ka \leqslant 10$.

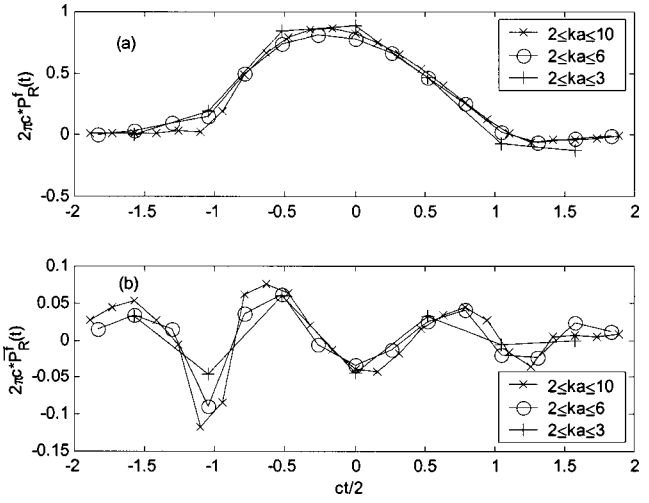


FIG. 9. Ramp responses of the submerged ellipsoid when the noise-contaminated backscattered frequency data (γ=10%) are available in different frequency bands: (a) ramp response P_R^f obtained by the present method and (b) ramp response $\overline{P_R^f}$ obtained by the conventional method.

$\leqslant 10$ or $2 \leqslant ka \leqslant 10$) were used, the ramp responses $\overline{P_R}$ and $\overline{P_R^f}$ obtained by the conventional method could not be used to identify the shape of the ellipsoid [see Figs. 6(b) and 7(b)]. But, when the effect of the incomplete frequency information was considered and the Fredholm integral equation [Eq. (9)] was applied to compute the ramp response, the numerical results P_R and P_R^f could be used to determine the shape of the ellipsoid [see Figs. 6(a) and 7(a)] whether the data were noise contaminated or not. It indicates that the ramp response obtained by solving the Fredholm integral equation and after filtering regularization could be successfully used to identify the shape of an underwater scatterer, even when the backscattered frequency data are band limited and noise contaminated.

If we continue to increase $ka_{\min}$ to 2.4, the numerical results of P_R and P_R^f for the ellipsoid are shown in Fig. 8(a). It was clear that when the backscattered data did not contain noise, the ramp response P_R was still nearly the same with the profile function of the ellipsoid and could be used to estimate its shape. However, if we used the noise-contaminated data to calculate the ramp response P_R^f by the present method, then we could not retrieve any shape information of the ellipsoid from the plot of P_R^f. This phenomenon is probably due to the fact that when the minimum nondimensional wavenumber $ka_{\min}$ increases, the condition number cond(C) in Eq. (9) will also increase. Meanwhile, although the filtering regularization can reduce the error of the noise-contaminated data, the error cannot be eliminated totally. Therefore, when the condition number cond(C) in Eq. (9) increases to quite large, then it will lead to significant relative error in the output ramp response by the present method. Furthermore, if the minimum nondimensional wavenumber $ka_{\min}$ increases to 3, then the ramp response of the ellipsoid shown in Fig. 8(b) becomes totally different with the exact profile function and we cannot get the shape information from it, although the ramp response is obtained using the noise-free data. It suggests that if the minimum nondimensional wavenumber $ka_{\min}$ were too large, then the ramp response obtained by the present method would not be proportional to the profile function of the object even for the ideal data. It implies that the low-frequency information is very important to calculate a perfect ramp response P_R.

On the other hand, we fix the minimum nondimensional wavenumber $ka_{\min}$ as 2, and vary $ka_{\max}$ to compute the ramp responses of the ellipsoid using the present method. The numerical results are shown in Fig. 9. In Fig. 9, the input backscattered frequency data have been added 10% Gauss noise. It was found that when $ka_{\max}$ decreased from 10 to 3, the ramp responses P_R^f obtained by the present method were nearly the same and could be used to determine the shape of the ellipsoid. Especially in the case of $2 \leqslant ka \leqslant 3$ (solid line with "+" marks in Fig. 9), the relative length of the frequency band was only $\frac{1}{3}$. But the numerical result of the function P_R^f in the case of $2 \leqslant ka \leqslant 3$ was fairly good, and very close to the results in other two cases. It indicates that a perfect ramp response could be obtained in a narrow frequency band when $ka_{\min}$ is not very large. If we, however, continue to decrease the value of $ka_{\max}$, we will get a bad ramp response result. Because when the FFT algorithm is employed to calculate the ramp response of an object, the time interval of two adjacent points in the ramp response plot is determined by the maximum frequency value of the backscattered frequency data. The time interval will be enlarged with a decrease of $ka_{\max}$. Thus if $ka_{\max}$ is too small, the time interval of two adjacent points in the ramp response plot will be quite large. Consequently, the points, which fall into the range of interest in the ramp response plot, such as $[-1,1]$ in our case, will decrease. In Fig. 9(a), there are only five points in the range of interest for the case of $2 \leqslant ka \leqslant 3$. So if we continue to decrease $ka_{\max}$, then there will be fewer points fallen in this range, and these so few points are not enough to describe the shape of an object. In short, the value of $ka_{\max}$ should be chosen properly and could not be too small.

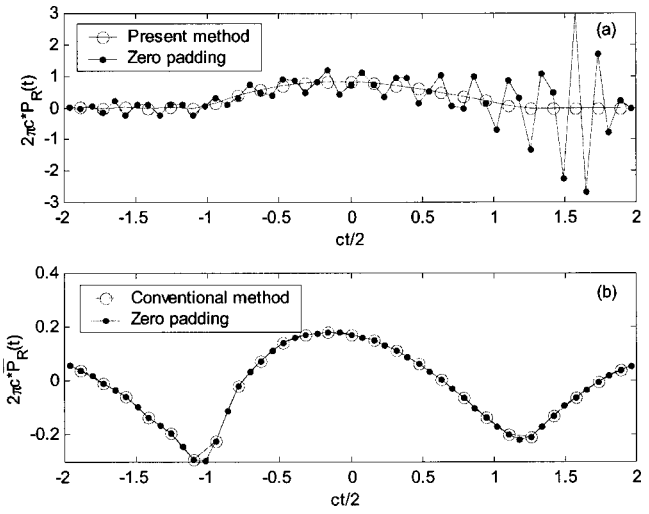


FIG. 10. Comparison of the ramp response results between the FFT method and the zero padding method.

C. Discussions of the zero padding method instead of the FFT method

According to the analysis in the last section, it was found that when the maximum frequency $ka_{\max}$ of the backscattering data decreased, then the time resolution of the ramp response was also decreased. In order to improve the time resolution of the ramp response, a zero padding method instead of the FFT method may be used to calculate the ramp response. The main idea of the zero padding is to extend a spectrum (or signal) with zeros to extend its frequency band (or time) limits. In this work, this method may also be used to calculate the ramp response of a rigid object.

Here, an underwater rigid spheroid with $c/a=2$ is used as the computational model. The original backscattering frequency data from this spheroid are calculated from $ka=1$ to 10. By the zero padding method, the backscattering frequency spectrum is extended from 10 to 20. Then the new frequency band of ka is from 1 to 20. In this section, only noise-free data are considered.

Figure 10 shows the numerical results of the ramp responses obtained by the zero padding method as well as the FFT method. It was quite clear that for the conventional ramp response signals $\bar{P}_R$, the zero padding output was nearly the same with the FFT output as shown in Fig. 10(b). As the band limit was extended, the time resolution of the zero padding output was one time higher than that of the FFT output. When the function $\bar{P}_R$ was used as the input data of Eq. (9) to solve the ramp response P_R, the FFT result was clearly in excellent agreement with the exact profile function. But the zero padding result was greatly unstable and could not be used to estimate the shape of the spheroid as shown in Fig. 10(a). The main reason for this phenomenon is that the condition number of the coefficient matrix cond(C) in Eq. (9) becomes very large, when the zero padding method is used. In order to obtain the zero padding result of the ramp response, it is necessary to calculate the zero padding results of the kernel function $h(t)$ and the function $\bar{P}_R(t)$. As a result, the size of the coefficient matrix in Eq. (9) is enlarged. It will lead to increase the value of cond(C) largely. In this case, the condition number cond(C) was only 6.6014E+01, when the FFT method was used. However, for the zero padding result, the value of cond(C) increased to 1.6171E+10. It is apparent that so large a condition number will definitely lead to serious instability in the result. Therefore, the zero padding method is not suitable to deal with the backscattering frequency data here, when the band-limited ramp response technique is used to estimate the profile function of a rigid object.

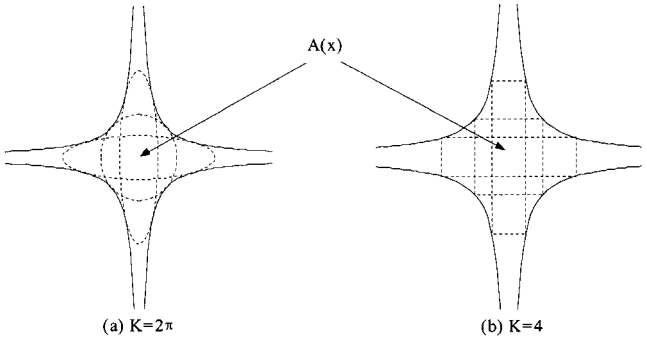


FIG. 11. Hyperbolic contour of the approximate limiting surface.

D. Image reconstruction

Once the ramp response of an underwater object at one looking angle is obtained using the method presented in this paper, then its profile function of the object could be obtained directly at this looking angle. If the profile functions at any three different looking angles are known, we can reconstruct a 3D image of the object by using the ''approximate limiting surface'' technique. This synthetic imaging technique will generate limiting surfaces for each one of the profile functions. The limiting surfaces supposedly enclose the object with the tightness of fit varying with target shape and looking angle, and the limiting contour is assumed to be a hyperbola shown in Fig. 11, which is defined as follows:

$$K|yz| = A(x), \tag{13}$$

where K is equal to 4 or 2π.

It can be proved that any rectangle with area $A(x)$ is inscribed of the hyperbolic contour surface for $K=4$, while any ellipse with area $A(x)$ is inscribed of the hyperbolic contour surface for $K=2\pi$. The common part to all limiting surfaces of different looking angles is the image result. In order to improve the image quality, the adjustable parameters are applied to modify the limiting surface. Modifications in-

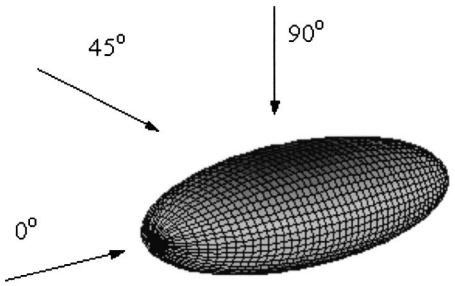


FIG. 12. 3D image of the ellipsoid and illustration of three incident angles.

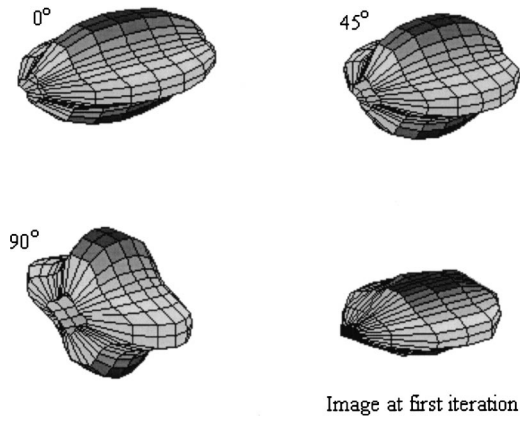


FIG. 13. Approximate limiting surfaces corresponding to the three looking angles and image of the ellipsoid obtained at the first iteration of image reconstruction.

clude a limitation on the length of the legs of the hyperbolic contour, and rotation and change of axis of each looking angle limiting surface. The approximate limiting surface in each looking angle is remolded iteratively till the best fit is obtained according to the preset error criteria. If convergence is obtained, the intersection of the final approximate limiting surfaces is the synthetic image of the object.

The rigid spheroid with $c/a=2$ is used as our model, and it is shown in Fig. 12. The incident waves propagate in three different directions. All angles of the incident waves are measured with respect to the rotational axis of symmetry of the spheroid, and are chosen as 0°, 45°, 90° (see Fig. 12). The backscattered data in the $1 \leqslant ka \leqslant 5$ range are used to calculate the ramp responses for each one looking angles. Also $\gamma=10\%$ Gauss noise is considered to add on the impulse responses. The approximate limiting surfaces for three looking angles and their intersection at the first stage of iteration are presented in Fig. 13. It is clear that the image quality at first iteration is very bad. The main reason is that the three looking angles are not orthogonal each other. It implies that some modifications are necessary when the looking angles are not orthogonal. After five iterative steps, the final image result is shown in Fig. 14, and the image quality is satisfying.

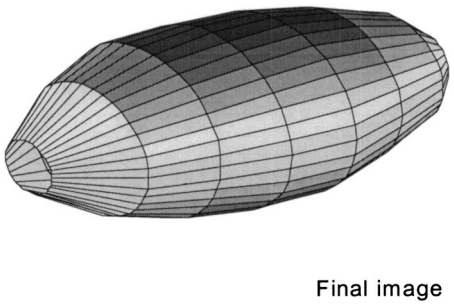


FIG. 14. Final 3D image of the ellipsoid obtained after five iterations of image reconstruction.

V. CONCLUSIONS

Based on the physical optics approximation, it is found that the ramp response of an underwater object is proportional to its profile function. This technique is used to solve the inverse scattering problems and reconstruct the shape of an underwater object. However, if the backscattered data are only available in a certain frequency band, then the ramp response obtained by the conventional method is not proportional to the profile function, and cannot be used to estimate the shape of the object. In this study, a new numerical method is proposed to compute a perfect ramp response for the incomplete scattering information case, and the Fredholm integral equation is developed based on the ramp response technique. To simulate the practical situation, the computer-generated Gauss noise is added directly on the impulse response. When the input band-limited data are noise contaminated, this inverse scattering problem becomes an ill-posed problem, and the filtering regularization is proposed for obtaining a perfect ramp response result. The numerical results show that by the process of the filtering regularization, most of the noise components could be effectively filtered out, and a perfect ramp response could be obtained using the present method when the backscattered data are band limited. However, in order to get an excellent ramp response, the lower limit of the frequency band should not be too large. Also, the decrease of the upper limit of the frequency band will lead to lower the time resolution of the ramp response.

When the ramp responses at three different looking angles are known, the 3D image of the object could be reconstructed by the approximate limiting surface method, and the image quality is satisfying after several iterative steps. Therefore, the present method could be successfully applied for the shape reconstruction of a single underwater rigid object. For multiple objects, the ramp response technique would fail to find the shape information of these objects. Furthermore, many underwater objects have very complicated boundary conditions, such as fluid–fluid, fluid–gas, or fluid–elastic boundary conditions. Therefore, further research works are necessary for more complicated cases.

ACKNOWLEDGMENTS

The work received support from the National University of Singapore, Singapore, and Institute of High Performance Computing, Singapore. We greatly appreciate the valuable comments provided by the reviewers.

[1] G. R. Liu and X. Han, *Computational Inverse Technique in Nondestructive Evaluation* (CRC, Boca Raton, 2003).
[2] E. M. Kennaugh and R. L. Cosgriff, "The use of impulse response in electromagnetic scattering problems," in IRE Nat. Conv. Rec., Part 1 (1958).
[3] E. M. Kennaugh and D. L. Moffatt, "Transient and impulse response approximations," Proc. IEEE **53**, 893–901 (1965).
[4] B. Y. Foo, S. K. Chaudhuri, and W. M. Boerner, "Polarization correction and extension of the Kennaugh-Cosgriff target-ramp response equation to the Bistatic case and application to electromagnetic inverse scattering," IEEE Trans. Antennas Propag. **38**, 964–972 (1990).
[5] C. C. Chen and L. Peters, Jr. "Ramp response signatures for UXOs," *Geoscience and Remote Sensing Symposium, 2000. Proceedings. IGARSS 2000. IEEE 2000 International* (2000), Vol. 4, pp. 1436–1438.
[6] S. Nag and L. Peters, Jr. "Ramp response signatures of dielectric targets,

especially land mines,'' *Geoscience and Remote Sensing Symposium Proceedings. 1998. IGARSS'98 IEEE 1998 International* (1998), Vol. 1, pp. 213–215.
[7] S. Nag and L. Peters, Jr., ''Ramp response signatures for dielectric targets,'' *Proceedings of SPIE on Detection and Remediation Technologies for Mines and Minelike Targets III* (1998), pp. 703–713.
[8] J. D. Young, ''Radar imaging from ramp response signature,'' IEEE Trans. Antennas Propag. **AP-24**, 276–282 (1976).
[9] S. J. J. Tsao, ''Image reconstruction using ramp response signatures,'' Ph.D. dissertation, The Pennsylvania State University, 1985.
[10] X. M. Zhang, W. Li, and G. R. Liu, ''A new technique in ramp response for acoustic imaging of underwater objects,'' Appl. Acoust. **63**, 453–465 (2002).
[11] G. R. Liu, W. Li, X. M. Zhang, and V. K. Varadan, ''Shape identification of underwater objects using backscattered frequency signals,'' J. Acoust. Soc. Am. **113**, 3146–3154 (2003).
[12] V. V. Varadan, S. J. Tsao, V. K. Varadan, and B. R. Tittmann, ''Image reconstruction of flaws using ramp response signatures,'' J. Wave-Mater. Interact. **10**, 67–78 (1995).
[13] N. N. Bojarski, ''A survey of the physical optics inverse scattering identity,'' IEEE Trans. Antennas Propag. **AP-30**, 980–989 (1982).
[14] R. H. Stolt and A. B. Welglein, ''Migration and inversion of seimic data,'' Geophysical. **50**, 2458–2472 (1985).
[15] A. B. Weglein, F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, ''An inverse-scattering series method for attenuating multiples in seismic reflection data,'' Geophysical. **62**, 1975–1989 (1997).
[16] K. Matson, D. Corrigan, A. B. Weglein, and C-Y Young, ''Inverse scattering internal multiple attenuation: results from complex synthetic and field data examples,'' *SEG Annual Meeting Expanded Abstracts 1999* (1999), pp. 1060–1063.
[17] P. C. Waterman, ''New formulation of acoustic scattering,'' J. Acoust. Soc. Am. **45**, 1417–1429 (1969).
[18] V. Varatharajulu and Y.-H. Pao, ''Scattering matrix for elastic wave. I. Theory,'' J. Acoust. Soc. Am. **60**, 556–566 (1976).
[19] V. V. Varadan and V. K. Varadan, ''Scattering matrix for elastic wave. III. Application to spheroids,'' J. Acoust. Soc. Am. **65**, 896–905 (1979).
[20] J.-H. Su, V. V. Varadan, V. K. Varadan, and L. Flax, ''Acoustic wave scattering by a finite elastic cylinder in water,'' J. Acoust. Soc. Am. **68**, 686–691 (1980).
[21] A. Sarkissian, C. F. Gaumond, and L. R. Dragonette, ''T-matrix implementation of forward scattering from rigid structures,'' J. Acoust. Soc. Am. **94**, 3448–3453 (1993).
[22] G. C. Bishop and J. Smith, ''Scattering from rigid and soft targets near a planar boundary: Numerical results,'' J. Acoust. Soc. Am. **105**, 130–143 (1999).
[23] R. Lim, K. L. Williams, and E. I. Thorsos, ''Acoustic scattering by a three-dimensional elastic object near a rough surface,'' J. Acoust. Soc. Am. **107**, 1246–1262 (2000).
[24] G. Dahlquist and Å. Björck, *Numerical Methods*, translated by N. Anderson (Prentice–Hall, Inc., Englewood Cliffs, NJ, 1974).
[25] A. Tikhonov and V. Arsenin, *Solution of Ill-posed Problems* (Wiley, New York, 1977).
[26] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Inc., Englewood Cliffs, NJ, 1989), pp. 311–312.

Propagation of sound through a spicy ocean, the SOFAR overture

Matthew Dzieciuch,[a] Walter Munk, and Daniel L. Rudnick
Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093

(Received 20 February 2003; accepted for publication 21 May 2004)

Using a closely sampled 1000-km hydrographic section in the eastern North Pacific, the sound-speed finestructure is separated into two component fields: (*i*) isopyncal tilt dominated by internal waves (the traditional view) and (*ii*) ''spicy'' (cold-fresh to hot-salty) millifronts associated with upper ocean stirring. Numerical transmission experiments show significant scatter within the mixed layer from the spicy fronts. This scattered energy arrives near the start of the SOFAR sequence, and is superimposed on a triplication of the channel dispersion at the transition from reflected to upper ocean refracted energy. (This SOFAR *overture* is totally different from the *finale* which has been prominent for over 50 years.) The critical dependence of the overture on mixed layer processes suggests a scheme for acoustically monitoring the upper oceans at surface-conjugate depths (3 to 5 km), offering some advantages over *in situ* monitoring. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1772397]

PACS numbers: 43.30.Re, 43.30.Qd [WJS] Pages: 1447–1462

I. INTRODUCTION

Perhaps the outstanding characteristic of long-range transmissions in the ocean is the high degree of variability. In the early days of SOFAR transmissions, attempts to account for the variability based on laboratory concepts of homogenous isotropic turbulence were favored for their analytical convenience. But ocean variability is neither homogenous nor isotropic. Munk and Zachariasen (1976) found such ''reliance on laboratory concepts...though fashionable, to be entirely misplaced,'' and then proposed, ''...internal waves and internal tides (as the) principal source of fluctuations.'' Their scattering estimates referred to (but did not include) ''intrusive and other forms of finestructure.''

For nearly 30 years internal waves have occupied a position near to monopoly in the literature on acoustic variability (Flatté *et al.*, 1979). But the consideration in this paper of frontal structure is not just a second item in an infinite menu of ocean variability; we suggest that the combination of internal waves and frontal activity can account for a major part of the measured variability.

Ewart *et al.* (1977) estimated the effect of non-internal-wave-related finestructure from data obtained in summer 1977 during the Mid-Ocean Acoustic Transmission Experiment (MATE) near Cobb Seamount in the northeastern Pacific. The temperature field was monitored from time series of moored sensors and vertical CTD profiles (Levine and Irish, 1981, 1986). The term ''finestructure'' was defined to represent a vertical structure in the temperature field with horizontal scales large compared to those of the internal waves (see also Garrett and Munk, 1971).

Ewart (1980) performed a numerical simulation of the effect of such finestructure on acoustic transmissions. Phases (travel times) remained unaffected but intensities were greatly modified. Here we examine acoustic transmissions through the ocean as measured by the towed vehicle SeaSoar. Ewart's vertical CTD finestructure profiles differ from the gradients measured with the SeaSoar horizontal tows in that many of the temperature ''jumps'' occur on scales small compared to the internal wave lengths. However, the spacing of fronts is comparable to internal wave scales, and spatial filtering does not readily separate the two fields. We perform the separation by a procedure that recognizes the intrinsic distinction between the two processes.

A cartoon (Fig. 1) will prepare the reader for the forthcoming analysis of ocean transects. The unperturbed ocean has a surface mixed layer with water of uniform density (as indicated by the density of dots) overlying a sharply increasing density structure $\rho(z)$. The isopycnals $z(\rho)$ are horizontal. Internal waves vertically displace the water particles, as indicated by the tilting isopycnals. However, the layers of constant density are not horizontally homogeneous: steep fronts separate regions of cold and fresh water (blue) from hot and salty water (red), with green indicating temperature and salinity averages of the blue and red waters, taken horizontally over many fronts. Temperature and salinity transitions are compensated, so that the density does not vary across the ''spice'' front. Accordingly the fronts do not play a role in ocean dynamics, but they do in ocean acoustics. Internal waves and spice fronts form two essential components of the sound-speed finestructure.

We have derived a scheme for separating the tilt and spice fields in a recent 1000-km section of temperature and salinity profiles in the North East Pacific. Numerical acoustic transmission experiments are conducted along (*i*) the measured profile, (*ii*) with spice removed, (*iii*) with tilt (internal waves) removed, and (*iv*) for the smoothed profile without any finestructure. The results are shown in the traditional timefront displays as they would be measured by a vertical receiver array at 1000 km range. Early timefronts are modified by the mixed layer processes, with different modification for different processes.

[a]Electronic mail: mad@ucsd.edu

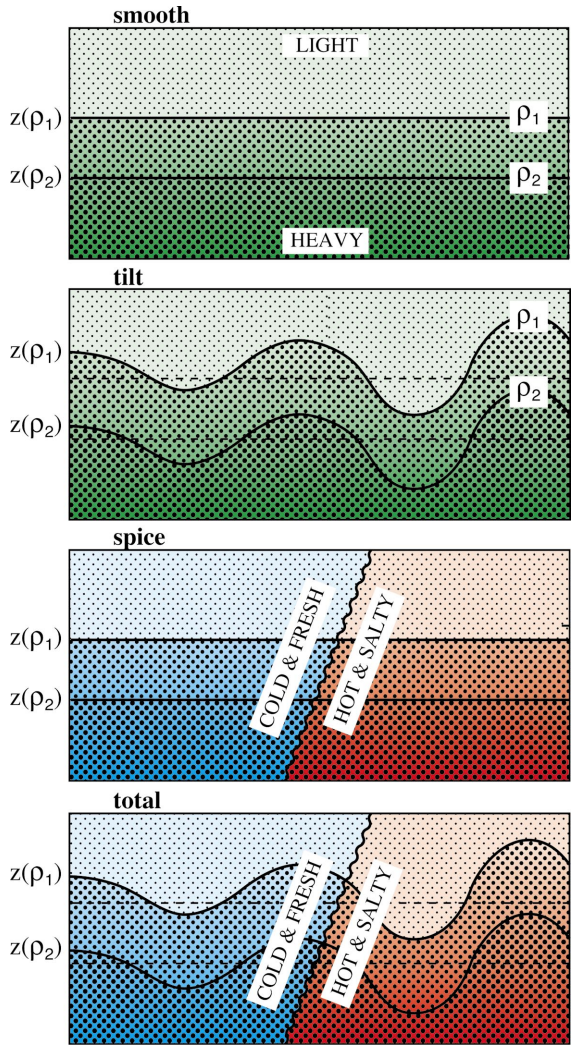


FIG. 1. Cartoon illustrating the separation of internal wave and spice related finestructure (see text).

To interpret the results, we consider a simple acoustic model which allows for (*i*) the early transition from surface reflected to upper ocean refracted rays, and (*ii*) the seasonal formation of the surface mixed layer. The resulting time-fronts are further modified by scattering within and beneath the mixed layer. The sensitive dependence of the overture on mixed layer processes suggests a scheme for acoustic monitoring; conducting the acoustic measurements at surface-conjugate[1] depths (3 to 5 km) has some advantages over *in situ* monitoring of upper ocean processes.

II. UPPER OCEAN FINESTRUCTURE

Figure 2 shows a potential density section in the upper 320 m of the North Pacific (Rudnick and Ferrari, 1999; Ferrari and Rudnick, 2000). Data were taken using a towed vehicle, SeaSoar, that is actively controlled from the ship by varying the pitch of its wings. A series of tows were done along 140 °W between 25° and 35 °N. First, SeaSoar was flown in a sawtooth pattern between 5 and 320 dbar, completing a cycle at least once per 12 min. A tow speed of 4 $\mathrm{m\,s^{-1}}$ produced a cycle in 2.88 km. Subsequent tows targeted surfaces identified in the first section. The second and third

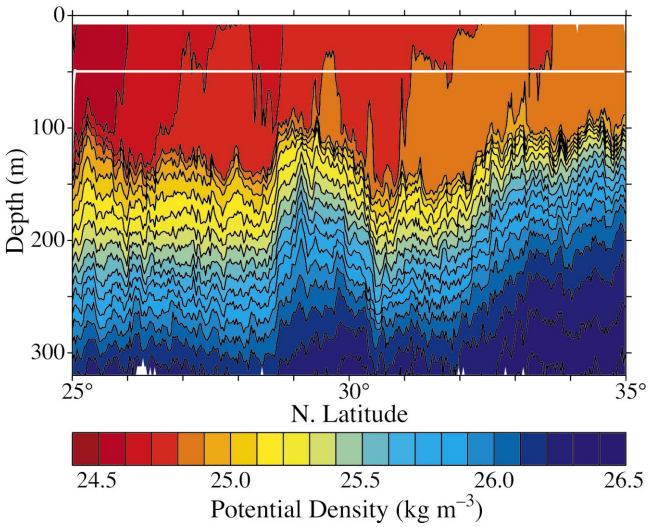


FIG. 2. Potential density in the upper ocean along 140 °W, measured using the towed vehicle SeaSoar (Rudnick and Ferrari, 1999). Contour interval is 0.1 $\mathrm{kg\,m^{-3}}$. Data are binned in depth and time prior to contouring, yielding a 3 km by 8 m resolution. Note the homogeneity of density in the upper mixed layer.

tows were along pressure surfaces in the mixed layer, 50 dbar, and in the thermocline, 200 dbar. It was found that the large-scale structure in the section were persistent over at least two weeks. Finescale features may change over the time required to complete the section, but the statistics should not, so the section offers a plausible realization of the finestructure.

A mixed layer of 100- to 150-m thickness lies atop the thermocline (Fig. 2). Density in the mixed layer is remarkably uniform, with differences between the surface and the bottom of the mixed layer of less than one part in 10^5. An outstanding feature in the mixed layer is the presence of fronts of all scales in which potential temperature and salinity transitions are density compensated (Fig. 3).

Fronts are common features of geophysical fluids, characterized by localized strong horizontal gradients in any of a variety of properties. Here, we are particularly interested in spice fronts, whose locally strong gradients are caused by the straining effects of stirring. As spice fronts have no density signature, these fronts have no effect on dynamics, but are ideal passive tracers. Young (1994) has given the interesting interpretation that the density compensation is a matter of survival: noncompensated fronts are gravitationally unstable and will slump; they are subsequently annihilated by the vertical mixing processes that form the upper homogenous layer. The surviving fronts are a manifestation of the stirring processes, which generate mean-square gradients. Little is known about the relevant time scales, but the dearth of finescale density fronts in the mixed layer suggests that the slumping process is fast relative to the processes (surface flux and vertical entrainment) that cause density anomalies.

The familiar horizontal variations of sound-speed caused by internal waves are associated with the prominent undulations of isopycnal contours beneath the mixed layer (Fig. 2). Less familiar are the transitions associated with the temperature and salinity fronts in and beneath the mixed layer. These frontal transitions are very different from the Gaussian-like

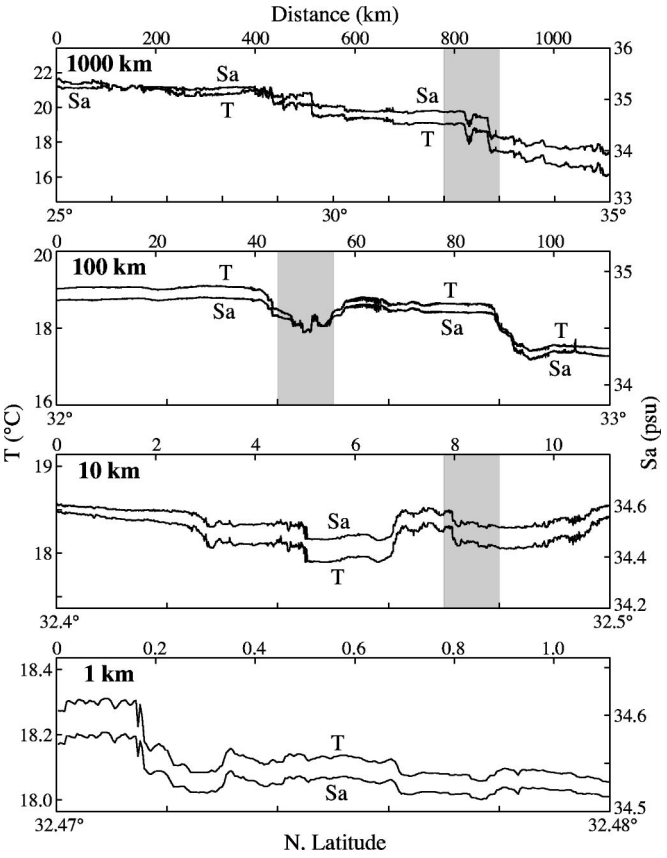


FIG. 3. Potential temperature (blue) and salinity (green) along 50 dbar (white line in Fig. 2). Vertical axes are so scaled that equal excursions of temperature and salinity imply equal but opposite effects on density. The upper panel covers the entire tow; succeeding panels are magnified by a factor of 10 of the shaded region of the panel above.

variability associated with internal waves. Although the temperature T and salinity Sa transitions are compensated with regard to density, this does not imply that the sound-speed is horizontally uniform; in fact the parameters that are density compensating are sound-speed enhancing. It is useful to define the field orthogonal to the density field in a T/Sa diagram (Fig. 4). This is called spice as it varies from cold and fresh to hot and salty.[2] Munk (1981) used the term spice to distinguish density-compensated ''globs'' from internal wave-related finestructure. Let

$$\frac{\Delta C}{C}=a\Delta T+b\Delta Sa, \quad \frac{\Delta\rho}{\rho}=-\alpha\Delta T+\beta\Delta Sa,$$

$$\frac{\Delta\mu}{\rho}=\beta\Delta T+\alpha\Delta Sa \tag{1}$$

for the variation in potential sound-speed, density, and spice as functions of potential temperature and salinity. Typical numerical values are

$$a=2.0\times10^{-3}\ ^\circ\mathrm{C}^{-1}, \quad \alpha=0.25\times10^{-3}\ ^\circ\mathrm{C}^{-1},$$

$$b=0.74\times10^{-3}\ \mathrm{PSU}^{-1}, \quad \beta=0.75\times10^{-3}\ \mathrm{PSU}^{-1},$$

where °C and PSU are temperature and salinity units. Write

$$\frac{\Delta C}{C}=a\Delta T(1+m), \quad \frac{\Delta\rho}{\rho}=-\alpha\Delta T(1+n) \tag{2}$$

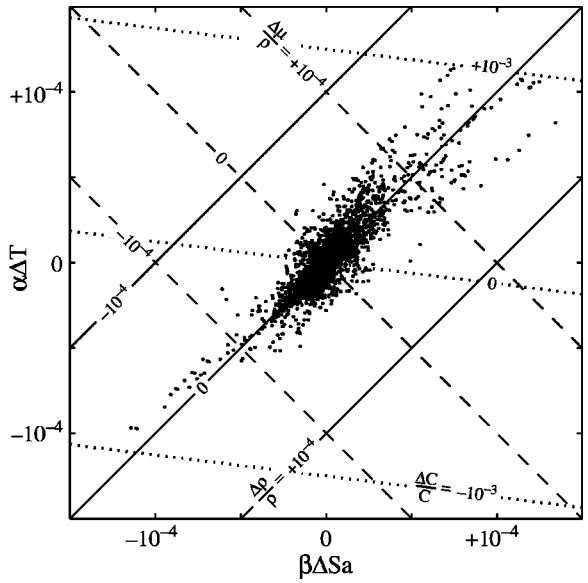


FIG. 4. A scatter plot of $\alpha\Delta T$ versus $\beta\Delta Sa$ throughout the mixed layer (Fig. 1). Points are horizontal differences over the resolution length of 2.88 km. Three sets of contour lines are drawn according to Eq. (1); density (ρ) and spice (μ) contours have slopes $R=\pm1$ (by definition). Data align along $\Delta\rho/\rho=0$, indicating that temperature and salinity differences are compensating in their effect on density. $\Delta C/C$ contours have a slope of -0.12 (see text).

so that

$$m=\frac{-\alpha b}{a\beta}n=-0.12n. \tag{3}$$

Total compensation corresponds to a ''density ratio''[3] of $n=\beta\Delta Sa/(-\alpha\Delta T)=-1$ and $m=+0.12$. Accordingly, in the spicy fronts of the upper North Pacific, the frontal salinity transitions lead to a near 100% *reduction* in the temperature-associated density transitions, and a 12% *augmentation* in the temperature-associated sound-speed transitions.

Beneath the mixed layer the temperature effects dominate, with the density ratio diminishing abruptly from $n=-1$ to nearly $n=-\frac{1}{2}$, a curiosity related by Schmitt (1981, 1994) to salt fingering (see also Stommel, 1993; Rudnick and Martin, 2002). Regardless of the numerical values, in a spicy environment with temperature fronts the density transitions are greatly reduced but the sound-speed transitions are somewhat enhanced by the salinity transitions.

Figure 5 shows sound-speed variability along constant density surfaces. This is equivalent to displaying sound-speed in an ocean without internal waves.

A. Separation into tilt- and spice-related processes

We are quite accustomed to filtering in space and time. But here we have two processes, internal waves and spicy fronts, with overlapping scales. For separation we depend on the intrinsic distinction between the two processes.

By mapping and filtering the recorded temperature and salinity fields in density space, we can separate processes that cause spice variability, such as stirring, from those associated with isopycnal displacement, such as internal waves. Oceanographers would stop there, but acousticians require these filtered fields in physical space be appropriate to sound

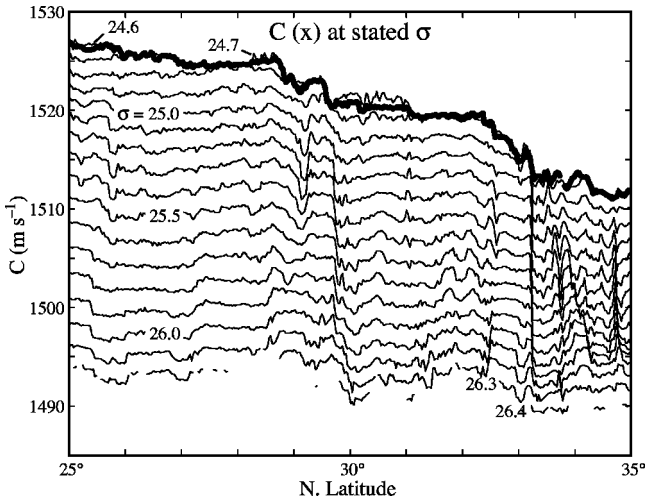


FIG. 5. Sound-speed profiles along constant potential density surface by steps of 0.1 kg m^{-3}. Shorter curves at the top are from outcropping isopycnals; discontinuous curves at the bottom extend below the deepest observations. The heavy purple line near the top is from a tow at near the middle of the mixed layer.

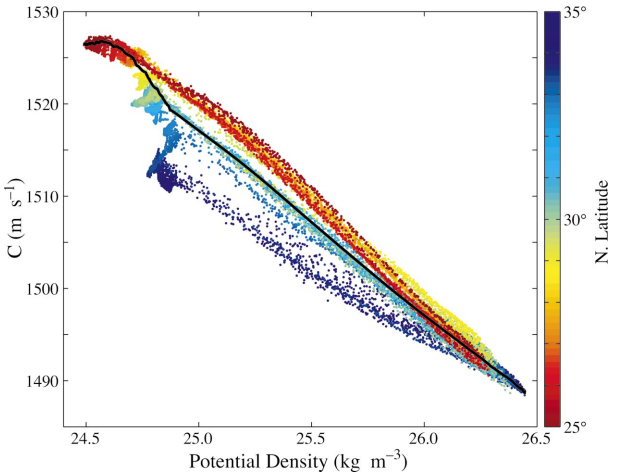


FIG. 6. Scatter plot of sound-speed against potential density for the upper ocean. The relation varies with latitude; the black curve is the latitude-averaged sound-speed.

transmission. Let $T(x,z)$ and $\rho(x,z)$ designate the measured fields of potential temperature and density measured as a function of latitude and depth. We replace the independent depth variable with potential density in the expression for temperature, yielding $T(x,\rho)$. Similarly, we can obtain $Sa(x,\rho)$ and hence $C(x,\rho)$; any contribution from internal waves and other tilt-generating processes is eliminated in the x, ρ-display. To return to physical space, we place the ρ-coordinate at the position of its average depth (Fig. 1, spice panel). The "spice" ocean can be imagined as the straightening-out of isopycnals by pulling them horizontally.

The "tilt" ocean is obtained by computing the x-average, holding ρ constant, $T_{\text{tilt}}(\rho)=\langle T(x;\rho)\rangle_x$, and mapping back into physical space by using

$$T_{\text{tilt}}(x,z)=T_{\text{tilt}}(\rho(x,z)) \tag{4}$$

and similarly for $Sa_{\text{tilt}}(x,z)$ and $C_{\text{tilt}}(x,z)$.

The cartoon in Fig. 1 neglects some essential considerations: we do not wish to throw out permanent oceanographic features with the elimination of internal waves. This is accomplished as follows: the measured field $\rho(x,z)$ is inverted to yield the isopycnals $z(x;\rho)$. Next, the isopycnals are operated on by a low-pass filter LP (Hanning filter with 50 km horizontal scale) yielding a "regional mean depth"

$$z_{\text{LP}}(x;\rho)=\text{LP}[z(x;\rho)], \tag{5}$$

and its inverse $\rho_{\text{LP}}(x,z)=z_{\text{LP}}^{-1}(x,\rho)$. The spice contribution is then obtained from

$$T_{\text{spice}}(x,z)=T(x;\rho_{\text{LP}}(x,z)). \tag{6}$$

This eliminates internal waves while retaining large-scale hydrography. Transient nearly geostrophically balanced motions, often called mesoscale eddies, are a dominant source of variability at length scales of tens of kilometers. Mesoscale eddies cause isopycnals to tilt, and stir spice gradients. As such, they contribute to both the tilt and spice fields.

Finally, we require an ocean with smooth isopycnals and no spice. This is accomplished by mapping the horizontally averaged temperature back into physical space using the low-passed density field:

$$T_{\text{smooth}}(x,z)=T_{\text{tilt}}(\rho_{\text{LP}}(x,z)). \tag{7}$$

This spice-free ocean, containing only large-scale isopycnal tilt, is a reference against which to evaluate the effect of spice and internal waves. The transformation between density space and physical space is nonlinear, and so the total field of any variable cannot be a sum of the smooth, spice, and tilt fields. For a meaningful analysis all sound-speed fields must have a sound channel.

Figure 6 is a scatter plot for the entire section. This functional relationship $C(\rho)$, indicated by the black line, between sound and potential density follows the center of mass (measured in the vertical) of the points in the scatter plot. The vertical scatter is a direct consequence of spice. The white spaces between points are caused by spice fronts (for example, 25 kg m^{-3} at 31 °N). Our examinations of the tilt field, and previous studies that ignore spice, essentially use a characterization of sound-speed that is a function only of density, equivalent to the black line. The scatter plot shows the sound-speed variability lost in such a characterization.

Figures 7 and 8 show the contributions of tilt and spice to the upper ocean fields of salinity and sound-speed. The bottom panel shows the total fields, as directly measured using SeaSoar. By construction, contours in the tilt-only sound field parallel the isopycnals in Fig. 2; in particular, the spice variations in the northern mixed layer are absent. The spice field is superimposed on smooth isopycnals. A careful comparison of the total and spice fields reveals the spice fronts in the total sound-speed. A smooth sound-speed field where the only variations are due to large-scale tilt is shown in the top panel. This is our approximation to an ocean with no internal waves and no spice. The smoothed profile contains the large-scale ocean variability along this 25 °N to 35 °N ocean section.

Strong, vertically coherent spice fronts can be seen near 29° and 33 °N (Fig. 3). These fronts are likely permanent

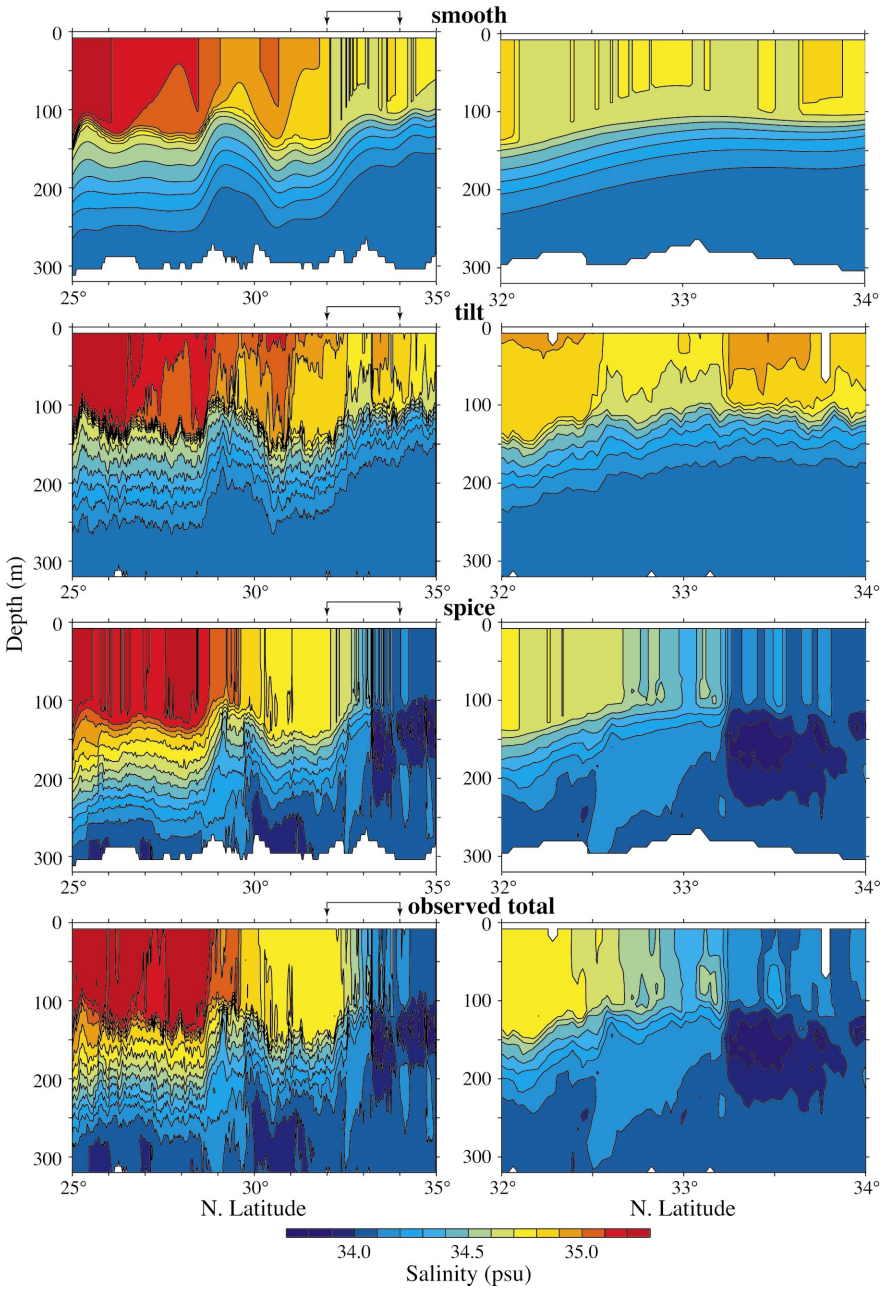


FIG. 7. Fields of salinity with 0.1 psu contour intervals for the entire section (left) and a subsection (right).

features of the region (Yuan and Talley, 1992), although their exact positions may vary. There are also variations of relatively short vertical scale, as for example, near 32 °N and 1505 m s^{-1}. This small vertical scale variability changes significantly between successive occupations of the section several days apart.

In the thoroughly mixed upper layer the sound-speed perturbations associated with internal-wave-induced vertical displacements are small, and the effect of spice fronts dominates. In the mixed layer at 48 m, $\Delta C = 0.24$ m/s for the rms difference at the resolved 3-km spacing. Beneath the mixed layer at 200 m we find $\Delta C = 0.74$ m/s and 0.35 m/s associated with internal waves and spice, respectively. For comparison, the pronounced frontal feature at 29 °N has differences of order 0.5 m/s per 3 km. A mean slowing by -0.1 m/s in 3 km is associated with the south-to-north cooling along the 1000 km profile. The statistical distribution of the horizontal gradients associated with isopycnal tilting are nearly Gaussian (as has long been known), whereas the frontallike distribution of spice leads to large departures from a Gaussian distribution.

B. Patching the upper ocean to Levitus climatology

We have patched Levitus February climatology (Levitus, 1994; Levitus *et al.*, 1994) to the upper 320-m section taken by SeaSoar 23 January–20 February 1997 (Fig. 9). There is no good way of splicing a single section to a decadal climatology. After many tries we have settled on a least-square vertical cubic spline applied separately to the temperature and salinity fields, and subsequently converted to sound-speed. The cubic spline allows for the disparate error bars,

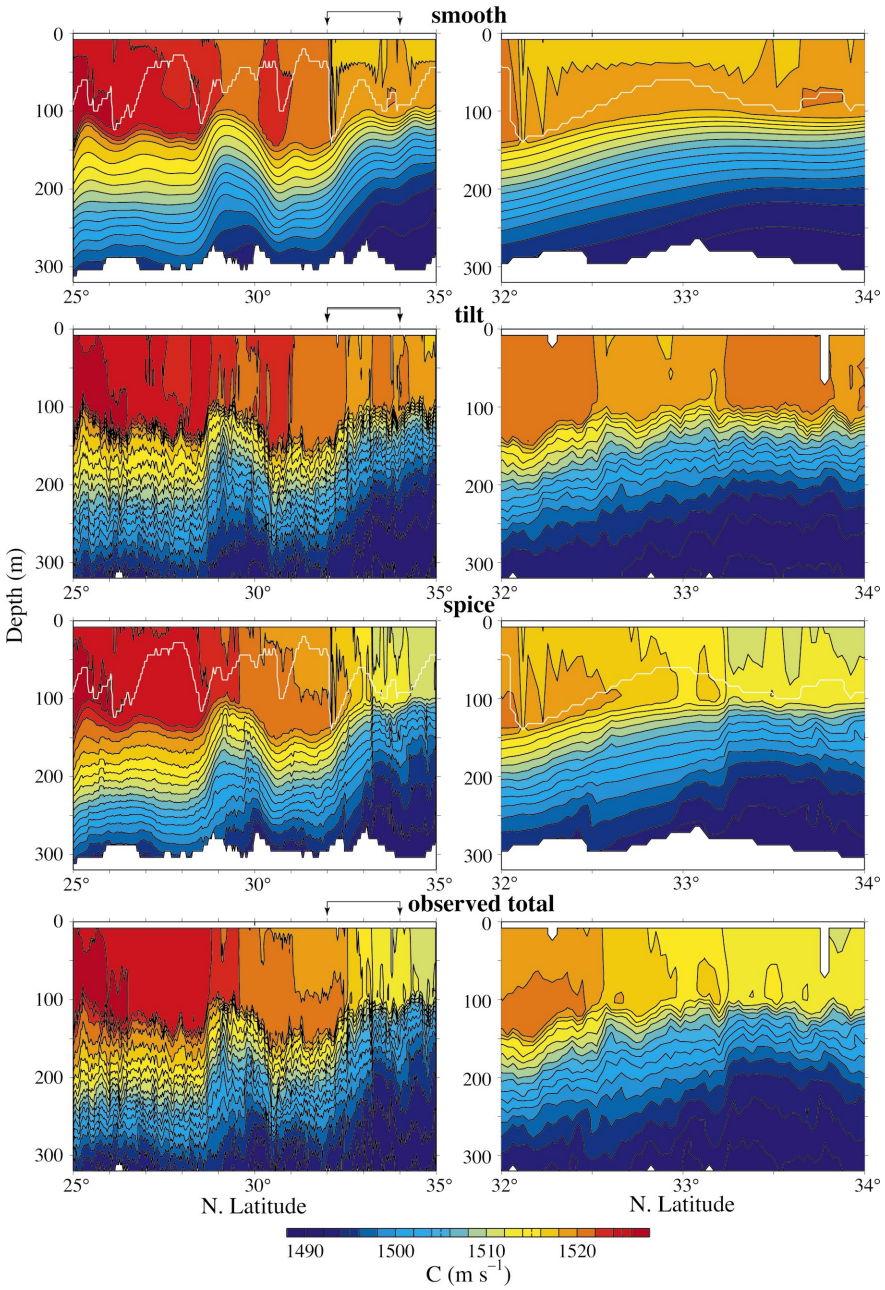


FIG. 8. Fields of sound speed with 2 m s^{-1} contour intervals for the entire section (left) and a subsection (right). Left bottom panel is the observed field (shown in Fig. 2). See text for separation into tilt-only, spice-only, and smoothed (no tilt, no spice) fields. The fields serve as basis for the following numerical transmission experiments. The white line indicates the region above which fields are determined by extrapolation to the surface as described in the text.

0.001 °C, 0.001 PSU for SeaSoar and 0.1 °C, 0.1 PSU for the climatology. Since the sampling grid of the two data sets are different, care has to be taken to avoid having the spline impose excessive curvature on the interpolated profile. The adopted procedure is to relax the SeaSoar error bars at the two deepest grid points (312 m, 320 m) by a factor of 100 to match the climatology error bars. This method avoids dC/dz discontinuities with resulting acoustic caustics.

The climatological mean field suppresses the internal wave perturbations and all but the major permanent frontal features. Horizontal surveying of the deep ocean is difficult, and we know little about the deep spice structure. However, spice heterogeneity is known to decay with depth (Ferrari and Rudnick, 2000), and so we suspect that spicy fronts are features mainly of the upper oceans. On the other hand, internal waves are known to pervade the entire water column, and their neglect here must lead to an underestimate in the scattering.

III. A NUMERICAL EXPERIMENT IN ACOUSTIC SCATTER

In order to investigate the effects of small-scale structure on acoustic propagation, we devised a series of numerical experiments resembling typical field experiments see Colosi (1999), (Fig. 10). A broadband sound source at 100±12.5 Hz was placed on the sound channel axis (730 m) at 25 °N. Sound pressure as function of time and depth was calculated

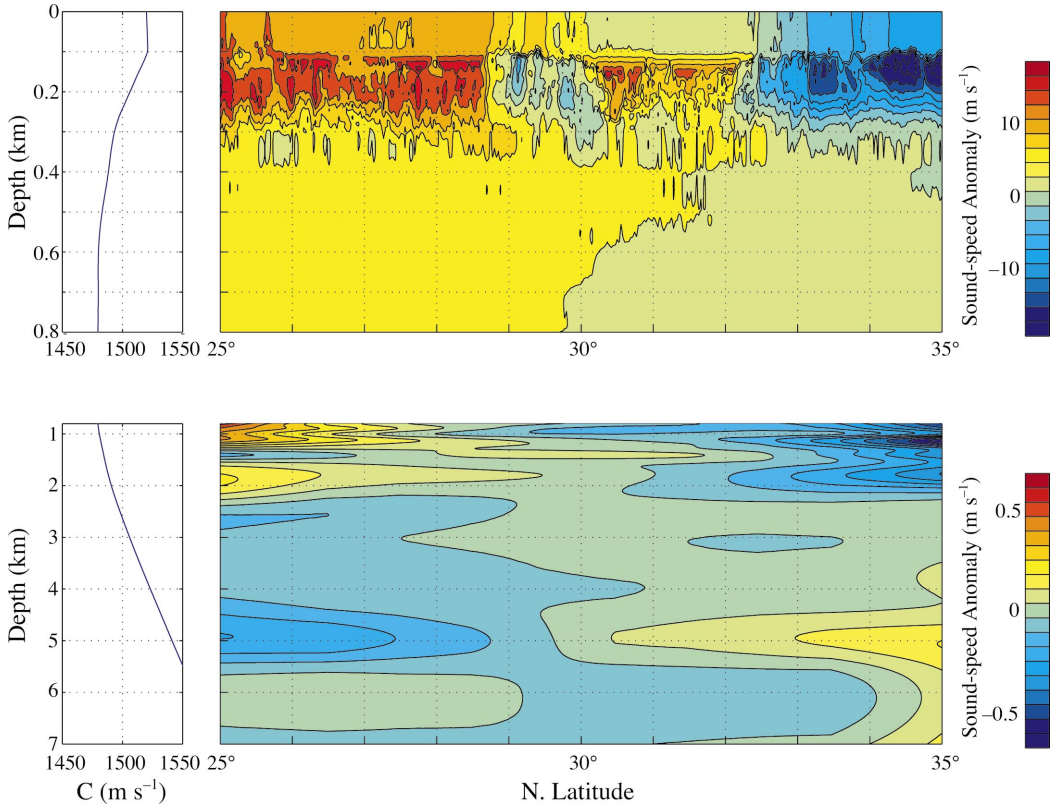


FIG. 9. Contours of sound-speed anomalies relative to mean profile (left panels). Measurements from SeaSoar tows in the upper 300 m are patched to the Levitus Ocean Atlas climatology. Note the intensive sound-speed finestructure in the upper layers, in contrast to the almost uniform potential density (Fig. 2).

at a range of 1000 km for each of the four ocean realizations described in the previous section. A further calculation was performed for the range-averaged smoothed profile.

A. Parabolic equations

The standard parabolic equation (PE) method was applied to underwater acoustics by Tappert (1977), but the split-step Pade PE (Collins, 1989), available in a computer code called RAM, served as a basis for all numerical experiments since deep-water profiles with a mixed layer channel require PE methods that handle wide propagation angles.

The range-averaged profile in the left panel of Fig. 9, with a mixed layer in the upper few hundred meters, is an essential feature in the scattering process (as we shall see). A grid spacing of 0.2 m in the vertical, 50 m in range, with four Pade terms, was found to converge and the result is shown in the top panel of Fig. 10. We are concerned with water-borne energy, and so placed a thick absorbing layer beneath the water column at 5-km depth to trap any energy that might interact with the bottom and obscure the mixed layer arrivals. In the range-independent case of the top panel, a comparison with an accurate mode computation by the Chebyshev method (Dzieciuch, 1993) was made and the mismatch was too small to be visible. Ray codes also matched the travel time predicted by the RAM PE and the modal calculation, except for the earliest arrivals which have turning depths just below the surface and thus finite-frequency boundary effects are important. It is interesting that the range-averaged case shows more early structure than the range-dependent cases.

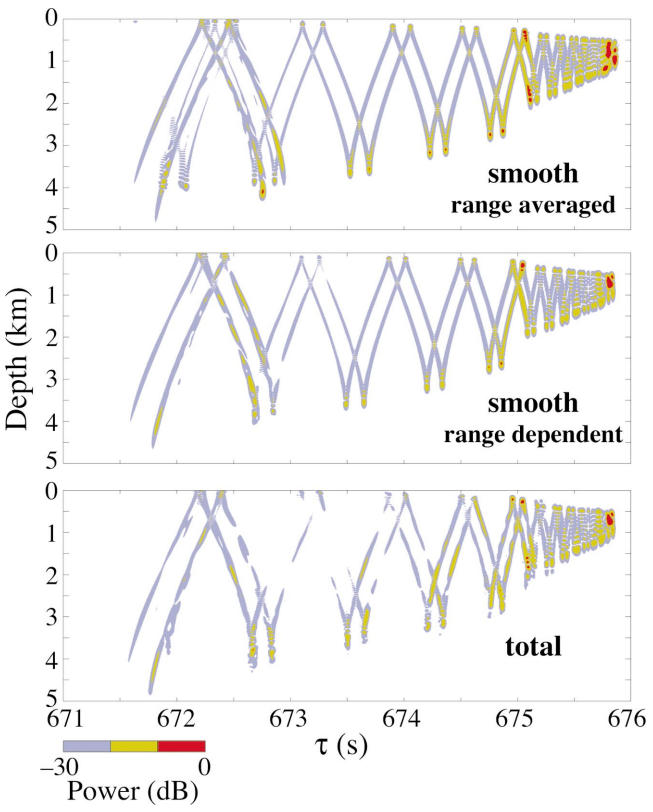


FIG. 10. Intensity field for an acoustic transmission from an axial source at 25 °N to a receiving array at 35 °N. The upper panel corresponds to the range-averaged sound-speed profile (Fig. 9, left). The second and third panels are for transmission through the smooth and observed fields of Fig. 8. All contours are relative to the axial climax in the smooth range-dependent transmission.

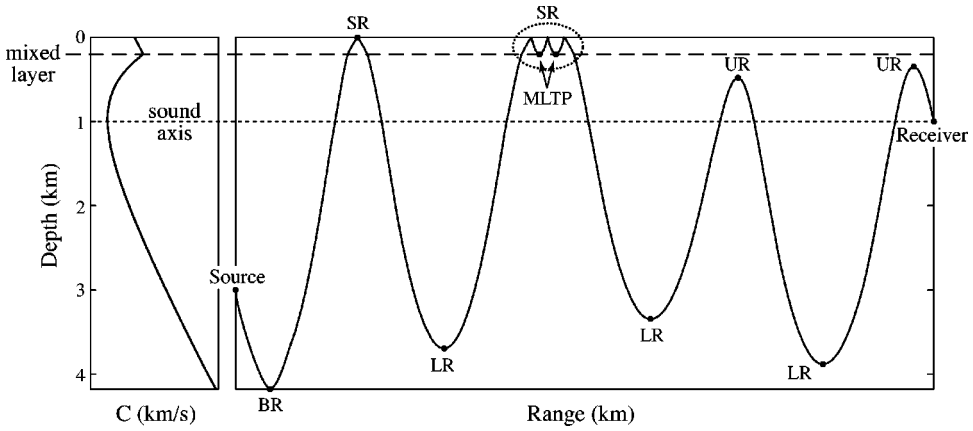


FIG. 11. Definition plot. UTP=SR+UR=2+2 refers to the number of upper turning points, consisting of SR (surface reflected) and UR (upper refracted) turning points. Similarly LTP=BR+LR=1+3 is the number of lower turning points. MLTP=2 is the number of lower turning points in the mixed layer; one or more trapped mixed layer loops counts as a single SR (see figure). For the above ray path, Id: −(UTP+LTP)=−8, %SR=100 SR/UTP=50%, %ML=100 MLTP/(MLTP+LTP)=33%, %BR=100 BR/LTP=25%.

B. Timefronts from an axial source

Figure 10 shows the computed intensity contours for transmission through the total measured field of sound-speed and a smoothed (no spice, no tilt) field (see Figs. 8 and 9). Intensities (down to −30 dB) refer to same maximum in the finale of the center panel. The top panel corresponds to a range-averaged profile of the smoothed field. All the numerical experiments yield the now familiar accordionlike timefronts (Munk and Wunsch, 1979, Fig. 4). The range-averaged field does not differ dramatically from the smooth range-dependent field in spite of the significant latitudinal change in the upper layers along the transmission path. Scattering from finestructure leads to some added complexity in the early arrival pattern. Using the climatological (rather than measured) deep ocean leaves the late arrivals unscattered, unlike the measured pattern of long-range transmissions (Colosi *et al.*, 1994).

We can recognize three distinct phases:

(*i*) the final half second of unresolved arrivals of increasing intensity leading to the abrupt cutoff (the classical SOFAR finale),

(*ii*) the central two seconds of resolved reflected and refracted arrivals which form the basis of tomographic inversions, and

(*iii*) the leading second of steep surface-reflected arrivals with fractured fronts and sensitivity to scattering.

Our emphasis will be on the SOFAR overture, from the earliest arrivals to the transition from reflected to refracted energy. This phase is particularly sensitive to the upper ocean processes under consideration here. For that purpose all subsequent results will be for a source at 3 km which places later arrivals (with turning points beneath 200 m) into a geometric shadow. The ray designations SR, UR, LR,... for surface reflected, upper refracted, lower refracted,... (an unsatisfactory compromise with traditional designations) are defined in Fig. 11. Upper turning points can be surface reflected or refracted, UTP=SR+UR; similarly LTP=BR+LR.

C. Abyssal source

Figure 12 illustrates the principal results of this paper. The four panels in Fig. 12(a) give the intensity fields corresponding to the four sound-speed fields in Fig. 8. The tilt-related scatter does not contribute to the penetration into the mixed layer. The principal effect is to extend the latest timefront into the acoustic shadow and to increase the turning sound-speed, resulting in a somewhat deeper lower turning point and a slightly higher upper turning point. (We remind the reader that internal wave scatter is here restricted to the upper 320 m.)

Spice enhances the intensity of the leading two fractured timefronts, and raises the two upper turning points at 673.5 s from beneath the mixed layer to the surface (see Fig. 13). This sharp modification in turning elevation of the −38 timefront is not accompanied by a significant change in turning sound-speed; accordingly, there is no appreciable lowering of the lower turning point at 674.0 s. However, the preceding SR timefront −36 is lowered from 3.7 to 3.9 km.

The question immediately arises as to the identification of the modified timefronts. Here we resort to the classical method of ray construction (4888 rays were launched at 5 millidegree intervals) with ray arrivals at the receiving array indicated by dots in Fig. 12(b); for comparison, the PE intensities above −30 dB are indicated by the underlying gray bands. Identification follows MWW nomenclature;[4] for example, +35 indicates a upward launch angle with 18 upper and 17 lower turning points. The distance between dots is an indication of intensity (often the dots are so closely spaced as to appear as a line).

We ignore internal reflections; a sharp interface reflects significant energy for glancing incidence at angles less than $\sqrt{\Delta C/C}$. At the bottom of the mixed layers, ΔC of order 1 m/s can occur within a few meters of depth. Yet the agreement between the PE-derived and ray-derived timefronts is generally satisfactory.

In the smooth ocean, the earliest[5] recorded timefront, −33a, is a weak arrival with only 136 rays (2.8%), each consisting of 16 upper and 17 lower turning points. For the

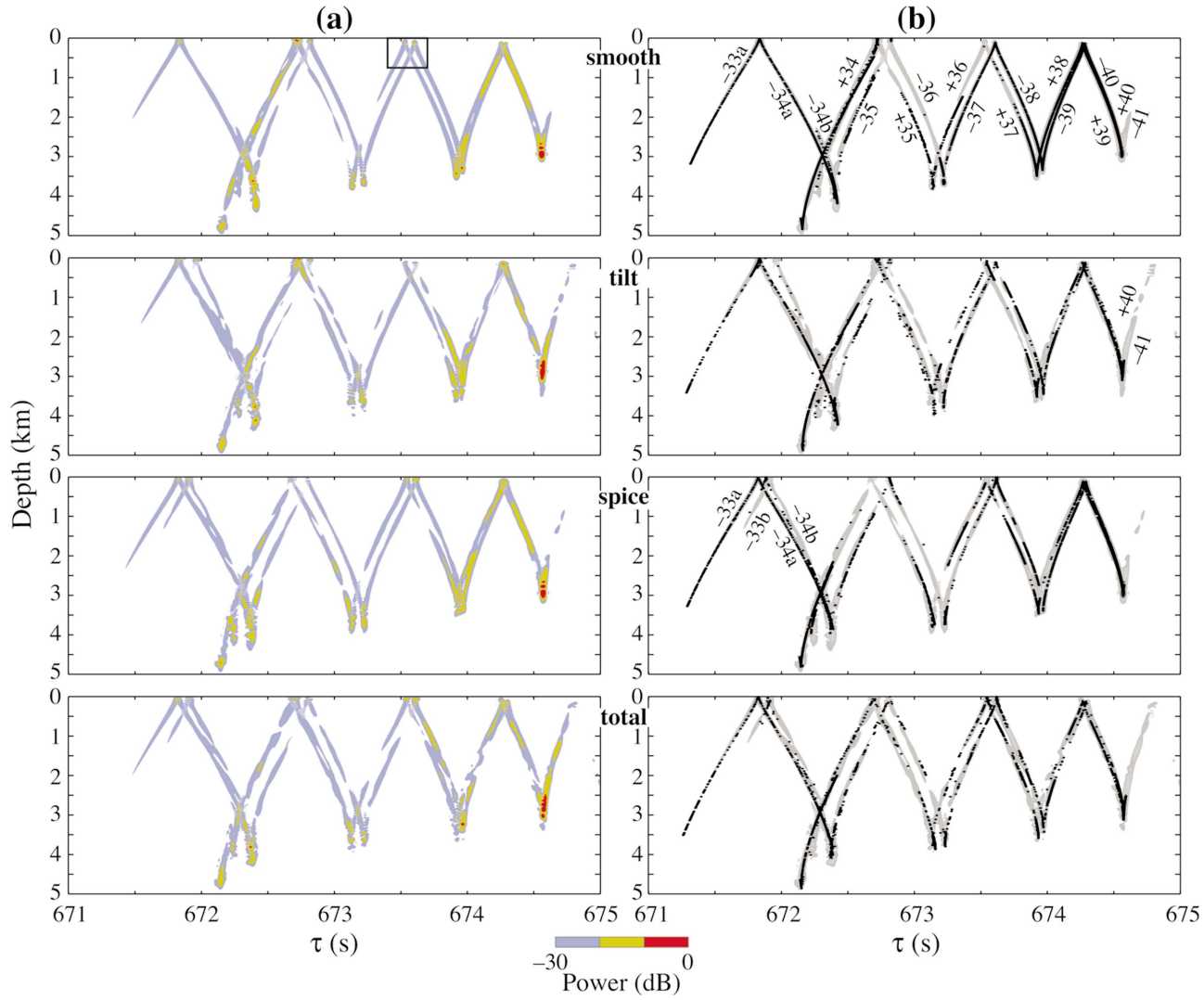


FIG. 12. (a) Computed PE intensity contours relative to the late ''smooth'' arrival from an abyssal source (3 km depth, conjugate to 200 m) at 25 °N to a receiving array at 35 °N for the smooth, spice, tilt (mostly internal waves), and total observed sound-speed profiles in Figs. 8 and 9. An enlarged version of the field within the rectangle is shown in Fig. 13. (b) Dots designate time and depth of ray arrivals at the receiver range, with PE intensities above −30 dB shaded. The ray designator −35 (for example) implies a negative launch angle with 17 upper and 18 lower turning points.

combined 136 rays along the −33a timefront, UTP=136 ×16=2176; all are surface reflected. Of the LTP=136×17 =2312 lower turning points, only 7 (0.3%) are bottom reflected. In the spicy ocean this is followed in 80 ms by a second arrival, −33b, with the same ray identifier −33. The two arrivals have dramatically different turning depths, 4.8 versus 3.9 km. The spicy rays have 5.6% of loops trapped in the mixed layer. The −33b arrivals are extremely weak in the tilt-only ocean.

The transition from surface reflected to upper refracted is abrupt for a smooth range-independent ocean; tilt and spice both extend the time of surface reflections. Rays trapped in the mixed layer are a feature of the transition; they are absent in the earliest and later arrivals.

With regard to the trailing timefronts +40 and −41, we note that the ray arrivals are concentrated at the lower turning points. The extension of intensity (without rays) into the upper ocean must be associated with diffraction. The prominent role by tilt is not understood. Finescale structure in the deep ocean (here neglected) would further extend the trailing timefronts (Colosi, private communication, 2003).

IV. A MODEL FOR THE SOFAR OVERTURE

We wish to interpret the results of the numerical experiment; in particular, the early arrival structure is very complex and sensitive to the formation of mixed layers. Here we look to an analytical model for guidance. The *temperate* profile (also known as the *canonical* profile; see MWW 2.18) is the simplest (but not a simple) representation of the sound channel at temperate latitudes. A global compilation of sound-speed profiles (MWW, Appendix B) exhibits a wide degree of variability, yet the temperate profile provides a rough description for conditions in the central North Pacific. The model is range independent; the resemblance of the upper two panels in Fig. 10 suggests that the principal features of the arrival pattern are retained in a range-averaged representation.

A convenient starting point is the *action variable* (Wunsch, 1987; Brekhovskikh and Lysanov, 1982; MWW, 2.5)

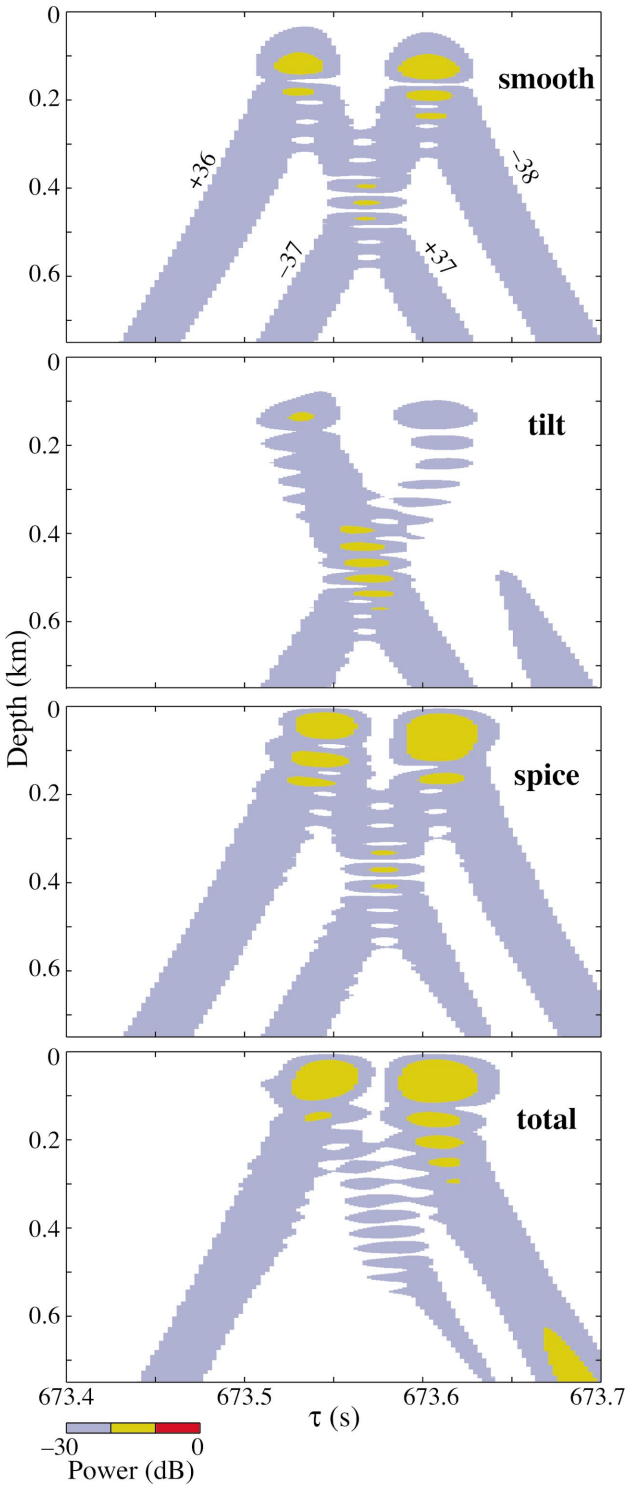


FIG. 13. An enlarged view of the intensity field near the surface within the rectangle in Fig. 12.

$$A(\tilde{S}) = 2\int_{\tilde{z}^-}^{\tilde{z}^+} dz (S^2 - \tilde{S}^2)^{1/2}, \tag{8}$$

where $S(z) = 1/C(z)$ is the sound *slowness*, and $\tilde{S}$ is the slowness at the turning depths $\tilde{z}^{\pm}$. Range and period of a ray loop are conveniently given in terms of the action and its derivative $A' \equiv dA/d\tilde{S}$:

$$R = -A', \quad T = A + R\tilde{S}, \tag{9}$$

with group-slowness and phase-slowness

$$s_{gr} = T/R = \tilde{S} - A/A', \quad s_{ph} = \tilde{S}. \tag{10}$$

The topology of the transition from refraction to surface-reflection is determined by the combined action $A(\tilde{S})$ (Fig. 14) in the sound channel plus the overlying mixed layer (see the Appendix for the cumbersome details). We choose phase versus group slowness for a representation of the dispersion (Fig. 15). The ordinate s_{gr} is readily interpretable as the traveltime $\tau = r s_{gr}$ (from 664 to 667 s); the abscissa $s_{ph} = S(\tilde{z}) \equiv \tilde{S}$ is a measure of turning depths $\tilde{z}^{\pm}$ above and beneath the axis. Alternately, going back to Snell's law, $\tilde{S} = S_{AX} \cos \theta_{AX}$ is a measure of the ray inclination at the axis.

We start with the case of no mixed layer (Fig. 15). Point **A** refers to the axial finale with $s_{ph} = s_{gr} = S_{AX}$. Preceding (refracted) arrivals of increasing steepness are of lower phase-slowness (increasing separation of turning depths from the axis). The dispersion follows a smooth curve up to point **B** where the upper refracted turning point reaches the surface. Steeper rays are surface reflected and of increasing group-slowness until **C** when the situation once again reverses. For arrival times between 665.3 and 665.7 s there are three branches (Fig. 15, top), and this triplication of arrivals accounts for the complexity of the early timefronts. Bottom interactions may cause further modifications (see Fig. 15, top, for the 5-km depth); it is an accident of nature (depending as it does on the temperature profiles and depths of the world oceans) that this interaction occurs in the vicinity of the triplication.

We can interpret the triplication in the terms of the properties of the temperate action profile. Normal dispersion means that steeper rays (smaller $\tilde{S} = s_{ph}$) travel faster (smaller s_{gr}). A transition between normal and abnormal dispersion infers that [using Eq. (10)]

$$s'_{gr} = 1 - \frac{(A')^2 - AA''}{(A')^2} = \frac{AA''}{(A')^2} = 0 \tag{11}$$

or $A'' = 0$, the definition of a caustic. The range of a double-loop is the sum of an upper and lower loop range: $R = R^+ + R^-$. Going from **A** to **B** in the direction of *decreasing* $\tilde{S}$, the loop range *increases* $[dR/d(-\tilde{S}) \equiv -R' = +A'' > 0]$; for the temperate model R increases from 42 km at **A** to a maximum of 53 km at **B**. In this interval the increase of R^- dominates over the decrease of R^+ (the normal situation). At point **B** the upper loop changes from refracted to surface-reflected, followed by a sharp decrease in R^+ dominating the total range R which reaches a minimum at **C**. With further steepening the contribution of the shrinking upper loop to the double-loop range becomes increasingly irrelevant and the lengthening lower loop dominates, restoring normal dispersion. Timefronts at a fixed range and fixed time determine the dispersion at discrete points which may or may not resolve the triplication.

The situation is similar for mixed layers (Fig. 15), with one important difference: a discontinuous decrease in group-slowness when the upper turning point reaches the lower boundary of the mixed layer.

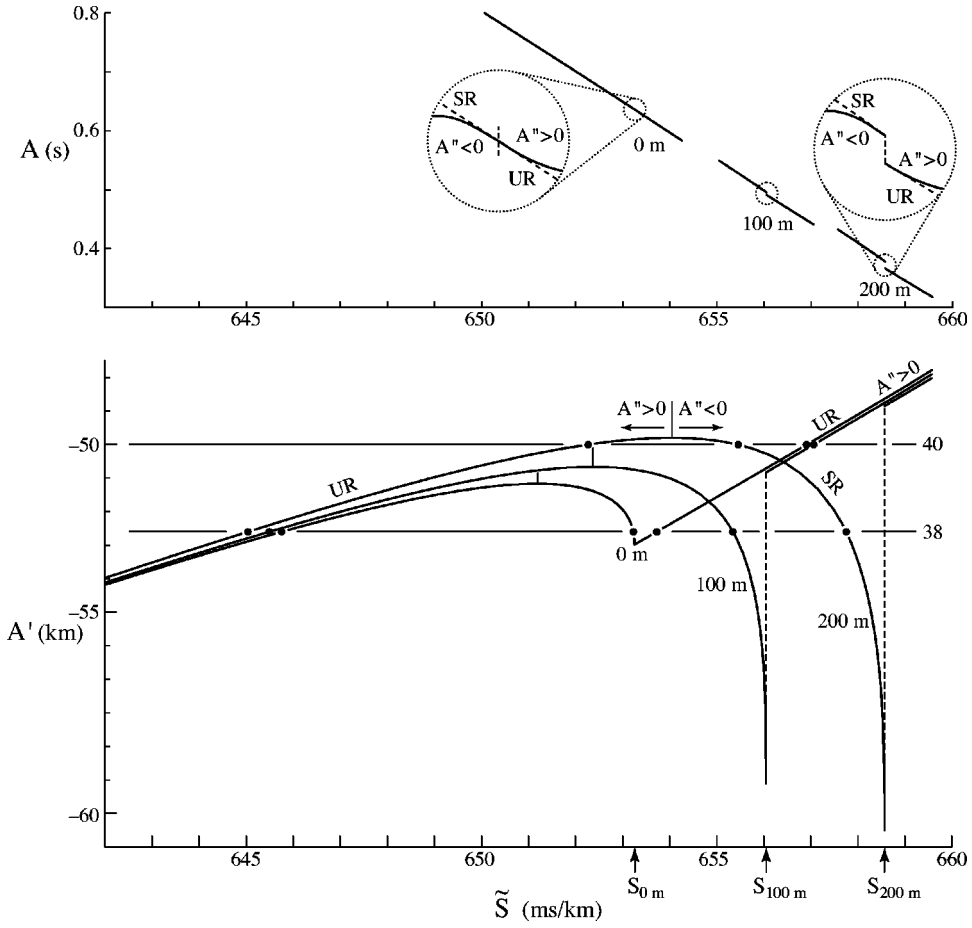


FIG. 14. The action variable A and its derivative $A'=dA/d\tilde{S}$ for the transition from refracted (UR) to surface reflected (SR) propagation, with mixed layer depths of 0, 100 and 200 m, respectively. A and A' are continuous for the case of no mixed layer, and discontinuous otherwise; in all cases A'' changes discontinuously from positive to negative as the ray path changes from UR to SR, and then back to positive (left of the peaks), with profound implications to the arrival structure. Dots indicate eigenrays 38 and 40.

V. THREE FEATURES OF INTEREST

The concentration of energy along the discrete timefronts, a consequence of Fermat's principle, is well known but nonetheless noteworthy: more than 95% of τ, z-space (Fig. 12) is at less than the -30 dB level, and more than 95% of the emitted rays are concentrated along the timefronts. Scattering processes lead to only a slight broadening of the timefronts, and to a slight bias (to be treated in a separate paper).

A further concentration is achieved by subjecting the timefronts to the turning-point filter TPF (Dzieciuch *et al.*, 2001):

$$P(\theta_{AX},\tau_{AX})=\sum_i [p_i(\tau_{AX}+\Delta\tau(z_i,\theta_{AX})], \tag{12}$$

where

$$\Delta\tau(z_i,\theta_{AX})=\int_{z_A}^{z_i} dz\sqrt{S^2(z)-S_A^2\cos^2\theta_{AX}} \tag{13}$$

is the time delay relative to the axial arrival τ_{AX} for a receiver at elevation z_i, assuming an axial inclination θ_{AX}. [At constant slowness this reduces to the linear beamformer $\Delta\tau=(z_i-z_{AX})S\sin\theta$.] The filtered intensity is plotted in s_{ph}, s_{gr}-space, with $s_{ph}=S_{AX}\cos\theta_{AX}$ and $s_{gr}=\tau_{AX}/r$ (Fig. 16). This is our preferred space for the acoustic monitoring of ocean processes.

There are many features to be accounted for. We shall consider three features in particular; they appear in all presentations but are highlighted in the three boxes of the TPF outputs.

Diffraction. Late refracted arrivals $+40$, -41 have PE intensities but no associated rays beyond the lower turning point (Figs. 12(a) and (b)]. The feature is more prominent at 50 Hz than at 200 Hz, as one would expect. All this is consistent with diffraction into the geometric shadow. In the real ocean with deep finestructure the late scattered arrivals are even further extended (Dushaw *et al.*, 1999), but the high intensity of the shadow arrivals has not been accounted for.

Orbit splitting. The early timefronts 33, 34 are split into multiple fronts with the same ray identifier but significantly different turning slowness; we refer to this as having different ray *orbits*. Orbit splitting is a manifestation of the triplication associated with the transition from surface reflection to refraction. The split is particularly pronounced for -34a,b with lower turning depths of 4.36 and 3.92 km, respectively. Timefront -34a is almost entirely surface reflected, whereas -34b has a significant mixed layer component. We associated -34a and -34b with the lower and central branches of the triplication (Fig. 17). Note that -34b is absent in the tilt

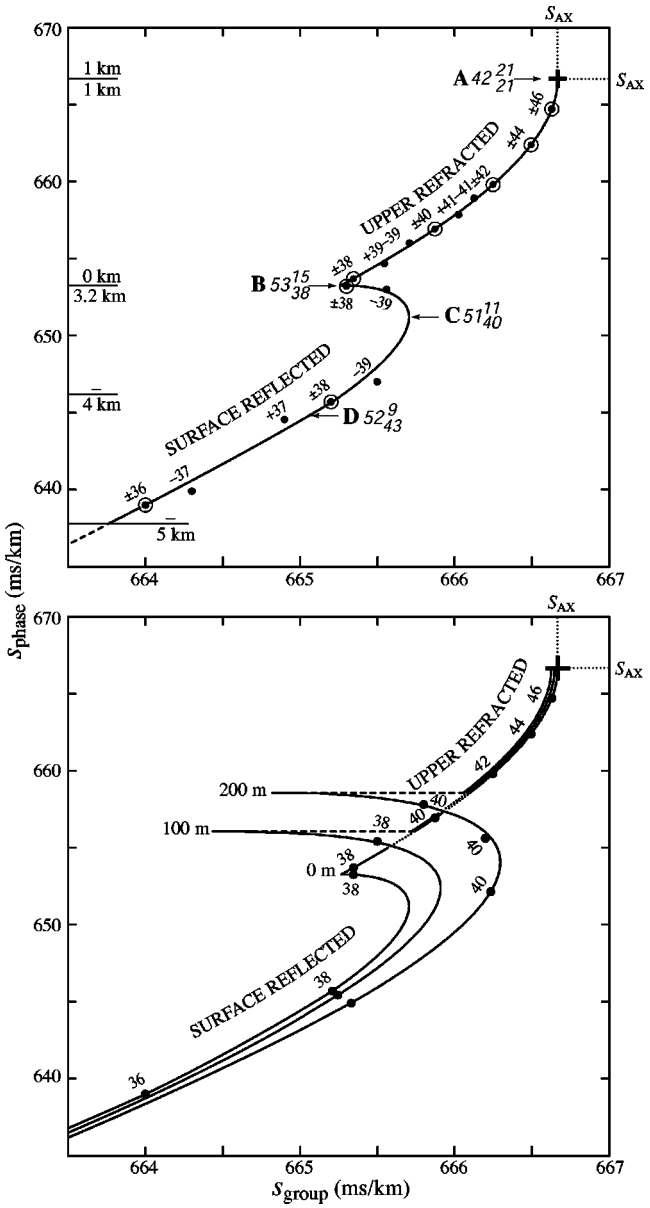


FIG. 15. Top: Dispersion for temperate profile for refracted (**A** to **B**) and surface reflected (**B** to **C** to **D**) rays. The dispersion is "normal" ($A''>0$) from **A** to **B** and from **C** to **D**, and abnormal ($A''<0$) from **B** to **C**. Numbers following **B** denote a 53 km loop length, consisting of a 15 km upper loop and 38 km lower loop, and similarly for **A**, **C**, **D**. Even identifiers ± 38, ± 40 designate upward/downward launch angles with equal numbers 19/19, 20/20 of upper/lower turning points; odd identifiers $+39$, -39 designate 20/19, 19/20 upper/lower turning points and lying slightly beneath/above the "even" dispersion curve. Some upper/lower turning depths are indicated; for an ocean depth of 5 km, the dispersion encounters bottom interaction near the lower limit (dashed). Bottom: Dispersion for 0, 100, and 200 m mixed layer depths. Refracted and surface-reflected dispersions are discontinuous for mixed layers.

ocean but clearly identified in the spice ocean.

The TPF representation (Fig. 16) exhibits $+34$ splitting and a weak and deep $+33$ arrival with 100% surface reflection and no mixed layer trapping. We detect a cluster of shallow $+33$ ray points associated with 36% surface reflection and 38% mixed layer interaction. Finally, some weak early ray arrivals are identified with -32 timefronts, 52% surface reflection and 42% mixed layer interactions.

The simulation clearly supports orbit splitting but does not convincingly demonstrate the expected triplication.

Trajectory splitting. The situation is very different for $+36$. Here we find multiple configurations of ray arrivals with no appreciable variation in the orbital turning slowness and lower turning depth. We are dealing here with a ray whose upper turning point is very near the bottom of the mixed layer, so that a tiny steepening leads to a penetration of the mixed layer and an upward refraction towards the surface. The extra loop is associated with a time delay (marked B in Fig. 16). This is more clearly seen in a traditional ray diagram (Fig. 17): ray trajectories $+36$ all turn beneath the mixed layer up to about mid-range where they encounter the sharp transition; here 20% of the rays (labeled **B**) are diverted to the surface for a single reflection. The remainder (**A**) retain their upper refracted turning points.[6] (Down-range at 800 km all $+36$ rays become surface reflected.) Some early evidence of trajectory splitting was found in a 900-km transmission in the northwest Atlantic (Brown *et al.*, 1980).

We conclude that trajectory splitting is along the s_{gr} axis (time axis), whereas orbital splitting has a major s_{ph} (depth) component. The triplication is an intrinsic feature of the SOSUS *overture* and needs to be taken into account in the acoustic interpretation of upper ocean processes. This offers interesting opportunities; for example, perturbations in travel times are of opposite sign in the central branch ($A''<0$) rather than in the other branches, suggesting *differential* times as sensitive measures of internal wave and spicy frontal activities. The perturbations are subtle and much is yet to be learned.

VI. DISCUSSION

The abrupt finale of SOFAR transmissions has been the subject of an extensive literature ever since its discovery in 1944 by Ewing and Worzel (1948); in the words of the authors (p. 4), "the end of the sound channel transmission was so sharp that it was impossible for the most unskilled observer to miss it." (We have been trying to find an objective specification ever since.) They also note (p. 8) that "a fair estimate of the distance between source and receiver can be obtained...from the over-all sound channel duration... ."

Kuperman *et al.* (2001) were the first to point out that there is also a definitive start time.[7] The physics of the overture and finale are altogether different. The former depends on the transition from refraction to surface reflection and is very sensitive to mixed layer formation and scattering; the latter depends on details of the axial sound-speed profile and is sensitive to internal wave scattering. An observation of trapping in the mixed layer goes back to Worcester (1977). A demonstration of the seasonal transition was provided by a 48-day tomography transmission in the northwest Atlantic in the fall of 1978 (Brown *et al.*, 1980). Sutton *et al.* (1997) recorded deepening of the mixed layer using acoustic modes. Simmen *et al.* (1999) have studied the arrivals associated with a 270 km transmission in April 1990 off Florida; they associate a folding of the early timefront with a complex upper ocean profile quite independent of surface reflection.

The sensitivity of the overture to upper ocean processes suggests an acoustic monitoring of seasonal (and other time)

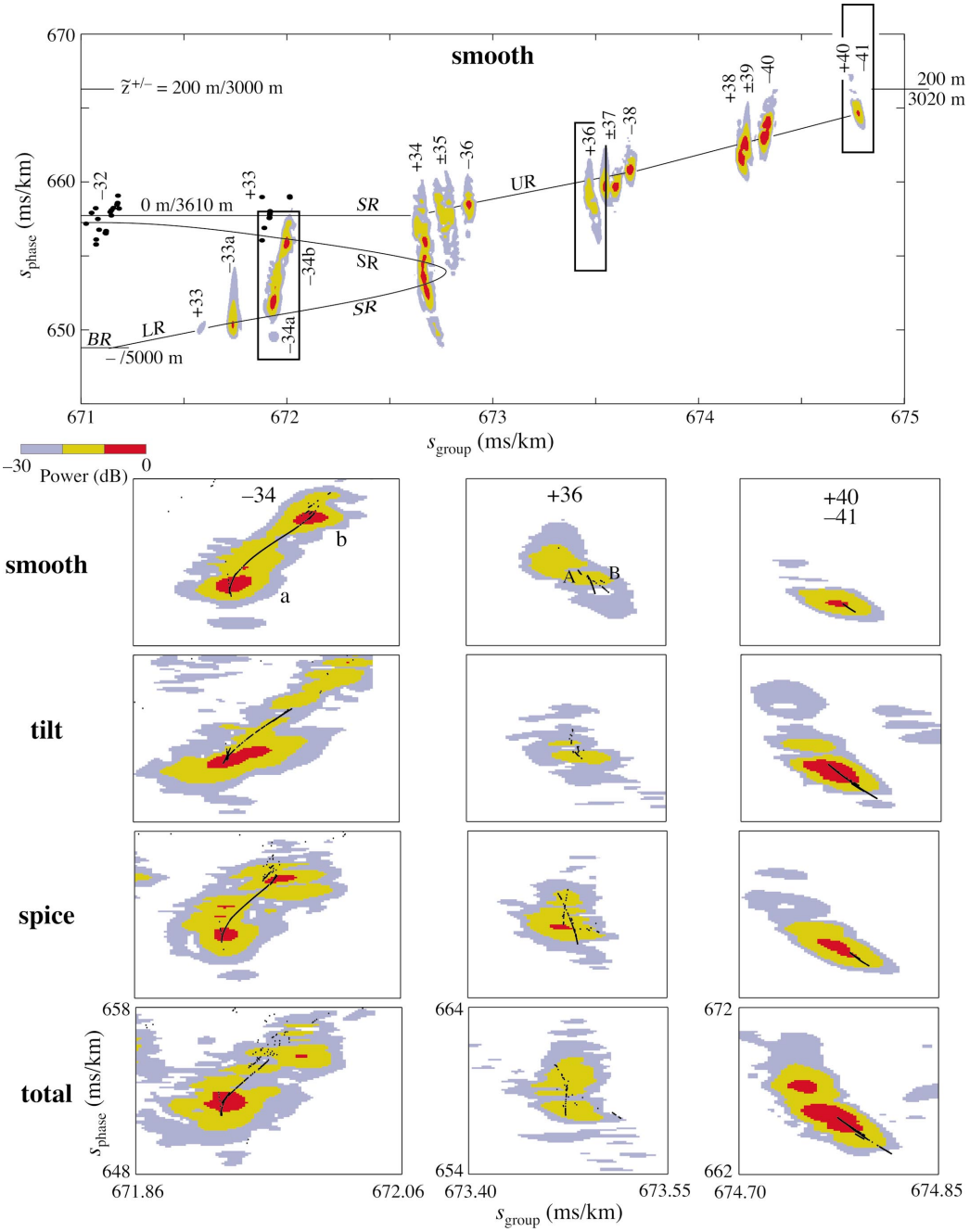


FIG. 16. Dispersion in slowness space, corresponding to $\tau, \tilde{z}^{\pm}$ space. In the lower ocean the turning depth varies from $\tilde{z}^{-}=5000$ to 3000 m, as indicated in the upper ocean from about $\tilde{z}^{+}=200$ m downward to the surface. Acoustic power (color contours) and ray arrivals (dots) define the dispersion relation, with a triplication at the transition from upper refracted (UR) to surface reflected (SR) arrivals. The three rectangles (expanded in the lower three columns) illustrate the cases of orbital splitting, trajectory splitting, and diffracted arrivals (see text).

variability of upper ocean dynamics. The problem is to determine the temporal variability of a measured mean state. (Seismologists are not so fortunate; they are confined to an invariable and unaccessible mean state.). A thin, wind-stirred layer caps much of the ocean area during winter and spring. The layer is remarkably uniform in density, with fractional differences between top and bottom of order 10^{-5}. At the same time it is populated with spice fronts across which the speed of sound changes abruptly by 10^{-3} parts. Photosynthesis in the mixed layer plays a major role in the biological productivity of the entire ocean environment. This flimsy surface film (typically 100 m out of 5000 m) puts its peculiar signature upon the acoustic transmission within the entire water column, offering the opportunity of ''viewing'' the important near-surface processes in the quiet of the deep ocean.

The standard procedure has been to place the source on the axis; this offers the entire record, from overture to finale. For the study of the early surface-interacting arrivals one can place the source at great depth, as we have in this paper. The deep ocean offers a benign environment of near-uniform temperatures and weak currents; a bottom-moored source permits short cables with small watch circles. Depth perturbations associated with orbit changes are more easily de-

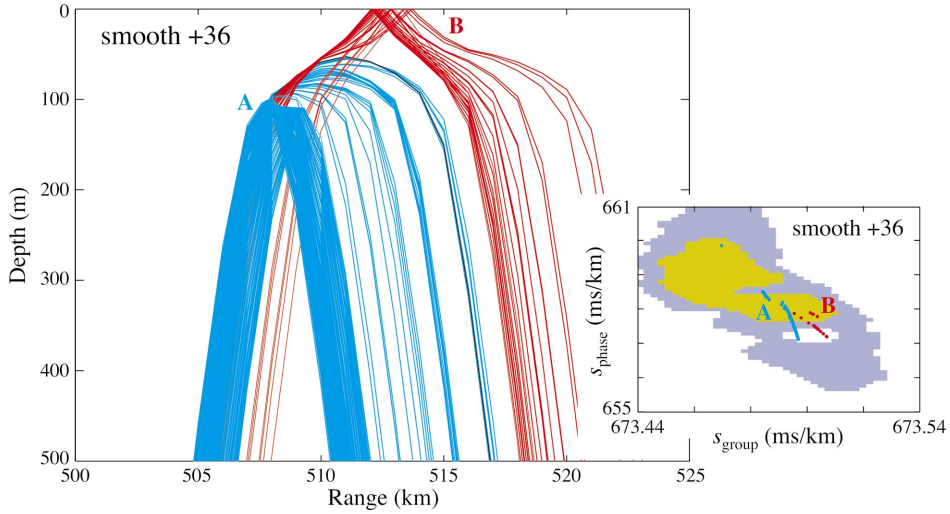


FIG. 17. Ray trajectories +36 all turn beneath the mixed layer up to mid-range, where they encounter a sharp transition: ray bundle **A** continues to turn beneath the surface, whereas **B** becomes surface reflected. The bundles are associated with two ray clusters in the dispersion (inset), an enlargement of the +36 panel in Fig. 16.

tected in the abyssal ocean; vertical displacements are inversely proportional to

$$dC/dz \approx C_A \gamma_a (N^2 - N_{AX}^2)/N_{AX}^2 \tag{14}$$

or roughly in the ratio $(N_{\text{axis}}/N_{\text{surface}})^2 = 10$ to 100 between bottom and top.

In this paper we had to rely on a *computed* sound intensity field transmitted through a *measured* sound-speed field. Plans for a 2004 experiment provide for coordinated measurements of both fields, with emphasis on the lower caustic. There is of course nothing new in monitoring upper ocean events at conjugate depths. It has been the basis of submarine detection for half a century.

ACKNOWLEDGMENTS

This paper is based largely on the application of the parabolic equation method, which was introduced into acoustics by Fred Tappert (Hardin and Tappert, 1973; Tappert, 1977). We acknowledge his influence here as well as on many previous occasions. (Tappert died while this paper was in preparation.) P. Worcester, W. Kuperman, and J. Colosi have made helpful suggestions. B. Betts and A. Santos helped prepare the manuscript. M.D. is supported by ONR Grant Nos. N00014-97-1-0258 and N00014-03-1-0182. W.M. holds the Secretary of the Navy Chair in Oceanography and is supported by ONR Grant No. N00014-03-1-0613. DLR is supported through NSF Grant Nos. OCE98-19521 and OCE00-02598.

APPENDIX: TEMPERATE PROFILE WITH SURFACE MIXED LAYER

We follow the notation in MMW 2.17 modified to include mixed layers and surface reflection. The slowness profile consists of a temperate sound channel with axis at depth $h_{SC} = 1$ km capped by an adiabatic mixed layer of thickness h_{ML} (Fig. 18). The temperate profile exhibits the up/down asymmetry of the measured profile (left panels of Fig. 10), and approaches the adiabatic gradient

$$dS/dz = S\gamma_a, \quad \gamma_a = 0.0114 \ \text{km}^{-1} \tag{A1}$$

at great depth.

The total action variable has three components:

$$A(\tilde{S}) = A_{SC}^{+} + A_{SC}^{-} + A_{ML}, \tag{A2}$$

where

$$A_{SC}^{\pm}(\tilde{S}) = S_{AX} h_{SC} \Gamma [a(1 - f_a^{\pm})\tilde{\phi}^2 \pm b(1 - f_b^{\pm})\tilde{\phi}^3 + c(1 - f_c^{\pm})\tilde{\phi}^4] \tag{A3}$$

for the upper/lower sound channel, with

$$\tilde{\phi}(\tilde{S}) = \Gamma^{-1}\sqrt{1 - (\tilde{S}/S_{AX})^2} = \Gamma^{-1} \sin\theta_{AX}, \quad \Gamma = \sqrt{\gamma_a h_{SC}}, \tag{A4}$$

$$a = \frac{\pi}{2\sqrt{2}}, \quad b = -\frac{2}{9}, \quad c = \frac{\pi}{48\sqrt{2}}. \tag{A5}$$

In the lower sound channel $f_{a,b,c}^{-} = 0$. In the upper sound channel

$$f_a^{+}(\sigma_{SC}) = (2\sigma_{SC} - \sin 2\sigma_{SC})/\pi,$$
$$f_b^{+}(\sigma_{SC}) = \sin^3 \sigma_{SC}, \tag{A6}$$
$$f_c^{+}(\sigma_{SC}) = (4\sigma_{SC} - \sin 4\sigma_{SC})/2\pi,$$

$$\sigma_{SC}(\tilde{S}, S_{ML}) = \sin^{-1}\left(\frac{S_{ML}^2 - \tilde{S}^2}{S_{AX}^2 - \tilde{S}^2}\right)^{1/2}, \quad \tilde{S} < S_{ML} = 0, \quad \tilde{S} > S_{ML} \tag{A7}$$

with

$$S_{ML} = S_{SU}(1 - \gamma_a h_{ML}) \tag{A8}$$

designating the slowness at the *bottom* of the mixed layer, S_{SU} being the surface slowness. For the rays with a refracted upper turning point, $\tilde{S} > S_{ML}$ hence $\sigma_{SC} = 0$ and $f_{a,b,c}^{+} = 0$; we return to the previous case of pure refraction. For the case $S_{ML} = S_{AX}$, $\sigma = 1/2\pi$ and $f_{a,b,c}^{+} = 1$: there is no upper loop.

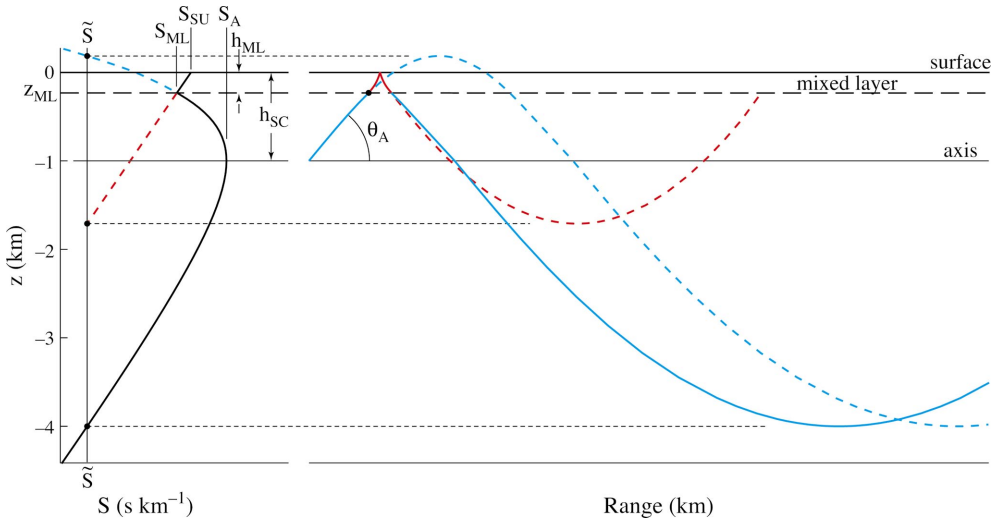


FIG. 18. Canonical sound-speed profile with adiabatic surface mixed layer. The upward refracted surface-reflected ray (blue–red–blue, solid) is drawn for a lower turning depth of 4 km with slowness $\tilde{S}$. The downward refracted ray through an upward extension of the canonical profile into the mixed layer and across the surface (blue, dashed) turns again at $\tilde{S}$, and so does the upward refracted ray through a downward extension of the adiabatic mixed layer (red, dashed).

For the case $\tilde{S}<S_{ML}$ there is a fractional upper loop within the sound channel (Fig. 18) and this is accommodated by "fractional action" as determined by $0<f^{+}_{a,b,c}<1$. The upper turning point lies above the surface; the turning slowness is determined by an upward extension of $S_{SC}(z)$ or a downward extension of $S_{ML}(z)$ as shown in Fig. 18.

For the mixed layer (MWW, 103–106)

$$A_{ML}=\frac{2}{3}\gamma_a^{-1}S_{SU}\frac{S_{SU}^2-\tilde{S}^2}{S_{SU}^2}(1-\sin^3\sigma_{ML}),$$
$$\sin\sigma_{ML}(\tilde{S},S_{ML})=\sqrt{\frac{S_{ML}^2-\tilde{S}^2}{S_{SU}^2-\tilde{S}^2}} \tag{A9}$$

for $\tilde{S}<S_{ML}$ and $A_{ML}=0$ otherwise. For a thinning mixed layer $h_{ML}\to 0$ we have $S_{ML}\to S_{SU}$ and $A_{ML}\to 0$; surface reflection is accommodated by $f^{+}_{a,b,c}(\sigma_{SC}(\tilde{S},S_{SU}))$ for $\tilde{S}<S_{SU}$.

[1]"Conjugate" refers to depths of equal sound-speed above and beneath the axis of minimum sound-speed.
[2]Spice ($\mu\pi\alpha\chi\alpha\rho\iota\kappa o$ in Greek) is denoted by μ.
[3]Usually denoted by $R=-1/n$.
[4]For Munk, Worcester, and Wunsch (1995).
[5]An even earlier timefront, +33, is detected by the turning point filter (Fig. 16).
[6]This may account for the puzzling structure of the +36 timefront (Fig. 12). The smooth timefront shows densely packed rays (**A**) but no contoured intensity between 1500 and 3000 m (the reverse holds above 1500 m). Near the turning point between 3300 and 3800 m there are both rays (**B**) and contoured intensity, the usual situation. We suspect that the anomalous occurrence of a dense ray configuration with no PE intensity is because ray theory misses the diffraction effects involved with the turning of acoustic energy near an interface. Note that the pattern is altered in the presence of scattering.
[7]The existence of both start and end times permits the estimate of range from a single station; Kuperman *et al.* (2001) attained an astounding precision of 100 km for a 3500-km transmission. A simple rule of thumb follows from the structure of ray arrivals (MWW, p. 48): the interval between ray arrivals is independent of range, but the number of ray arrivals increases linearly with range. Referring to the total record in Fig. 12, a single receiver at 200-m (or 3 km) depth receives five major pulses at roughly 1-s intervals (each pulse consists of multiple subpulses). We have $r=n^{+}R^{+}+n^{-}R^{-}$ where $n^{\pm}$ is the number of upper/lower loops of range $R^{\pm}$, respectively. For the case of an equal number of upper and lower loops, $r=nR$ where $n=n^{+}=n^{-}$ is the number of double-loops and $R=R^{+}+R^{-}$ is the double-loop range. The sound-speed profile at the receiver is consistent with $R_1=60$ km and $R_2=53$ km for the lengths of the ray double-loops with turning points at the surface and 200 m, respectively. From $r=n_1R_1=n_2R_2$ it follows that $r=(n_2-n_1)\Delta$, $\Delta=(R_2^{-1}-R_1^{-1})^{-1}=454$ km. There are two major pulses for each double-loop arrival, and we estimate $r=5/2\ 454=1135$ km. The estimates can be refined by a subsequent matched filter analysis. The above procedure assumes a source beneath 200 m.

Brekhovskikh, L., and Lysanov, Y. (**1982**). *Fundamentals in Ocean Acoustics* (Springer-Verlag, Berlin).
Brown, M. G., Munk, W. H., Spiesberger, J. L., and Worcester, P. F. (**1980**). "Long-range acoustics in the northwest Atlantic," J. Geophys. Res. **85**, 2699–2703.
Collins, M. D. (**1989**). "A higher-order parabolic equation method for wave propagation in an ocean overlying an elastic bottom," J. Acoust. Soc. Am. **86**, 1459–1464.
Colosi, J. A., Flatté, S. M., and Bracher, C. (**1994**). "Internal-wave effects on 1000 km oceanic acoustic pulse propagation: Simulation and comparison with experiment," J. Acoust. Soc. Am. **96**, 452–468.
Colosi, J. A., the ATOC Group: Scheer, E., Flatté, S. M., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999**). "Comparison of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific," J. Acoust. Soc. Am. **105**, 3202–3218.
Dushaw, B. D., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1999**). "Multimega-meter Range Acoustic Data Obtained by Bottom-Mounted Hydrophone Arrays for Measurement of Ocean Temperature," IEEE J. Ocean. Eng. **24**, 202–214.
Dzieciuch, M. (**1993**). "Numerical solution of the acoustic wave equation using Chebyshev polynomials with application to global acoustics," Proceedings of the IEEE Oceans '93 (IEEE, Piscataway, NJ), Vol. 1, pp. 267–271.
Dzieciuch, M., Worcester, P., and Munk, W. (**2001**). "Turning point filters: Analysis of sound propagation on a gyre-scale," J. Acoust. Soc. Am. **110**, 135–149.
Ewart, T. E. (**1980**). "A numerical simulation of the effect of oceanic fine-structure on acoustic transmission," J. Acoust. Soc. Am. **67**, 496–503.
Ewart, T. E., Irish, J. D., and Olsen, L. O. (**1977**). "The MATE experiment, a preliminary report," Internal Report (University of Washington, Applied Physics Laboratory).

Ewing, M., and Worzel, J. L. (**1948**). ''Long-range sond transmission,'' Geol. Soc. Am. Memoir **27**, 1–35.
Ferrari, R., and Rudnick, D. (**2000**). ''Thermohaline variability in the upper ocean,'' J. Geophys. Res. **105**, 16857–16883.
Flattè, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean* (Cambridge U.P., Cambridge).
Garrett, C. J. R., and Munk, W. H. (**1971**). ''Internal waves in the presence of finestructure,'' J. Phys. Oceanogr. **1**, 196–202.
Hardin, R. H., and Tappert, F. D. (**1973**). ''Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations,'' SIAM Rev. **15**, 423.
Kuperman, W. A., D'Spain, G. L., and Heaney, K. D. (**2001**). ''Long range source localization from single hydrophone spectrograms,'' J. Acoust. Soc. Am. **109**, 1935–1943.
Levine, M. D., and Irish, J. D. (**1981**). ''A statistical description of temperature in the presence of internal waves,'' J. Phys. Oceanogr. **11**, 676–691.
Levine, M. D., and Irish, J. D. (**1986**). ''Simultaneous spatial and temporal measurements of the internal wave field during MATE,'' J. Geophys. Res. **91**, 9709–9719.
Levitus, S. (**1994**). *World Ocean Atlas, Vol. 4 Temperature*, NOAA Atlas NESDIS No. 4 (NOAA, Washington, DC).
Levitus, S., Burgett, R., and Boyer, T. (**1994**). *World Ocean Atlas, Vol. 3 Salinity*, NOAA Atlas NESDIS No. 3 (NOAA, Washington, DC).
Munk, W. (**1981**). ''Internal Waves and small-scale processes,'' in *Evolution of Physical Oceanography—Scientific Surveys in Honor of Henry Stommel*, edited by B. Warren and C. Wunsch (MIT, Boston), pp. 264–291.
Munk, W., Worcester, P., and Wunsch, C. (**1995**). *Ocean Acoustic Tomography* (Cambridge U.P., Cambridge). To be designated MWW.
Munk, W. H., and Wunsch, C. (**1979**). ''Ocean acoustic tomography: a scheme for large scale monitoring,'' Deep-Sea Res., Part A **26**, 123–161.
Munk, W. H., and Zachariasen, F. (**1976**). ''Sound propagation through a fluctuating stratified ocean: Theory and Observation,'' J. Acoust. Soc. Am. **59**, 818–838.
Rudnick, D. L., and Ferrari, R. (**1999**). ''Compensation of horizontal temperature and salinity gradients in the ocean mixed layer,'' Science **283**, 526–529.
Rudnick, D. L., and Martin, J. P. (**2002**). ''On the horizontal density ratio in the upper ocean,'' Dyn. Atmos. Oceans **36**, 3–21.
Schmitt, R. W. (**1981**). ''Form of the temperature-salinity relationship in the central water: Evidence for double-diffusive mixing,'' J. Phys. Oceanogr. **11**, 1015–1026.
Schmitt, R. W. (**1994**). ''Double diffusion in oceanography,'' Annu. Rev. Fluid Mech. **26**, 255–285.
Simmen, J. A., Flatté, S. M., DeFerrari, H. A., Nguyen, H., and Williams, N. J. (**1999**). ''Near-caustic behavior in a 270-km acoustical experiment,'' J. Acoust. Soc. Am. **105**, 3231–3244.
Stommel, H. (**1993**). ''A conjectural mechanism for determining the thermohaline structure of the oceanic mixed layer,'' J. Phys. Oceanogr. **23**, 142–158.
Sutton, P. J., Morawitz, W., Worcester, P., and Cornuelle, B. (**1997**). ''Temperature evolution of the upper ocean in the Greenland Sea January to March 1989,'' J. Geophys. Res. **102**(C13), 27861–27874.
Tappert, F. D. (**1977**). ''The parabolic approximation method,'' in *Wave Propagation and Underwater Acoustics*, edited by J. Keller and J. Papadakis (Springer-Verlag, Berlin), pp. 224–287.
Worcester, P. F. (**1977**). ''Reciprocal acoustic transmissions in a mid-ocean environment,'' J. Acoust. Soc. Am. **62**, 895–905.
Wunsch, C. (**1987**). ''Acoustic tomography by Hamiltonian methods including the adiabatic approximation,'' Rev. Geophys. **25**, 41–53.
Young, W. R. (**1994**). ''The subinertial mixed layer approximation,'' J. Phys. Oceanogr. **24**, 1812–1826.
Yuan, X., and Talley, L. D. (**1992**). ''Shallow salinity minima in the North Pacific,'' J. Phys. Oceanogr. **22**, 1302–1316.

Resonance frequency of microbubbles: Effect of viscosity

Damir B. Khismatullin[a)]
Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708

(Received 4 January 2004; revised 20 April 2004; accepted 14 June 2004)

The transmitted frequency at which a gas bubble of millimeter or submillimeter size oscillates resonantly in a low-viscosity liquid is approximately equal to the undamped natural frequency (referred to as the Minnaert frequency if surface tension effects are disregarded). Based on a theoretical analysis of bubble oscillation, this paper shows that such an approximation cannot be validated for microbubbles used in contrast-enhanced ultrasound imaging. The contrast-agent microbubbles represent either encapsulated bubbles of size less than 10 μm or free (nonencapsulated) bubbles of submicron size. The resonance frequency of the microbubbles deviates significantly from the undamped natural frequency over the whole range of microbubble sizes due to the increased viscous damping coefficient. The difference between these two frequencies is shown to have a tremendous impact on the resonant backscatter by the microbubbles. In particular, the first and second harmonics of the backscattered signal from the microbubbles are characterized by their own resonance frequencies, equal to neither the microbubble resonance frequency nor the undamped natural frequency. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1778835]

PACS numbers: 43.35.Ei, 43.80.Qf, 43.35.Bf [AJS] Pages: 1463–1473

I. INTRODUCTION

Due to its safety and low cost, ultrasound scanning, or echosonography, is actively used in obstetrics and gynecology. However it loses other imaging modalities in regard to the image quality. An excellent way to increase the quality of sonograms is the use of micron-size bubbles as ultrasound contrast agents,[1–3] i.e., as scatterers of ultrasound waves inside the human body. When injected into the bloodstream, the microbubbles, less than 10 μm in size, move with blood cells toward the measuring site. Once reaching the blood vessel exposed to ultrasound, they break into oscillation and thus scatter the incoming ultrasound waves in all directions. Due to the large acoustic impedance difference at the interface between the gas and blood, scattering by microbubbles is of high intensity and results in a strong signal at the receiver, i.e., it enhances the echo from blood. Moreover, because microbubbles are nonlinear scatterers, the backscattered signal from them contains higher harmonics of the transmitted frequency. This has made possible the contrast harmonic imaging technique,[4,5] which, in the Doppler mode, can image blood flow in smallest blood capillaries, allowing quantification of myocardial perfusion.[3]

Although the overall effect of contrast agents on the backscattered signal can be detected easily, it is very difficult to measure oscillations of an individual microbubble in the population situated within the insonated blood vessel. The microbubble behavior can be recorded by optical imaging with a frame rate of several MHz. Such fast-speed imaging of contrast agent microbubbles has been reported by de Jong and co-workers.[6] This group of scientists measured the radius-time curves for microbubbles but did not relate these results to the contribution of each microbubble to the backscattered signal. In the absence of experimental data, we can only rely upon theory to realize how a given microbubble interacts with ultrasound. Particularly, a theoretical analysis of microbubble dynamics can be used (a) to evaluate the transmitted frequency at which scattering by a microbubble of a given size reaches the maximum and (b) to determine the intensity (scattering cross-section) and damping rate of the scattered wave produced by the microbubble. The acoustic response of the microbubble population located near the measuring site can be calculated from these functions of bubble size, if the size distribution of microbubbles in the population is known or has been measured.

All theoretical works concerning oscillations of ultrasound contrast agents are based on the Raleigh–Plesset equation.[7] A number of corrections to this model have been made, including the effects of the encapsulating shell,[8,9] blood compressibility[10] and viscoelasticity,[11] and interaction between microbubbles.[12] Hilgenfeldt *et al.*[13] analyzed theoretically the effects of nonlinearity and polychromaticity (pulsed driving) on the resonant backscatter by free microbubbles. Except for Refs. 11 and 13, these works analyze backscattering by contrast agents from the viewpoint of large bubbles. If a bubble is covered with a high-viscosity shell and its size is less than 10 μm, its resonance frequency differs significantly from its natural frequency due to the increased viscous damping coefficient.[11] For a free bubble, the shift in resonance frequency due to damping effects were discussed by Leighton[14] and Brennen.[15] As shown here, the difference between these two frequencies has a tremendous impact on backscattering by ultrasound contrast agents. For example, it is believed[1] that contrast-enhanced ultrasound provides images with the highest quality if it operates at the frequency at which the major part of microbubble population resonates. To put it differently, if the backscattered signal has a peak at the transmitted frequency or at multiple of this frequency, it comes from bubbles which resonate at this fre-

a)Electronic mail: damir@duke.edu

quency. However, we will demonstrate that *the frequency at which the signal scattered by a microbubble has a resonance peak is not equal to the resonance frequency of the microbubble*.

The commercially available ultrasound contrast agents can be divided into two different types: (i) suspensions of encapsulated microbubbles and (ii) emulsions of liquid dodecafluoropentane (DDFP) and superheated drops. Encapsulated microbubbles are coated by a polymer,[1] protein,[1] or lipid layer[16] with the aim to increase stability of microbubbles in blood flow. The thickness and mechanical properties (viscosity and elasticity) of the layer differ for different contrast agents. An emulsion of liquid dodecafluoropentane[17] boils at 28 °C, i.e., it becomes a suspension of free (non-encapsulated) microbubbles after intravenous injection. A superheated emulsion[18] also transforms itself into a suspension of free microbubbles but under the action of ultrasound. In this paper, the resonant properties of both free and encapsulated microbubbles are studied theoretically.

II. ENCAPSULATED MICROBUBBLES

Free gas microbubbles dissolve away in blood very quickly. This is caused by the over-pressure in the gas inside the bubble compared to the pressure in blood due to surface tension at the gas-blood interface.[19] Encapsulation of the microbubble by a biocompatible surface-active layer lowers the over-pressure, and, as a result, makes the microbubble more stable against dissolution. The increased stability is the reason why the great bulk of available contrast agents represent encapsulated microbubbles. However, the encapsulating shell inevitably constrains the bubble oscillation due to the shell viscosity, which is much higher than the viscosity of blood plasma, and thus results in decreasing the amplitude of the backscattered signal. The shell-induced damping of the oscillation becomes more pronounced with increasing the shell thickness.[11]

The encapsulating shell shows viscoelastic properties. Elasticity of the shell material arises from changes in configurations of its macromolecules under flow.[20] The viscoelastic solid model of the shell, in the form of the Kelvin–Voigt constitutive equation, has been studied in Refs. 9 and 11. According to this model, the shell elasticity plays a positive role for scattering. An increase in the shell elasticity leads to increasing the magnitude of the resonance peaks in the scattering cross-section curves.[11] In this section, we analyze the effect of viscosity on the resonance frequency of and resonant backscatter from microbubbles encapsulated by a viscoelastic solid shell.

Let us consider a spherical gas bubble covered by a viscoelastic layer of finite thickness (Fig. 1). The bubble resides in an unbounded Newtonian liquid (blood plasma) and oscillates under the action of ultrasound. The radial oscillation of an encapsulated bubble is described by the equation[11]

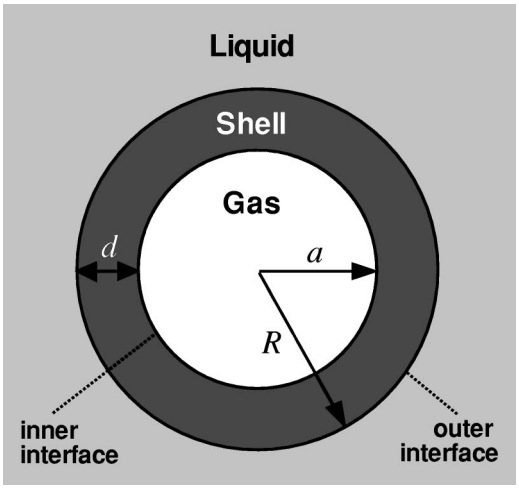


FIG. 1. Schematic sketch of an encapsulated microbubble of outer radius R. The microbubble resides in an unbounded Newtonian liquid. It is covered with a shell (viscoelastic layer of finite thickness d), i.e., it is characterized by two interfaces: inner (gas-shell) and outer (shell-liquid) interfaces.

$$\rho_{s0}\left\{\left(1+\frac{\Delta_\rho a}{R}-\frac{F}{C_l}\frac{da}{dt}\right)a\frac{d^2a}{dt^2}+\left[\frac{3}{2}+\frac{\Delta_\rho a}{R}\left(2-\frac{a^3}{2R^3}\right)-\frac{1}{2C_l}\frac{d}{dt}(aF)\right]\left(\frac{da}{dt}\right)^2\right\}=p_a-p_I+\frac{\rho_{l0}}{\rho_{s0}C_l}\frac{d}{dt}\left[\frac{a(p_a-p_I)}{1+\Delta_\rho a/R}\right], \tag{1a}$$

$$p_a=p_{g0}\left(\frac{a}{a_0}\right)^{-3\kappa}-2\left(\frac{\sigma_1}{a}+\frac{\sigma_2}{R}\right)+3\left[\int_a^R\frac{\tau_{rr}^{(s)}}{r}dr+\int_R^\infty\frac{\tau_{rr}^{(l)}}{r}dr\right]. \tag{1b}$$

Here t is time, a the inner radius of the bubble, R the outer radius of the bubble ($d=R-a$ is the shell thickness), ρ_{s0} the shell density, ρ_{l0} the initial density of the liquid (blood plasma), $p_{g0}=p_0$ the initial gas pressure, p_I the incident pressure in the liquid, and C_l the speed of sound in the liquid. Surface tensions at the inner (gas-shell) and outer (shell-liquid) interfaces are denoted by σ_1 and σ_2. The parameters Δ_ρ and F are defined as

$$\Delta_\rho=\frac{\rho_{l0}-\rho_{s0}}{\rho_{s0}},\quad F=\frac{\rho_{l0}}{\rho_{s0}}\frac{1+\Delta_\rho(a/R)^4}{1+\Delta_\rho a/R}. \tag{2}$$

A polytropic exponent κ is different from γ_g (ratio of constant-pressure to constant-volume specific heats for the gas) because of heat transfer through the bubble walls. It takes the value between 1 (isothermal behavior) and γ_g (adiabatic behavior). However, we disregard the energy dissipation due to thermal effects because the thermal damping coefficient for an encapsulated microbubble is three orders of magnitude less than the viscous damping coefficient if the ultrasound scanner operates at frequencies $f=1-10$ MHz.[11] In the case of free microbubbles, thermal damping is two orders of magnitude less than viscous damping.[11]

The integrals in (1b) represent the contributions of the shell and liquid viscoelasticities to the bubble oscillation. The radial component of the shear stress tensor (radial stress) is denoted by τ_{rr}. In the shell domain, $\tau_{rr}=\tau_{rr}^{(s)}$. In the liquid domain, $\tau_{rr}=\tau_{rr}^{(l)}$. The liquid elasticity weakly affects the bubble oscillation as compared to the shell elasticity.[11] Therefore, we assume that the liquid is Newtonian, i.e., $\tau_{rr}^{(l)}$ depends linearly on the radial component of the rate-of-strain tensor:

$$\tau_{rr}^{(l)}=2\mu_l\dot{\gamma}_{rr}=2\mu_l\frac{\partial v_r}{\partial r} \tag{3}$$

(μ_l is the liquid viscosity). Near the outer interface of the oscillating bubble, the radial velocity of the liquid is

$$v_r=\frac{a^2}{r^2}\frac{da}{dt}=\frac{R^2}{r^2}\frac{dR}{dt}. \tag{4}$$

Provided $\tau_{rr}^{(l)}|_{r\to\infty}=0$, substitution of Eqs. (3) and (4) into the second integral of Eq. (1b) gives

$$3\int_R^\infty\frac{\tau_{rr}^{(l)}}{r}dr=-\frac{4\mu_l a^2}{R^3}\frac{da}{dt}=-\frac{4\mu_l}{R}\frac{dR}{dt}. \tag{5}$$

We restrict our attention to the case of small-amplitude oscillation of the microbubble in the sinusoidal pressure field

$$p_I=p_0-P_A\sin(\Omega t)=p_0+\frac{P_A}{2}[i\exp(i\Omega t)+\text{c.c.}]. \tag{6}$$

Here p_0 is the undisturbed liquid pressure in the region far from the microbubble (blood pressure), P_A the acoustic pressure amplitude, Ω the angular transmitted frequency [the transmitted frequency $f=\Omega/(2\pi)$]. The symbol c.c. denotes complex conjugate. The acoustic pressure amplitude is far less than p_0, i.e., there exists a small parameter ϵ_p such that $P_A=\epsilon_p p_0 P$ and P is a real number of the order of 1. Then, the solution to Eq. (1) is expanded in powers of ϵ_p as[11]

$$a=a_0(1+x), \tag{7a}$$

$$R=R_0\left[1+\frac{a_0^3}{R_0^3}x+\frac{a_0^3}{R_0^3}\left(1-\frac{a_0^3}{R_0^3}\right)x^2+O(x^3)\right], \tag{7b}$$

$$x=x(t;\epsilon_p)=\epsilon_p x_1(t)+\epsilon_p^2 x_2(t)+O(\epsilon_p^3) \tag{7c}$$

(a_0 and R_0 are the equilibrium values of the inner and outer radii of the microbubble). The linear and second-order nonlinear acoustic responses of the microbubble (the first and second harmonics of microbubble oscillation) are specified by the functions $X_1(t)=\epsilon_p a_0 x_1(t)$ and $X_2(t)=\epsilon_p^2 a_0 x_2(t)$. They may be expressed as

$$X_1(t)=\frac{A_{1r}(\Omega)}{2}P_A\exp(i\Omega t), \tag{8a}$$

$$X_2(t)=A_{0r}(\Omega)P_A^2+\frac{A_{2r}(\Omega)}{2}P_A^2\exp(2i\Omega t), \tag{8b}$$

where $A_{0r}(\Omega)$, $A_{1r}(\Omega)$, and $A_{2r}(\Omega)$ are the complex-valued functions of the transmitted frequency. The magnitudes of these functions represent the zeroth-, first-, and second-harmonic responses of the microbubble to the ultrasound field.

To calculate the contribution of the microbubble to the backscattered signal, the scattering cross-sections of the microbubble at the transmitted frequency and at twice the transmitted frequency are considered. The scattering cross-section at the transmitted frequency, or the first-harmonic scattering cross-section, is denoted by σ_{s1}. It is related to the ratio of the total acoustic power W_1 scattered by the microbubble at the first harmonic to the intensity I_0 of the incident acoustic field,[9]

$$\sigma_{s1}=\frac{W_1}{I_0},\quad W_1=\frac{4\pi r^2|P_{s1}|^2}{2\rho_l C_l},\quad I_0=\frac{P_A^2}{2\rho_l C_l}. \tag{9}$$

The symbol σ_{s2} denotes the scattering cross section at twice the transmitted frequency, or the second-harmonic scattering cross section:

$$\sigma_{s2}=\frac{W_2}{I_0},\quad W_2=\frac{4\pi r^2|P_{s2}|^2}{2\rho_l C_l} \tag{10}$$

(W_2 is the total acoustic power scattered by the microbubble at the second harmonic). Here r is the radial coordinate, P_{s1} and P_{s2} are the first- and second-harmonic amplitudes of the pressure wave scattered by the microbubble. Based on the relationship between the scattered pressure field and the bubble radius[11,21–23]

$$p_s(t,r)=\frac{\rho_{l0}}{3r}\frac{d^2}{d\xi^2}a^3(\xi),\quad \xi=t-r/C_l, \tag{11}$$

the expressions for P_{s1} and P_{s2} can be found [see Eqs. (86) and (87) in Ref. 11]. Then, the first- and second-harmonic scattering cross sections depend on $A_{1r}(\Omega)$ and $A_{2r}(\Omega)$ as follows:

$$\sigma_{s1}=4\pi\rho_{l0}^2 a_0^4\Omega^4|A_{1r}(\Omega)|^2, \tag{12a}$$

$$\sigma_{s2}=64\pi\rho_{l0}^2 a_0^4 P_A^2\Omega^4\left|A_{2r}(\Omega)+\frac{A_{1r}^2(\Omega)}{2a_0}\right|^2. \tag{12b}$$

It is important to say that the expression for the second-harmonic scattering cross section given by Church [Eq. (26b) in Ref. 9] is in error because it has been derived without considering (a) the nonlinear relationship between the scattered pressure field and bubble oscillation and (b) the dependence of the scattered pressure on the second time derivative of the bubble radius. Due to the same reasons, it is also inappropriate to use the well-known formula for the scattered pressure[24]

$$p_s=i\Omega\rho_{l0}\left(a_0^2\frac{da}{dt}\right)\frac{\exp[-i(\Omega/C_l)r]}{r} \tag{13}$$

for evaluating nonlinear scattering by microbubbles.

If the shell behaves as a viscoelastic solid, its behavior can be described by the Kelvin–Voigt model. Because we ignore the shape deformation of the microbubble, this model is reduced to the following relation:[11]

$$\tau_{rr}^{(s)}=-\frac{4a^2}{r^3}\left[G_s(a-a_e)+\mu_s\frac{da}{dt}\right], \tag{14}$$

where G_s and μ_s are the elasticity modulus and viscosity of the shell, a_e is the unstrained inner radius of the bubble, which is different from the equilibrium radius a_0 due to radial prestress in the shell.[11] Under the assumption that the prestress is counterbalanced by the over pressure, the unstrained radius depends on the shell elasticity and surface tensions,[9]

$$a_e = a_0(1+Z), \quad Z = \frac{1}{2G_s}\left(\frac{\sigma_1}{a_0}+\frac{\sigma_2}{R_0}\right)\frac{R_0^3}{R_0^3-a_0^3}. \tag{15}$$

In view of Eq. (15), substitution of Eq. (14) into the first integral of Eq. (1b) gives

$$3\int_a^R \frac{\tau_{rr}^{(s)}}{r}dr = -\frac{4(R_0^3-a_0^3)}{R^3}\left[G_s\left(1-\frac{a_0}{a}\right)+\frac{\mu_s}{a}\frac{da}{dt}\right] + \frac{a_0R_0^3}{aR^3}\left(\frac{2\sigma_1}{a_0}+\frac{2\sigma_2}{R_0}\right). \tag{16}$$

Because $R^3 = a^3 + R_0^3 - a_0^3$ due to the shell incompressibility, Eqs. (1), (5), and (16) can be combined into a single ODE in the inner radius a (or in the outer radius R). The resulting differential equation represents the mathematical model for radial oscillation of an encapsulated microbubble in a compressible Newtonian liquid provided the encapsulating shell behaves as a viscoelastic solid. By applying the perturbation technique [see Eqs. (6)–(8)] to this model and using Eqs. (12), we obtain the following expressions for the first- and second-harmonic scattering cross-sections of the microbubble:

$$\sigma_{s1} = \frac{4\pi(1+\Delta_\rho a_0/R_0)^{-2}\rho_{l0}^2 a_0^2\Omega^4}{\rho_{s0}^2\{[\Omega_0^2-\Omega^2+S_{ac}(\Omega)]^2+4\Omega^2\beta^2(\Omega)\}}, \tag{17a}$$

$$\sigma_{s2} = \frac{16\pi(1+\Delta_\rho a_0/R_0)^{-4}\rho_{l0}^2 P_A^2\Omega^4\Gamma(\Omega)}{\rho_{s0}^4 a_0^2\{[\Omega_0^2-\Omega^2+S_{ac}(\Omega)]^2+4\Omega^2\beta^2(\Omega)\}^2}. \tag{17b}$$

Here $\Omega_0 = 2\pi f_0$ and f_0 is the undamped natural frequency, i.e., the natural frequency of microbubble oscillation upon ignoring viscous damping and liquid compressibility,

$$f_0 = \frac{1}{2\pi a_0[\rho_{s0}(1+\Delta_\rho a_0/R_0)]^{1/2}}\left\{3\kappa p_0 - \frac{2\sigma_1}{a_0} - \frac{2\sigma_2 a_0^3}{R_0^4} + 4G_s\left(1-\frac{a_0^3}{R_0^3}\right)\left[1+\left(1+\frac{3a_0^3}{R_0^3}\right)Z\right]\right\}^{1/2}. \tag{18}$$

The function $S_{\rm ac}(\Omega)$ is the contribution of acoustic radiation to the bubble stiffness,

$$S_{\rm ac}(\Omega) = \frac{T_{\rm ac}^2\Omega^4}{1+T_{\rm ac}^2\Omega^2}, \quad T_{\rm ac} = \frac{\rho_{l0}a_0}{\rho_{s0}C_l}\left(1+\frac{\Delta_\rho a_0}{R_0}\right)^{-1}. \tag{19}$$

If thermal damping is ignored, the total linear damping coefficient $\beta(\Omega) = \beta_{\rm vis} + \beta_{\rm ac}(\Omega)$, where

$$\beta_{\rm vis} = \frac{2}{\rho_{s0}a_0^2}\left(1+\frac{\Delta_\rho a_0}{R_0}\right)^{-1}\left[\mu_s\left(1-\frac{a_0^3}{R_0^3}\right)+\mu_l\frac{a_0^3}{R_0^3}\right], \tag{20}$$

$$\beta_{\rm ac}(\Omega) = \frac{T_{\rm ac}\Omega^2}{2(1+T_{\rm ac}^2\Omega^2)} \tag{21}$$

are the viscous damping and acoustic radiation damping coefficients. The function $\Gamma(\Omega)$ can be written as

$$\Gamma(\Omega) = \frac{[\Omega_n^2-\alpha\Omega^2+S_n(\Omega)]^2+16\Omega^2\beta_n^2(\Omega)}{[\Omega_0^2-4\Omega^2+S_{\rm ac}(2\Omega)]^2+16\Omega^2\beta^2(2\Omega)}. \tag{22}$$

Here

$$\alpha = \frac{3[1+\Delta_\rho(a_0/R_0)^4]}{2(1+\Delta_\rho a_0/R_0)}, \tag{23}$$

$$S_n(\Omega) = S_{\rm ac}(2\Omega) - \frac{4\alpha}{3}\frac{T_{\rm ac}^2\Omega^4(4+T_{\rm ac}^2\Omega^2)}{(1+4T_{\rm ac}^2\Omega^2)(1+T_{\rm ac}^2\Omega^2)}, \tag{24}$$

$$\Omega_n = \frac{1}{a_0[\rho_{s0}(1+\Delta_\rho a_0/R_0)]^{1/2}}\left\{\frac{9\kappa(\kappa+1)}{2}p_0 - \frac{4\sigma_1}{a_0} - \frac{4\sigma_2 a_0^6}{R_0^7} + 4G_s\left(1-\frac{a_0^3}{R_0^3}\right)\times\left[2+\frac{3a_0^3}{R_0^3}+\left(2+\frac{3a_0^3}{R_0^3}+\frac{9a_0^6}{R_0^6}\right)Z\right]\right\}^{1/2}, \tag{25}$$

$$\beta_n(\Omega) = \frac{3}{\rho_{s0}a_0^2}\left(1+\frac{\Delta_\rho a_0}{R_0}\right)^{-1}\left[\mu_s\left(1-\frac{a_0^6}{R_0^6}\right)+\mu_l\frac{a_0^6}{R_0^6}\right] + \beta_{\rm ac}(2\Omega) - \frac{\alpha}{2}\frac{T_{\rm ac}\Omega^2(1-T_{\rm ac}^2\Omega^2)}{(1+4T_{\rm ac}^2\Omega^2)(1+T_{\rm ac}^2\Omega^2)}, \tag{26}$$

F and $T_{\rm ac}$ are given in Eqs. (2) and (19), respectively.

By definition,[14] the linear resonance frequency of bubble oscillation $f_{\rm res}$ is the frequency at which the bubble first-harmonic response (linear amplitude-frequency response) has a local maximum. If the bubble is covered with a shell, its first-harmonic response is of the form

$$|A_{1r}(\Omega)| = \frac{\rho_{s0}^{-1}a_0^{-1}(1+\Delta_\rho a_0/R_0)^{-1}}{\{[\Omega_0^2-\Omega^2+S_{\rm ac}(\Omega)]^2+4\Omega^2\beta^2(\Omega)\}^{1/2}}. \tag{27}$$

Even if the surrounding liquid is assumed to be incompressible, the total damping coefficient is sufficiently large to influence the resonance frequency of microbubbles encapsulated by the high-viscosity shell. Damping has a negligible effect on $f_{\rm res}$ if

$$\beta(\Omega) \ll \Omega_0/2. \tag{28}$$

This condition is satisfied if the Reynolds numbers for the shell $\mathrm{Re}_s = a_0\rho_{s0}U/\mu_s$ and for the liquid $\mathrm{Re}_l = a_0\rho_{l0}U/\mu_l$ are much more than unity ($U = \sqrt{p_0/\rho_{l0}}$ is the characteristic velocity). It is clear that the difference between the resonance and undamped natural frequencies increases with decreasing the bubble size and increasing viscosity. For microbubbles of radius 1 μm suspended in water and encapsulated by the shell, which density is equal to the water density, $\mathrm{Re}_s < 1$ if $\mu_s > 0.01$ Pa s $= 10$ cP. Meanwhile, the shell viscosity of con-

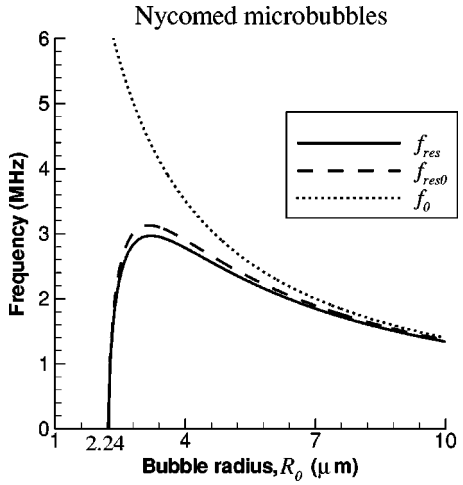


FIG. 2. Resonance frequency (solid) of an air microbubble encapsulated by a polymeric shell (Refs. 25 and 27) as a function of the outer radius of the microbubble, compared with the undamped natural frequency (dotted). The microbubble resides in human blood plasma. The dashed line is the resonance frequency under the assumption that blood plasma is incompressible [Eq. (29)]. The solid and dotted lines are calculated from Eqs. (30) and (18), respectively. The encapsulating shell is characterized by viscosity of 0.45 Pa s and elasticity of 12 MPa. Its thickness is 5% of the bubble outer radius. Other parameters, $\mu_l = 0.0012$ Pa s, $C_l = 1543$ m s^{-1}, $\rho_{l0} = 1027$ kg m^{-3}, $\rho_{s0} = 1100$ kg m^{-3}, $p_0 = 0.1$ MPa, $\sigma_1 = 0.04$ N m^{-1}, $\sigma_2 = 0.005$ N m^{-1}, and $\kappa = 1.1$.

trast microbubbles is much higher than 10 cP.[8,25] Therefore, using the term ''resonance frequency'' as a synonym of ''natural frequency'' is validated only for large bubbles.

Upon neglecting the liquid compressibility, the simple expression for the resonance frequency is obtained from the maximization of $|A_{1r}(\Omega)|$,

$$f_{\rm res} = f_{\rm res\,0} = \sqrt{f_0^2 - \frac{\beta_{\rm vis}^2}{2\pi^2}}. \tag{29}$$

As follows from Eq. (27), microbubbles resonate in a compressible viscous liquid at the frequency

$$f_{\rm res} = \frac{1}{2\pi T_{\rm ac}} \left\{ -1 + \left[1 + \frac{4T_{\rm ac}^2 (2\pi^2 f_0^2 - \beta_{\rm vis}^2)}{1 + 4T_{\rm ac}\beta_{\rm vis}(1 + T_{\rm ac}\beta_{\rm vis})} \right]^{1/2} \right\}^{1/2}. \tag{30}$$

Note that the acoustic contribution to the resonance frequency of free bubbles was discussed by Prosperetti.[26]

Figure 2 compares the resonance frequency with the undamped natural frequency for air microbubbles encapsulated by the shell of viscosity $\mu_s = 0.45$ Pa s and elasticity $G_s = 12$ MPa. The shell thickness d is 5% of the initial outer radius R_0. As estimated by Hoff *et al.*,[25] these values of parameters are typical of ''polymeric microbubbles'' prepared by Nycomed.[27] We assume that the microbubbles reside in blood plasma (viscosity $\mu_l = 0.0012$ Pa s, density $\rho_{l0} = 1027$ kg m^{-3}, and speed of sound $C_l = 1543$ m s^{-1}).[28] The shell density is supposed to be slightly higher than the plasma density: $\rho_{s0} = 1100$ kg m^{-3}. The initial pressure in the liquid $p_0 = 0.1$ MPa. The surface tensions $\sigma_1 = 0.04$ N m^{-1} and $\sigma_2 = 0.005$ N m^{-1} (see Ref. 9). In the case of encapsulated microbubbles, the best matching between the numerical solution, which accounts for heat conduction through the bubble wall, and the analytical formula is observed at $\kappa = 1.1$.[11] This value of the polytropic exponent is used in all calculations here.

As illustrated in Fig. 2, there is a notable difference between the resonance and undamped natural frequencies over the whole range of microbubble sizes. In the case of Nycomed suspension, this difference varies from -0.39 MHz (15.9%) at $R_0 = 5$ μm to -3.30 MHz (141.8%) at $R_0 = 2.5$ μm. In addition, encapsulated microbubbles are characterized by a critical size below which they never resonate. The critical size is in the size range of contrast agents. The critical radius of Nycomed microbubbles is 2.24 μm. From both Eqs. (29) and (30) it follows that microbubbles cease to resonate if

$$\beta_{\rm vis} \geqslant \pi\sqrt{2} f_0. \tag{31}$$

The greater the shell and liquid viscosities, the greater the critical radius (and hence the difference between the resonance and undamped natural frequencies). Although the critical radius decreases with increasing the shell elasticity, the shell viscosity is a more important parameter: a tenfold decrease in shell viscosity has approximately the same effect on the critical radius of Nycomed microbubbles as a 100-fold increase in shell elasticity.

Figure 2 also shows that ignoring the liquid compressibility leads to overestimation of the resonance frequency: the microbubble situated in a less compressible liquid resonates at a higher frequency. This happens due to acoustic radiation damping. The difference between $f_{\rm res\,0}$ and $f_{\rm res}$ reaches 0.165 MHz (≈5.6%) for Nycomed bubbles of radius 3.1 μm. Another result of Fig. 2 is that there exists an upper limit to the resonance frequency of microbubbles: the Nycomed suspension does not resonate if the transmitted frequency exceeds 3 MHz. This result is, however, not applicable to the backscattered signal produced by the suspension, as shown later.

Even in the case of Albunex®, which is a suspension of protein-encapsulated microbubbles, there is a significant deviation of the resonance frequency from the undamped natural frequency for all possible sizes of microbubbles.[11] As estimated by Church,[9] the protein (albumin) shell of these microbubbles is thin ($d = 15$ nm), very viscous ($\mu_s = 1.77$ Pa s) and elastic ($G_s = 88.8$ MPa). Figure 3 shows that the critical radius of Albunex® bubbles $R_c = 1.46$ μm. This value is less than the critical radius of Nycomed bubbles mainly due to a difference in shell thickness between these two contrast agents. The upper limit to the resonance frequency is higher for Albunex® ($f_{\rm res}^{\rm max} = 5.47$ MHz) than for Nycomed ($f_{\rm res}^{\rm max} = 2.97$ MHz). This can be explained only by a higher elasticity of the albumin shell. (A decrease in shell thickness leads to a decrease in the critical radius but has no effect on the upper limit to the resonance frequency.)[11]

The widely used assumption that the resonances of the backscattered signal (resonance peaks in the output-level versus frequency curve) are produced by microbubbles resonating at the transmitted frequency may lead to misinterpretation of experimental data. The first-harmonic resonance of the signal scattered by an encapsulated microbubble occurs at the frequency f_1, which is different from the resonance

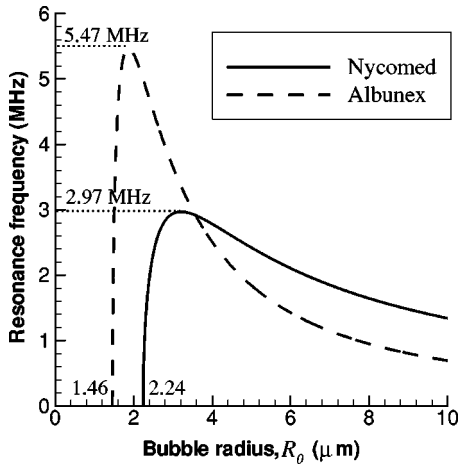


FIG. 3. Resonance frequencies of Nycomed and Albunex® microbubbles (solid and dashed) as functions of the outer radius [based on Eq. (30)]. The Albunex® microbubble is encapsulated by a protein shell of viscosity $\mu_s = 1.77$ Pa s, elasticity $G_s = 88.8$ MPa, and constant thickness $d = 15$ nm (see Ref. 9). Both microbubbles reside in human blood plasma. The parameters of the Nycomed microbubble and blood plasma are as in Fig. 2.

frequency of the microbubble f_{res}. If the surrounding liquid is considered to be incompressible, the maximization of the first-harmonic scattering cross-section σ_{s1} [see Eq. (17a)] gives the following formula for f_1:

$$f_1 = \frac{f_0^2}{f_{\text{res}\,0}} = \frac{f_0^2}{\sqrt{f_0^2 - \beta_{\text{vis}}^2/(2\pi^2)}}. \tag{32}$$

In the case of large bubbles, i.e., when the difference between the resonance and undamped natural frequencies of bubble oscillation is negligible small, it is correctly reasoned that $f_1 = f_{\text{res}} = f_0$. In such a situation, resonating bubbles are responsible for the resonance peaks in the output-level versus frequency curve. In the case of small bubbles, the viscous damping coefficient is large and, therefore, nonresonating bubbles make the major contribution to the resonant backscatter.

Considering the liquid compressibility, the first-harmonic resonance frequency of the backscattered signal $f_1 = \sqrt{X}$. The quantity X, as is evident from the maximization of σ_{s1}, is one of the positive real roots of the cubic equation

$$X^3 + c_2 X^2 + c_1 X + c_0 = 0 \tag{33}$$

with

$$c_0 = \frac{f_0^4}{2\pi^2 T_{\text{ac}}^2 (1 + 2T_{\text{ac}}\beta_{\text{vis}})^2}, \tag{34a}$$

$$c_1 = \frac{8\pi^2 T_{\text{ac}}^2 f_0^4 + \beta_{\text{vis}}^2/(2\pi^2) - f_0^2}{2\pi^2 T_{\text{ac}}^2 (1 + 2T_{\text{ac}}\beta_{\text{vis}})^2}, \tag{34b}$$

$$c_2 = \frac{2(4\pi^2 T_{\text{ac}}^2 f_0^4 + \beta_{\text{vis}}^2/\pi^2 - 2f_0^2)}{(1 + 2T_{\text{ac}}\beta_{\text{vis}})^2}. \tag{34c}$$

Figure 4(a) plots the resonance and undamped natural frequencies of the Nycomed microbubble (dashed–dotted and dashed lines) situated in blood plasma as well as the first-harmonic resonance frequency of the signal scattered by

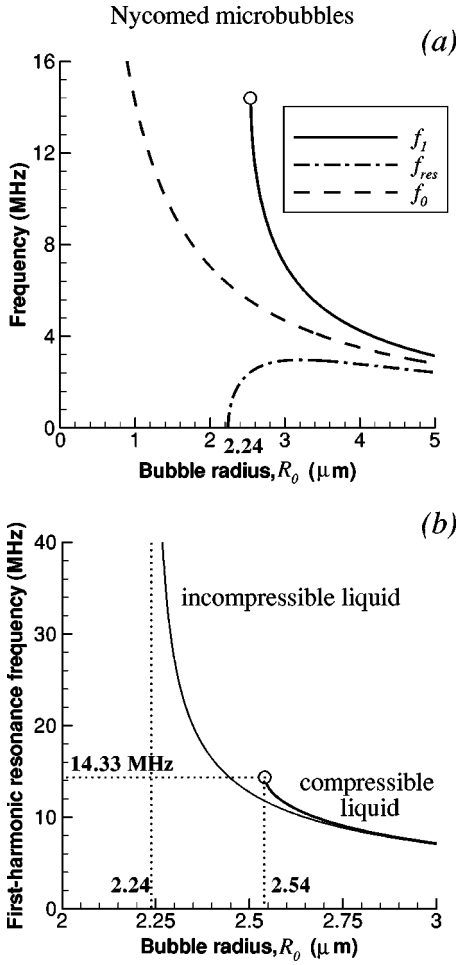


FIG. 4. (a) Resonance and undamped natural frequencies of Nycomed microbubbles (dashed–dotted and dashed) and first-harmonic resonance frequency of the backscattered signal from the microbubbles (solid line) as functions of the bubble outer radius. The liquid (blood plasma) compressibility is taken into account. Open circle denotes the highest frequency at which the resonant peak in the first-harmonic scattering cross section can be observed. (b) The first-harmonic resonance frequency of the backscattered signal produced by Nycomed microbubbles surrounded by compressible and incompressible liquids (thick and thin solid lines) as a function of the bubble outer radius.

the microbubble (solid line) as functions of the outer radius R_0. The first-harmonic resonance frequency is calculated from the solution of Eq. (33) (see p. 17 in Ref. 29). Like the undamped natural frequency f_0, the first-harmonic resonance frequency of the backscattered signal f_1 decreases monotonically with increasing the bubble radius. However, f_0 tends to infinity when the inner radius a_0 approaches zero, while f_1 reaches the highest value (several tens of MHz) at $a_0 > 0$. If we continue to use the assumption $f_1 = f_{\text{res}} = f_0$ for bubbles of arbitrary small size, we will be led to the conclusion that the bubble resonance and the resonance peak in the first-harmonic scattering cross section always exist, no matter how bubbles are small. We will, then, argue that very small bubbles resonate at very high frequencies. Figure 4 demonstrates that this conclusion is wrong. Bubbles are characterized by the critical radius. Not only bubbles of subcritical size cease to resonate, but they also never produce the resonant backscatter at the transmitted frequency. (As will be shown later, bubbles of subcritical size can produce the reso-

nant backscatter at twice the transmitted frequency.) The difference between the first-harmonic resonance frequency of the backscattered signal and the microbubble resonance frequency always exceeds the difference between the undamped natural and resonance frequencies of the microbubble. For Nycomed bubbles of outer radius $R_0=5\ \mu$m, $f_1-f_{\rm res}\approx 0.73$ MHz (30.0%), while $f_0-f_{\rm res}\approx 0.38$ MHz (15.9%). If $R_0=2.54\ \mu$m, $f_1-f_{\rm res}\approx 11.89$ MHz (487.2%) and $f_0-f_{\rm res}\approx 3.09$ MHz (126.8%).

As indicated in Fig. 4(b), a restriction on the first-harmonic resonance frequency is solely the effect of the liquid compressibility. If blood plasma is treated as an incompressible liquid, f_1 is calculated from Eq. (32). According to this equation, f_1 tends to infinity when the bubble radius approaches the critical value ($R_{\rm crit}=2.24\ \mu$m for Nycomed microbubbles). However, Nycomed microbubbles oscillating in blood plasma are not able to produce the resonance peak in the first-harmonic scattering cross section if the transmitted frequency exceeds 14.33 MHz. The existence of the upper limit to f_1 results in another value of the critical bubble radius, specific for f_1. In particular, the resonant backscatter from Nycomed bubbles of radius less than $R_{\rm crit\,1}=2.54\ \mu$m seems not to exist.

In the most experimental works on contrast agents, the properties of microbubbles, such as the shell viscosity and elasticity,[25,30] are evaluated through frequency analysis of the backscattered echo, i.e., from the output-level versus frequency curves. It is generally believed[1,31] that the highest peaks in these curves are associated with the resonance frequency of microbubbles. In reality, the highest peaks are associated with the first-harmonic resonance frequency of the backscattered signal f_1, which is not coincident with the microbubble resonance frequency $f_{\rm res}$. Moreover, it is believed that the frequency at which the highest peak is observed depends on the microbubble parameters according to Eq. (18). In reality, Eq. (18) describes the undamped natural frequency f_0, which is different from the first-harmonic resonance frequency of the backscattered signal. The undamped natural frequency can be used instead of f_1 and $f_{\rm res}$ if the microbubble radius is more than 5 μm, i.e., beyond the size range of contrast agents. Figure 5 shows how the difference between f_1 (measured in experiments) and f_0 (used instead of f_1) depends on the microbubble radius for two types of contrast agents: Nycomed[25,27] and Albunex®.[9] Although this difference is relatively small for bubbles of radius $R_0=5\ \mu$m, in particular, $f_1-f_0=0.34$ MHz for Nycomed bubbles and 0.025 MHz for Albunex® bubbles, it becomes very large when the bubble radius approaches the critical value $R_{\rm crit\,1}$. Note that in the case of Albunex® bubbles the critical radius for f_1 is again more than the critical radius for $f_{\rm res}$: $R_{\rm crit\,1}=1.61\ \mu{\rm m}>R_{\rm crit}=1.46\ \mu$m. The maximal difference between f_1 and f_0 is 8.8 MHz for Nycomed bubbles and 14.1 MHz for Albunex® bubbles. Therefore, the use of Eq. (18) for the interpretation of experimental data on the resonant backscatter from microbubbles cannot be tolerated.

A monodisperse suspension of encapsulated microbubbles cannot produce the resonant backscatter at both the transmitted frequency and twice the transmitted frequency. It is not quite correct to argue that if bubbles of one radius resonate, all the harmonics, subharmonics, and ultraharmonics of the backscattered signal they produce will be resonant. This statement is validated only for large bubbles. In particular, the second harmonic of the backscattered signal from encapsulated microbubbles of radius R_0 has a resonance peak at the frequency f_2 (second-harmonic resonance frequency), which is equal to neither the first-harmonic resonance frequency f_1 nor the undamped natural frequency f_0. This is apparent from the maximization of the second-harmonic scattering cross section σ_{s2}. It should be noted that f_2 is defined as the *transmitted frequency* at which the second harmonic of the backscattered signal is resonant, i.e., in the output-level versus frequency curve, the second-harmonic resonance peak will be at $2f_2$. The σ_{s2} is a very complicated function of the angular frequency Ω [see Eq. (17b)]. It is impossible to obtain the analytical formula for f_2 even in the incompressible case. (If $T_{\rm ac}=0$, the maximization of σ_{s2} results in a fifth-order algebraic equation in Ω^2.) To calculate numerically the second-harmonic resonance frequency as a function of bubble radius, we use Maple™ (Maplesoft, a division of Waterloo-Maple Inc.). The results of the numerical calculation are presented in Figs. 6 and 7.

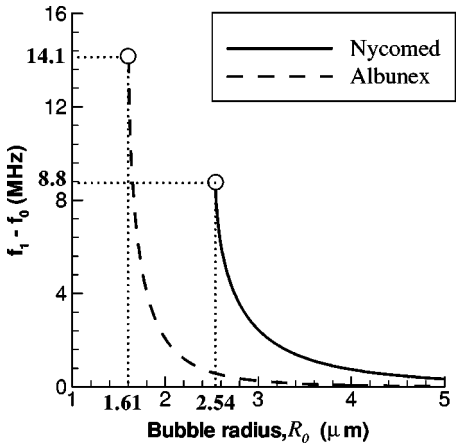


FIG. 5. The difference between the first-harmonic resonance frequency f_1 and the undamped natural frequency f_0 for Nycomed and Albunex® microbubbles (solid and dashed) as functions of the bubble outer radius. Nycomed microbubbles of radius $R_0<2.54\ \mu$m and Albunex® microbubbles of radius $R_0<1.61\ \mu$m do not produce resonant scattering at the transmitted frequency. The parameters are listed in Figs. 2 and 3.

Figure 6 compares the second-harmonic resonance frequency f_2 (solid line), the first-harmonic resonance frequency f_1 (dashed–dotted line), and the undamped natural frequency f_0 (dashed line) for Nycomed microbubbles of radius $R_0\leqslant 5\ \mu$m. While f_1 is always higher than f_0, f_2 is always less than f_0. Hence, the difference between f_1 and f_2 is more than between f_1 and f_0. It increases with decreasing the bubble radius. For instance, when the outer radius of Nycomed microbubbles is changed from 5 μm to 2.54 μm, f_2-f_1 increases from 0.9 MHz to 10.6 MHz. Also, there is no critical radius for the second-harmonic resonance. Even submicron bubbles may have a resonance peak at twice the transmitted frequency. Hence, ultrasound-based methods for detection of very small bubbles inside materials should exploit the second harmonic, not the first harmonic.

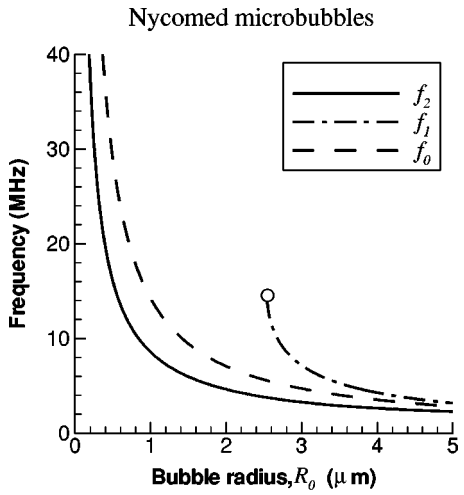


FIG. 6. First- and second-harmonic resonance frequencies of the backscattered signal (dashed–dotted and solid) and undamped natural frequency of Nycomed microbubbles (dashed) as functions of the outer radius. The liquid compressibility is taken into account. The second-harmonic resonance frequency is defined as the transmitted frequency at which the second harmonic of the backscattered signal is resonant. Open circle denotes the highest frequency at which the resonant peak in the first-harmonic scattering cross section can be observed. $P_A = 30$ kPa.

As illustrated in Fig. 7, overestimation of the optimal transmitted frequency for second-harmonic imaging by using f_0 reaches several tens of MHz for microbubbles of radius less than 1 μm. In this figure, the difference between f_2 and f_0 is plotted as a function of the outer radius for Nycomed and Albunex® microbubbles (solid and dashed lines). This difference remains relatively small (but cannot be ignored) for microbubbles of radius between 3 μm and 5 μm. In this range, $f_2 - f_0$ varies from -0.55 to -1.43 MHz for Nycomed bubbles and from -0.063 to -0.485 for Albunex® bubbles. The second-harmonic resonance frequency deviates significantly from the undamped natural frequency if the bubble radius is less than 3 μm (or its size is less than 6 μm).

The microbubble population is highly polydisperse. The polydisperse suspension is less sensitive to the transmitted frequency than the monodisperse suspension. If the ultrasound transducer operates at the frequency of several MHz (from 2 to 14 MHz), there always exist bubbles of one size, say R_1, that produce the resonant backscatter at the transmitted frequency and bubbles of another size, say R_2, that produce the resonant backscatter at twice the transmitted frequency. The backscattered signal from this suspension, therefore, has peaks at the transmitted frequency and multiple of this frequency. It is important to say that R_1 is not equal to R_2 in the case of encapsulated microbubbles.

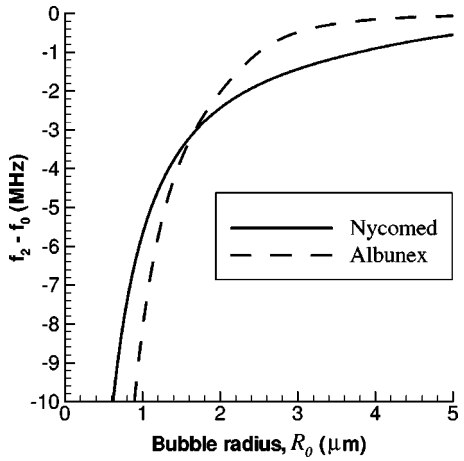


FIG. 7. The difference between the second-harmonic resonance frequency f_2 and the undamped natural frequency f_0 for Nycomed and Albunex® microbubbles (solid and dashed) as functions of the outer radius. $P_A = 30$ kPa.

To increase the quality of contrast-enhanced ultrasound imaging, one needs (a) to create a suspension of microbubbles with a mean radius equal to either R_1 (for fundamental mode imaging) or R_2 (for second-harmonic mode imaging) or (b) to operate the transducer at the frequency f_1 in the fundamental (first-harmonic) mode or at f_2 in the second-harmonic mode, where f_1 and f_2 are taken at the mean radius of microbubble population. The former way should be used if the transducer is characterized by a narrow range of operating frequencies. Unfortunately, neither Albunex® nor Nycomed (and other polymeric contrast agents) are good candidates for the fundamental mode imaging. The mean radius of Albunex® bubbles is about 1.7 μm,[9] which is very close to the critical radius for f_1. As shown in Fig. 5, $R_{\text{crit}\,1} = 1.61\,\mu$m for Albunex®. (The critical radius will be even higher if thermal damping is taken into account.) Polymeric microbubbles are worse than Albunex® because of the thick shell. The mean radius of Nycomed microbubbles is between 2.0 and 2.5 μm,[27] less than the critical radius ($R_{\text{crit}\,1} = 2.54\,\mu$m for Nycomed). These contrast agents can, however, be used for second-harmonic imaging. Figure 8(a) shows that if the transducer operates at frequency $f = 5.0$ MHz (this frequency was used in experiments with Nycomed bubbles by Hoff *et al.*),[25] a Nycomed suspension should be of mean radius $R_1 = 4.36\,\mu$m in order to provide optimal imaging in the fundamental (first-harmonic) mode. Such a high value of R_1 signifies that the major contribution to the first harmonic of the backscattered signal comes from 4–5 μm bubbles, i.e., from the largest microbubbles of the suspension. Smaller microbubbles plays an insignificant role for first-harmonic imaging. Optimal second-harmonic imaging is possible if the mean radius of the Nycomed suspension is $R_2 = 2.77\,\mu$m. This value is sufficiently close to the actual mean radius of the Nycomed suspension (between 2.0 and 2.5 μm), i.e., Nycomed microbubbles will scatter ultrasound more or less efficiently at twice the transmitted frequency. Nevertheless, one can increase the quality of second-harmonic imaging from Nycomed microbubbles by a slight decrease in the transmitted frequency. As illustrated in Fig. 8(b), the optimal transmitted frequency for second-harmonic imaging is $f_2 = 4.61$ MHz, if the mean radius of Nycomed microbubbles is 2.0 μm. Figure 8(b) also shows that there is no optimal frequency for first-harmonic imaging. This is because 2.0 μm is less than the critical radius $R_{\text{crit}\,1} = 2.54\,\mu$m. It is necessary to say that the present analysis is expected to hold quantitatively for experiments on scattering by microbubbles under low-intensity continuous ultrasound. Finite-amplitude effects and polychromaticity of the ultrasound signal can lead to additional shifts in resonance frequencies.[13]

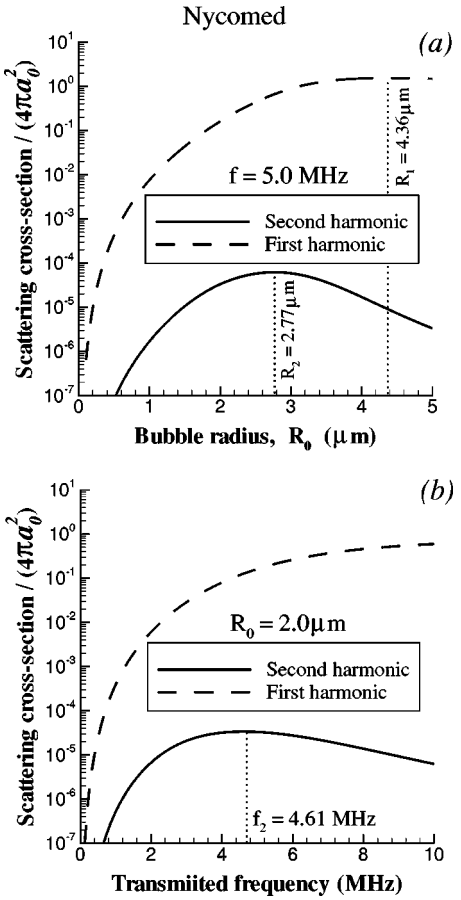


FIG. 8. First- and second-harmonic scattering cross sections normalized by $4\pi a_0^2$ (dashed and solid) as functions of (a) bubble radius R_0 at the transmitted frequency $f=5.0$ MHz and (b) transmitted frequency at $R_0=2.0$ μm. The parameters correspond to Nycomed microbubbles situated in blood plasma.

Because a lipid monolayer is very thin as compared to a polymeric or protein layer, the viscoelastic solid shell model does not work well for lipid-coated microbubbles. In this case, a Newtonian interface model suggested by Chatterjee and Sarkar[32] is more appropriate. This model introduces the interfacial dilatational viscosity κ_s instead of the shell viscosity μ_s. (In this model, an encapsulated shell has zero thickness.) Therefore lipid-coated microbubbles can be considered as free microbubbles with the decreased surface tension coefficient and the liquid viscosity replaced by

$$\mu_{lb}=\mu_l+\frac{\kappa_s}{a_0} \tag{35}$$

[see Eqs. (16) and (17) in Ref. 32]. Hence the viscous damping coefficient for lipid-coated microbubbles is calculated from Eq. (37), given below, where μ_{lb} is used instead of μ_l. The viscosity-induced shift in resonance frequency will clearly exist for these microbubbles because the second term on the right-hand side of Eq. (35) becomes very large when bubble sizes are in the micron or submicron range.

III. FREE MICROBUBBLES

Another type of the commercially available ultrasound contrast agents represents emulsions that transform itself into suspensions of free gas microbubbles inside the human body.[1] In particular, Echogen (Sonus Pharmaceuticals, Bothell, Washington) is an emulsion of liquid dodecafluoropentane (DDFP) which boils at about 28 °C.[1] In the liquid phase, such an emulsion is stable against dissolution. In the gas phase, it is more echogenic than a suspension of encapsulated microbubbles.

To evaluate backscattering by Echogen microbubbles at the transmitted frequency and twice the transmitted frequency, we can still use Eqs. (17) in which $R_0=a_0$, $\sigma_1=\sigma_2=\sigma$, $Z=0$, $p_{g0}=p_0+2\sigma/a_0$ (without the stabilizing layer, there is the over-pressure in the gas inside a microbubble), $\tau_{rr}^{(s)}=0$, $\rho_{s0}=\rho_{l0}$, and $\Delta_\rho=0$. In this case, the undamped natural frequency f_0 and the characteristic time $T_{\rm ac}$ are given by

$$f_0=\frac{1}{2\pi a_0}\sqrt{\frac{3\kappa p_0}{\rho_{l0}}+(3\kappa-1)\frac{2\sigma}{\rho_{l0}a_0}},\quad T_{ac}=\frac{a_0}{C_l}. \tag{36}$$

The viscous damping coefficient

$$\beta_{\rm vis}=\frac{2\mu_l}{\rho_{l0}a_0^2}. \tag{37}$$

The expressions for $S_{\rm ac}(\Omega)$ and $\beta_{\rm ac}(\Omega)$ remain unchanged [see Eqs. (19) and (21)].

The viscous damping coefficient of free microbubbles is smaller than that of encapsulated microbubbles. For example, a free bubble of radius 1 μm situated in blood plasma ($\rho_{l0}=1027\,{\rm kg\,m^{-3}}$, $\mu_l=0.0012\,{\rm Pa\,s}$) is characterized by $\beta_{\rm vis}\approx 2.337\times10^6\,{\rm s^{-1}}$, while the viscous damping coefficient of an Albunex® microbubble of the same (outer) radius is approximately $1.596\times10^8\,{\rm s^{-1}}$. Therefore, the deviation of the resonance frequency from the undamped natural frequency is expected to be less for Echogen than for Albunex®. However, it is not negligible. Like encapsulated bubbles, free bubbles are characterized by a critical radius for bubble resonance $R_{\rm crit}$. From (31) it follows that

$$R_{\rm crit}=\left(\frac{3\kappa-1}{3\kappa}\right)\frac{\sigma}{p_0}\left[-1+\sqrt{1+\frac{24\kappa\mu_l^2p_0}{(3\kappa-1)^2\sigma^2\rho_{l0}}}\right] \tag{38}$$

for free bubbles. If $\sigma=0$, the critical radius varies directly with the liquid viscosity μ_l,

$$R_{\rm crit}=\frac{2\sqrt{2}\mu_l}{\sqrt{3\kappa p_0\rho_{l0}}}. \tag{39}$$

The liquid viscosity has a more effect on the critical radius than the surface tension. According to Eq. (38), if the surface tension increases from 0 to $0.04\,{\rm N\,m^{-1}}$, the critical radius of bubbles oscillating almost isothermally ($\kappa=1.1$) in blood plasma decreases from 184 to 56 nm. An increase in the liquid viscosity from 0.0012 to 0.004 Pa s (high-shear-rate viscosity of blood) at $\sigma=0.04\,{\rm N\,m^{-1}}$ leads to increasing $R_{\rm crit}$ from 56 to 396 nm. Because the mean radius of Echogen microbubbles is about 75 nm,[17] a slight increase in viscosity and/or a decrease in surface tension, induced by soluble polymers and another surface-active molecules in blood, will make this suspension nonresonant. Experiments[17] show no first-harmonic resonance of the backscattered signal from

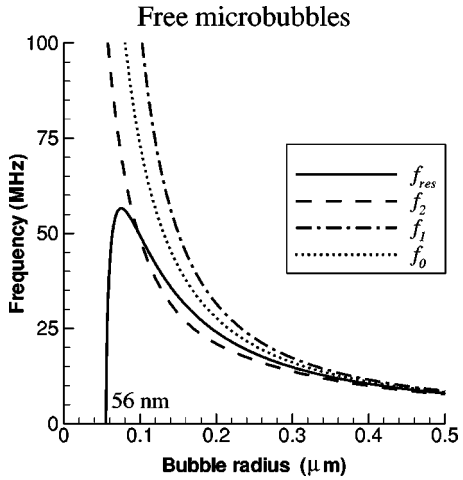


FIG. 9. Resonance and undamped natural frequencies of free microbubbles in blood plasma (solid and dotted) and first- and second-harmonic resonance frequencies of the backscattered signal from the microbubbles (dashed–dotted and dashed) as functions of the bubble radius. The blood plasma compressibility is taken into account ($C_l = 1543\ \mathrm{m\,s^{-1}}$ and $\rho_{l0} = 1027\ \mathrm{kg\,m^{-3}}$). The pressure wave amplitude $P_A = 30$ kPa. Other parameters: $\sigma = 0.04\ \mathrm{N\,m^{-1}}$, $\mu_l = 0.0012$ Pa s, $p_0 = 0.1$ MPa, and $\kappa = 1.1$. Free bubbles do not resonate if their size is less than $R_{\mathrm{crit}} = 56$ nm. The upper limit to the first-harmonic resonance frequency f_1 is 357 MHz (not shown here). The critical radius for the resonant backscatter at the first harmonic is $R_{\mathrm{crit1}} = 66$ nm.

these microbubbles. Note that the presence of a thin surfactant layer on the bubble surface, which is realized in lipid-shelled contrast agents and leads to a decrease in surface tension, will further magnify the critical radius and hence the difference between f_0 and f_{res}.

As illustrated in Fig. 9, the undamped natural frequency f_0 (dotted line) should not be used as the microbubble resonance frequency f_{res} (solid line) for free bubbles in blood plasma if the bubble radius is less than 0.5 μm. Also, in this size range, both the first- and second-harmonic resonance frequencies of the backscattered signal f_1 and f_2 (dashed–dotted and dashed lines) deviate strongly from the undamped natural frequency. In particular, $f_1 - f_0$ and $f_0 - f_2$ are equal to 0.28 MHz and 0.61 MHz at $a_0 = 0.5\ \mu$m but become 32.3 MHz and 24.8 MHz at $a_0 = 0.1\ \mu$m. The critical radius for f_1 at which the first resonance frequency reaches the upper limit (357 MHz) is 66 nm.

IV. CONCLUSION

The transmitted frequency at which the resonant excitation of microbubble oscillation occurs is taken to be equal to the undamped natural frequency in most experimental and theoretical works on scattering by contrast agents. We have shown that this assumption does not work for both free and encapsulated microbubbles used in contrast-enhanced ultrasound imaging. The resonance frequency of microbubbles deviates from the undamped natural frequency due to the increased viscous damping coefficient. The difference between these two frequencies rises as the bubble size decreases because the viscous damping coefficient is inversely proportional to the square of the bubble radius. The linear resonance of microbubble oscillation is characterized by a critical radius below which it cannot be observed. The critical radius of encapsulated microbubbles is typically between 1 and 3 μm and increases with increasing the shell thickness. The critical radius of free bubbles depends significantly on the liquid viscosity and surface tension. In blood plasma, free bubbles cease to resonate if their radius is below 56 nm.

The deviation of the resonance frequency from the undamped natural frequency results in the resonant backscatter by nonresonating microbubbles. The resonances of the backscattered signal (resonance peaks in the output-level versus frequency curves) are characterized by the frequencies different from both the microbubble resonance frequency and the undamped natural frequency. Every harmonic of the backscattered signal has its own resonance frequency. In particular, a monodisperse suspension of microbubbles cannot produce the resonant backscatter at both the transmitted frequency and twice the transmitted frequency. (Because a microbubble suspension is polydisperse, there exist bubbles of one size that produce the resonant backscatter at the transmitted frequency and bubbles of another size that produce the resonant backscatter at twice the transmitted frequency.) The difference between the first- and second-harmonic resonance frequencies of the backscattered signal f_1 and f_2 reaches several MHz in the case of encapsulated microbubbles of radius less than 3 μm. To increase the quality of contrast-enhanced ultrasound imaging, the transducer should operate at the frequency f_1 in the fundamental mode and at the frequency f_2 in the second-harmonic mode, where f_1 and f_2 are taken at the mean radius of the ultrasound contrast agent.

Also, there exists a critical radius for the first-harmonic resonance of the backscattered signal, larger than the critical radius for the bubble resonance. The microbubbles of this radius produce the resonant backscatter at the transmitted frequency equal to the upper limit to the first-harmonic resonance frequency. There is no critical radius for the second harmonic resonance of the backscattered signal.

The resonance peak in the first harmonic of the backscattered signal is not observed in many experiments with ultrasound contrast agents. It is worth to mention Refs. 33 (Figs. 4, 5, and 7), 34 [Figs. 4(a) and 9(b)], and 25 (Fig. 5, see also Fig. 9). Indirect evidence for the disappearance of the bubble resonance (and hence of the first-harmonic resonance of the backscattered signal) is presented in experiments by Chen *et al.*[35] As noted by Hoff *et al.*[25] "the resonance frequency increases and the resonance peak broadens and almost disappears due to the viscoelastic polymeric shell."

Our next concerns will be to incorporate bubble polydispersity and pulsed driving conditions (with different pulse shapes and lengths) into the theoretical model and performing a direct comparison between theoretical predictions and experimental data on the resonance frequencies of the backscattered signal from ultrasound contrast agents.

ACKNOWLEDGMENTS

The author wishes to thank Ronald A. Roy and Ali Nadim for fruitful discussions. This work was partially supported by the North Atlantic Treaty Organization under Grant No. DGE-0000779. Any opinions, findings, conclusions or

recommendations expressed in this publication are those of the author and do not necessarily reflect the view of NATO.

[1] D. Cosgrove, ''Echo enhancers and ultrasound imaging,'' Eur. J. Radiol. **26**, 64–76 (1997).
[2] F. Calliada, R. Campani, O. Bottinelli, A. Bozzini, and M. G. Sommaruga, ''Ultrasound contrast agents. Basic principles,'' Eur. J. Radiol. **27**, S157–S160 (1998).
[3] H. Becher and P. N. Burns, *Handbook of Contrast Echocardiography: Left Ventricular Function and Myocardial Perfusion* (Springer-Verlag, New York, 2000).
[4] P. N. Burns, D. H. Simpson, and M. A. Averkiou, ''Nonlinear imaging,'' Ultrasound Med. Biol. **26**, S19–S22 (2000).
[5] P. J. A. Frinking, A. Bouakaz, J. Kirkhorn, F. J. Ten Cate, and N. de Jong, ''Ultrasound contrast imaging: Current and new potential methods,'' Ultrasound Med. Biol. **26**, 965–975 (2000).
[6] N. de Jong, P. J. A. Frinking, A. Bouakaz, M. Goorden, T. Schourmans, X. Jingping, and F. Mastik, ''Optical imaging of contrast agent microbubbles in an ultrasound field with a 100-mhz camera,'' Ultrasound Med. Biol. **26**, 487–492 (2000).
[7] M. S. Plesset and A. Prosperetti, ''Bubble dynamics and cavitation,'' Annu. Rev. Fluid Mech. **9**, 145–185 (1977).
[8] N. de Jong and L. Hoff, ''Ultrasound scattering of Albunex® microspheres,'' Ultrasonics **31**, 175–181 (1993).
[9] C. C. Church, ''The effects of an elastic solid surface layer on the radial pulsations of gas bubbles,'' J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[10] C. T. Chin and P. N. Burns, ''Predicting the acoustic response of a microbubble population for contrast imaging in medical ultrasound,'' Ultrasound Med. Biol. **26**, 1293–1300 (2000).
[11] D. B. Khismatullin and A. Nadim, ''Radial oscillations of encapsulated microbubbles in viscoelastic liquids,'' Phys. Fluids **14**, 3534–3557 (2002).
[12] J. S. Allen, D. E. Kruse, P. A. Dayton, and K. W. Ferrara, ''Effect of coupled oscillations on microbubble behavior,'' J. Acoust. Soc. Am. **114**, 1678–1690 (2003).
[13] S. Hilgenfeldt, D. Lohse, and M. Zomack, ''Response of bubbles to diagnostic ultrasound: a unifying theoretical approach,'' Eur. Phys. J. B **4**, 247–255 (1998).
[14] T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[15] C. E. Brennen, *Cavitation and Bubble Dynamics* (Oxford University Press, New York, 1995).
[16] A. L. Klibanov, ''Targeted delivery of gas-filled microspheres, contrast agents for ultrasound imaging,'' Advanced Drug Delivery Reviews **37**, 139–157 (1999).
[17] E. C. Everbach, R. A. Roy, and J. T. Flaherty, ''Quantification of submicron DDFP gas bubbles using 30 mhz backscattered ultrasound,'' J. Acoust. Soc. Am. **110**, 2615 (2001).
[18] R. E. Apfel, ''Radiation and ultrasound-triggered bubble contrast agents based on superheated drops,'' J. Acoust. Soc. Am. **108**, 2518 (2000).
[19] P. S. Epstein and M. S. Plesset, ''On the stability of gas bubbles in liquid-gas solutions,'' J. Chem. Phys. **18**, 1505–1509 (1950).
[20] R. B. Bird, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids, Vol. 1: Fluid Mechanics* (Wiley, New York, 1987).
[21] I. S. Akhatov, N. K. Vakhitova, G. Y. Galeyeva, R. I. Nigmatulin, and D. B. Khismatullin, ''Weak oscillations of a gas bubble in a spherical volume of compressible liquid,'' PMM-J. Appl. Math. Mech. **61**, 921–930 (1997).
[22] R. I. Nigmatulin, I. S. Akhatov, N. K. Vakhitova, and J. R. T. Lahey, ''On the forced oscillations of a small gas bubble in a spherical liquid-filled flask,'' J. Fluid Mech. **414**, 47–73 (2000).
[23] A. Prosperetti and A. Lezzi, ''Bubble dynamics in a compressible liquid. Part 1. First-order theory,'' J. Fluid Mech. **168**, 457–478 (1986).
[24] W. T. Coakley and L. N. Nyborg, ''Cavitation: Dynamics of gas bubbles: Application,'' in *Ultrasound: Its Applications in Medicine and Biology: Part 1*, edited by F. J. Frey (Elsevier, New York, 1978), pp. 77–159.
[25] L. Hoff, P. C. Sontum, and J. M. Hovem, ''Oscillations of polymeric microbubbles: Effect of the encapsulating shell,'' J. Acoust. Soc. Am. **107**, 2272–2280 (2000).
[26] A. Prosperetti, ''Thermal effects and damping mechanisms in the forced radial oscillations of gas bubbles in liquids,'' J. Acoust. Soc. Am. **61**, 17–27 (1977).
[27] K. Bjerknes, P. Sontum, G. Smistad, and I. Agerkvist, ''Preparation of polymeric microbubbles: Formulation studies and product characterization,'' Int. J. Pharm. **158**, 129–136 (1997).
[28] T. D. Mast, ''Empirical relationship between acoustic parameters in human soft tissues,'' ARLO **1**, 37–42 (2000).
[29] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972).
[30] N. de Jong, ''Acoustic properties of ultrasound contrast agents,'' Ph.D. thesis, Erasmus University, Rotterdam, The Netherlands, 1993.
[31] D. L. Miller, ''Frequency relationships for ultrasonic activation of free microbubbles, encapsulated microbubbles, and gas-filled micropores,'' J. Acoust. Soc. Am. **104**, 2498–2505 (1998).
[32] D. Chatterjee and K. Sarkar, ''A newtonian rheological model for the interface of microbubble contrast agents,'' Ultrasound Med. Biol. **29**, 1749–1757 (2003).
[33] P. J. A. Frinking and N. de Jong, ''Acoustic modeling of shell-encapsulated gas bubbles,'' Ultrasound Med. Biol. **24**, 523–533 (1998).
[34] P. J. A. Frinking, N. de Jong, and E. I. Cespedes, ''Scattering properties of encapsulated gas bubbles at high ultrasound pressures,'' J. Acoust. Soc. Am. **105**, 1989–1996 (1999).
[35] W.-S. Chen, T. J. Matula, and L. A. Crum, ''The disappearance of ultrasound contrast agents: Observations of bubble dissolution and cavitation nucleation,'' Ultrasound Med. Biol. **28**, 793–803 (2002).

Frequency-dependent acoustic properties of a fluid/porous solid interface

A. A. Gubaidullin and O. Y. Kuchugurina
Tyumen Branch of the Institute of Theoretical and Applied Mechanics, SB RAS, Tyumen, 625000, Russia

D. M. J. Smeulders[a] and C. J. Wisse
Delft University of Technology, 2600 GA Delft, The Netherlands

(Received 1 October 2003; revised 28 May 2004; accepted 4 June 2004)

We investigate numerically the frequency dependence of the velocities and damping coefficients of the various surface waves that exist on the interface between a fluid half-space and a half-space of a water-saturated porous Biot medium including viscous attenuation. We find that there is a f^1 frequency dependence for pure radiation damping. For a sealed-pore surface configuration, the pseudo-Stoneley wave is radiating at high frequencies, but becomes purely viscosity-damped as frequency decreases. For the open-pore surface configuration, the pseudo-Rayleigh wave remains radiative over the entire frequency domain and is damped by a combination of radiation and viscous fluid–solid interaction. It is also found that the pseudo-Rayleigh wave may even become faster than the bulk shear mode, at low frequencies. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777856]

PACS numbers: 43.35.Pt [RAS] Pages: 1474–1480

I. INTRODUCTION

For the interpretation of acoustic borehole logs a thorough understanding of the effects of rock properties on acoustic wave propagation is required. Examples of these properties are the porosity, permeability, lithology, and hydrocarbon saturation. In conventional acoustic borehole logging the first arrival of the acoustic wave train can be used to determine the porosity. The full wave train is applied to determine mechanical properties, while one of the late arrivals, the Stoneley wave, is used to obtain permeability data.[1,2] Rosenbaum[3] combined the theory of wave modes with the Biot theory[4] in order to describe the wave modes in a borehole. Schmitt *et al.*,[5] Liu,[6] and Winkler *et al.*[7] demonstrated the permeability dependence of the Stoneley wave. It was argued that the permeability dependence of the Stoneley wave could be explained by its relation to the Biot slow wave. It is obvious that a thorough understanding of the borehole acoustics requires an adequate description of the macroscopic boundary conditions at an interface between a fluid and a porous solid, because these boundary conditions govern the amplitudes at which the fast, slow, and shear waves are generated. We will study these boundary conditions at a plane interface, which has the advantage that there is no apparent lengthscale in the problem other than the wavelength. The number of wave modes in this case is limited which is not the case for borehole configurations (see, e.g., Wisse *et al.*[8]). One common assumption for the boundary conditions is that the pores are open to the fluid, and that the acoustic pore pressure equals the acoustic pressure in the fluid: the so-called open-pore boundary condition. Another is that the surface is sealed so that the fluid cannot flow relative to the solid, at the surface: the so-called sealed-pore boundary condition. More generally, one may model the surface boundary by means of the so-called surface flow impedance T, in which $T\rightarrow 0$ and $T\rightarrow\infty$ are the special cases mentioned above.[9] The first to describe surface waves for the open pore case was Deresiewicz,[10] where it is made clear that the surface wave has components of all three bulk waves. The surface waves are, however, very sensitive to the value of T. These properties were derived by Feng and Johnson,[11] who show that there is an additional surface acoustic wave for large values of T that is not present for small values of T. Subsequently, the new surface acoustic wave was observed by Nagy[12] and Adler and Nagy.[13]

In order to define the various surface modes, we will first discuss some analogues. At the interface between a homogeneous isotropic solid and a vacuum there is always one and only one surface mode: the Rayleigh wave. It has a velocity slightly lower than the bulk shear wave and it is not attenuated as it propagates along the solid surface.[14] If the vacuum is replaced by a fluid whose *P*-wave velocity is less than the Rayleigh wave velocity, the Rayleigh wave is attenuated by radiating into the fluid and becomes a pseudo-Rayleigh wave, faster than the fluid bulk wave but slower than the other bulk modes.[15–17] The radiation condition requires the pseudo-Rayleigh mode to have an exponentially growing amplitude factor instead of a decaying one as one moves away from the interface into the fluid. Although this seems physically unrealistic, it was shown by van der Hijden[18] that if the pole of the secular equation is close to the real axis, its effect is still seen in the Green's function, it is experimentally observable[19] and hence relevant to consider this pseudosurface mode. The true surface mode, the Stoneley wave, has a velocity which is, in this case, slightly less than the fluid *P*-wave velocity and it is not attenuated as it propagates along the solid/fluid interface. Due to contributions of Scholte,[20] some authors refer to this wave as the Scholte wave.[18,21,22] The Stoneley wave between two solids exists only for limited range of elasticity parameters of the

[a]Electronic mail: d.m.j.smeulders@citg.tudelft.nl

 © 2004 Acoustical Society of America

solids, while Scholte[20] showed that the surface wave along a solid/liquid interface exists for arbitrary parameters of the solid and the fluid.

If, now, the solid is a porous solid that has its bulk slow wave slower than the fluid bulk wave, it was shown by Feng and Johnson[11] that the Stoneley wave loses energy by radiating into the bulk slow wave, and becomes a pseudo-Stoneley mode. It is, in this case, faster than the bulk slow wave but slower than the rest of the bulk modes. Depending on the boundary conditions and solid-to-fluid stiffness ratio, there can be true surface modes, pseudo-Rayleigh and pseudo-Stoneley modes, or there can be one or two of those.

In this paper, we analyze the frequency dependence of the velocities and attenuation coefficients of these acoustic surface modes. In this respect our work is a sequel to the work of Feng and Johnson,[11] who only considered the high-frequency limit where the Biot–Plona slow wave is propagatory and the porous medium is assumed lossless (in the lossless condition the pore fluid is ideal, i.e., its viscosity $\eta=0$ and thus all bulk modes are undamped). In this case the true surface modes are not damped as they propagate along the solid/fluid interface, and the radiation damping of the pseudosurface modes displays f^1 frequency behavior (as we will see). However at low frequencies, i.e., below the rollover frequency f_c defined in (14), the Biot–Plona slow wave is no longer propagatory but becomes diffusive. As the surface waves have components of the slow wave, it is therefore of practical interest[8] to investigate their frequency-dependent behavior. In order to do so, we will incorporate the viscous interaction between the fluid and the solid as described by Johnson *et al.*[23] This means that also nonradiating true modes will become damped at all frequencies and that radiating modes are damped by a combination of viscous and radiation damping. In the case that the fast and the slow wave are strongly decoupled (for air-saturated porous media), surface waves including viscous attenuation have been described before.[24–26] Measurements confirmed the existence of surface waves propagating slower than homogeneous plane waves.

In Sec. II, we describe the main features of the Biot formalism. The properties of the acoustic surface modes are discussed in Sec. III, and the boundary conditions and secular equations in Sec. IV. The main results of the paper are presented in Secs. V (sealed-pore boundary conditions) and VI (open-pore boundary conditions). We summarize our results in Sec. VII. In this paper we consider the velocities and damping coefficients of the various surface modes. We do not wish to consider the amplitudes with which they are generated.

II. ACOUSTIC BULK PROPERTIES

The equations of motion for a fluid-saturated porous medium were formulated by Biot.[4] It is assumed that for long-wavelength disturbances ($\lambda \gg a$, where λ is the wavelength and a is a characteristic pore scale) we can define average values of the local displacements $\mathbf{u}(\mathbf{r},t)$ in the solid and $\mathbf{U}(\mathbf{r},t)$ in the fluid. If we consider a cube of unit size of bulk material (porosity ϕ), we will denote τ_{ij} the forces per unit bulk area applied to that portion of the cube faces occupied by the solid. They are a result of both the fluid pressure p and the intergranular stresses σ_{ij}:

$$\tau_{ij} = -\sigma_{ij} - (1-\phi)p\,\delta_{ij}, \tag{1}$$

where the Kronecker delta reflects the fact that the pore fluid cannot exert nor sustain any shear forces. The total normal tension force per unit bulk area applied to the fluid faces of the unit cube is denoted τ,

$$\tau = -\phi p. \tag{2}$$

In the case of isotropic materials, the stress–strain relations for the solid and the fluid may be written[4]

$$\tau_{ij} = 2Ne_{ij} + Ae_{kk}\delta_{ij} + Q\epsilon_{kk}\delta_{ij}, \tag{3}$$

$$\tau = Qe_{kk} + R\epsilon_{kk}, \tag{4}$$

where $e_{ij} = 1/2(\partial u_i/\partial x_j + \partial u_j/\partial x_i)$, $\epsilon_{ij} = 1/2(\partial U_i/\partial x_j + \partial U_j/\partial x_i)$, and summation over repeated indices is assumed. A, Q, and R are generalized elastic parameters which can be related via Gedanken experiments[27] to porosity, bulk modulus of the solid K_s, bulk modulus of the fluid K_f, bulk modulus of the porous drained matrix K_b, and shear modulus N of both the drained matrix and of the composite

$$A = K_b - 2N/3 + K_f(1-\phi-K_b/K_s)^2/\phi_{\text{eff}}, \tag{5}$$

$$Q = \phi K_f(1-\phi-K_b/K_s)/\phi_{\text{eff}}, \tag{6}$$

$$R = \phi^2 K_f/\phi_{\text{eff}}, \tag{7}$$

where $\phi_{\text{eff}} = \phi + K_f/K_s(1-\phi-K_b/K_s)$ is an effective porosity. The equations of motion resulting from momentum conservation and the stress–strain relations can be written as[4]

$$N\nabla^2\mathbf{u} + (A+N)\nabla\nabla\cdot\mathbf{u} + Q\nabla\nabla\cdot\mathbf{U} = \frac{\partial^2}{\partial t^2}(\rho_{11}\mathbf{u} + \rho_{12}\mathbf{U}) + b_0 F\frac{\partial}{\partial t}(\mathbf{u}-\mathbf{U}), \tag{8}$$

$$Q\nabla\nabla\cdot\mathbf{u} + R\nabla\nabla\cdot\mathbf{U} = \frac{\partial^2}{\partial t^2}(\rho_{12}\mathbf{u} + \rho_{22}\mathbf{U}) - b_0 F\frac{\partial}{\partial t}(\mathbf{u}-\mathbf{U}). \tag{9}$$

Introducing the solid and fluid densities ρ_s and ρ_f, and the tortuosity α_∞, the effective densities are given by

$$\rho_{12} = -(\alpha_\infty - 1)\phi\rho_f, \tag{10}$$

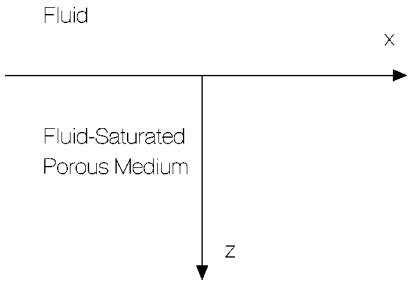


FIG. 1. Geometry of the problem. The half-space $z<0$ is filled with fluid. The half-space $z>0$ is filled with fluid-saturated porous material.

TABLE I. Parameter values used in computations.

Constained modulus K_p	16.08 GPa
Shear modulus N	5.36 GPa
Porosity ϕ	0.33
Permeability k_0	308 Darcy
Tortuosity α_∞	2
Grain bulk modulus K_s	36.5 GPa
Grain density ρ_s	2620 kg/m^3
Fluid bulk modulus K_f	2.2 GPa
Fluid density ρ_f	998 kg/m^3
Fluid viscosity η	0.001Pa·s
Rollover frequency f_c	85.43 Hz

$$\rho_{11}=(1-\phi)\rho_s-\rho_{12}, \tag{11}$$

$$\rho_{22}=\phi\rho_f-\rho_{12}=\alpha_\infty\phi\rho_f. \tag{12}$$

The viscous damping factor is denoted $b_0=\eta\phi^2/k_0$, with η the dynamic fluid viscosity and k_0 the zero-frequency permeability. The viscous correction factor F was formulated in its generalized form,[23]

$$F=\sqrt{1+\tfrac{1}{2}iMf/f_c}, \tag{13}$$

where M is the shape factor (usually taken 1). The rollover frequency is defined as

$$f_c=\frac{\eta\phi}{2\pi\alpha_\infty\rho_f k_0}. \tag{14}$$

Using standard decoupling techniques,[4,28] it can be shown that there are two longitudinal wave modes (the fast and the slow compressional waves), and one shear wave mode. The dispersion relation for the compressional wave numbers k can effectively be written in terms of $\zeta=k^2/\omega^2$,

$$(PR-Q^2)\zeta^2-(P\tilde{\rho}_{22}+R\tilde{\rho}_{11}-2Q\tilde{\rho}_{12})\zeta+(\tilde{\rho}_{11}\tilde{\rho}_{22}-\tilde{\rho}_{12}^2)=0, \tag{15}$$

where $P=A+2N$. The density factors and the damping factor b_0F always appear together. It is therefore convenient to define complex-valued, frequency-dependent, density factors

$$\tilde{\rho}_{12}=\rho_{12}+ib_0F/\omega, \tag{16}$$

$$\tilde{\rho}_{11}=\rho_{11}-ib_0F/\omega, \tag{17}$$

$$\tilde{\rho}_{22}=\rho_{22}-ib_0F/\omega. \tag{18}$$

The dispersion relation for the shear wave is given by

$$N\tilde{\rho}_{22}\zeta-(\tilde{\rho}_{11}\tilde{\rho}_{22}-\tilde{\rho}_{12}^2)=0. \tag{19}$$

TABLE II. High-frequency velocity values.

	Sealed pores	Open pores
$V_{\rm sh}$ (m/s)	1670.8	1670.8
V_{pR} (m/s)		1576.9
V_f (m/s)	1484.7	1484.7
V_{pS} (m/s)	1333.4	1371.9
V_{p2} (m/s)	943.1	943.1
V_{tS} (m/s)	911.0	

Each wave propagates simultaneously in the pore fluid and in the frame. The fluid-to-solid amplitude ratios β_{p1}, β_{p2}, and $\beta_{\rm sh}$ are given by

$$\beta_j=\frac{\tilde{\rho}_{11}-P\zeta_j}{Q\zeta_j-\tilde{\rho}_{12}}, \quad j=p1,p2, \tag{20}$$

$$\beta_{\rm sh}=-\frac{\tilde{\rho}_{12}}{\tilde{\rho}_{22}}. \tag{21}$$

III. ACOUSTIC SURFACE MODES

We consider the geometry depicted in Fig. 1. The half-space $z>0$ is filled with porous material saturated with fluid. The half-space $z<0$ is filled with that same fluid. Let $\mathbf{j}$ be the unit vector in the y direction perpendicular to the XOZ plane. Let φ_{p1} and φ_{p2} be the potentials for the compressional waves, and $\mathbf{j}\psi_{\rm sh}$ for the shear wave. Let φ_f be the potential for the fluid half-space. The displacements of the solid and the fluid in the lower half-space and the fluid displacement $\mathbf{U}_f(\mathbf{r},t)$ in the upper half-space are given by, respectively,

$$\mathbf{u}=\nabla\varphi_{p1}+\nabla\varphi_{p2}+\nabla\times\mathbf{j}\psi_{\rm sh}, \tag{22}$$

$$\mathbf{U}=\beta_{p1}\nabla\varphi_{p1}+\beta_{p2}\nabla\varphi_{p2}+\beta_{\rm sh}\nabla\times\mathbf{j}\psi_{\rm sh}, \tag{23}$$

$$\mathbf{U}_f=\nabla\varphi_f. \tag{24}$$

A true surface wave which propagates along the porous medium/fluid interface is a linear combination of all bulk waves with exponential decay in the z direction. We therefore express the potentials as

$$\varphi_{p1}=A_1e^{-\gamma_{p1}z}\exp i(\omega t-k_xx), \quad z>0, \tag{25}$$

$$\varphi_{p2}=A_2e^{-\gamma_{p2}z}\exp i(\omega t-k_xx), \quad z>0, \tag{26}$$

$$\psi_{\rm sh}=iBe^{-\gamma_{\rm sh}z}\exp i(\omega t-k_xx), \quad z>0, \tag{27}$$

$$\varphi_f=A_fe^{\gamma_f z}\exp i(\omega t-k_xx), \quad z<0, \tag{28}$$

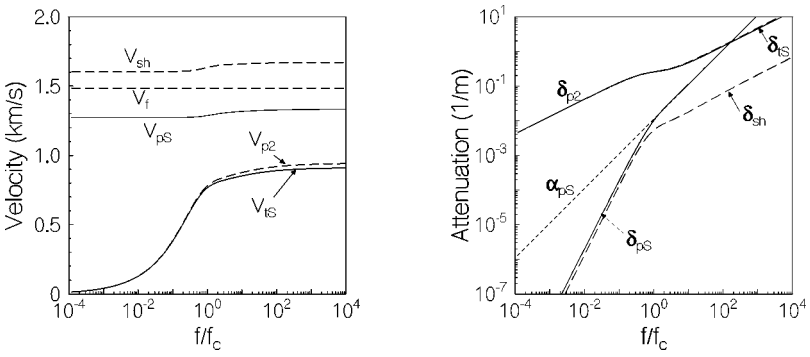


FIG. 2. Surface wave velocities (left, solid lines) and attenuation coefficients (right, solid lines) for the sealed-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I. Dashed lines refer to bulk modes. The dotted line α_{pS} (right) refers to frictionless attenuation (pure radiation).

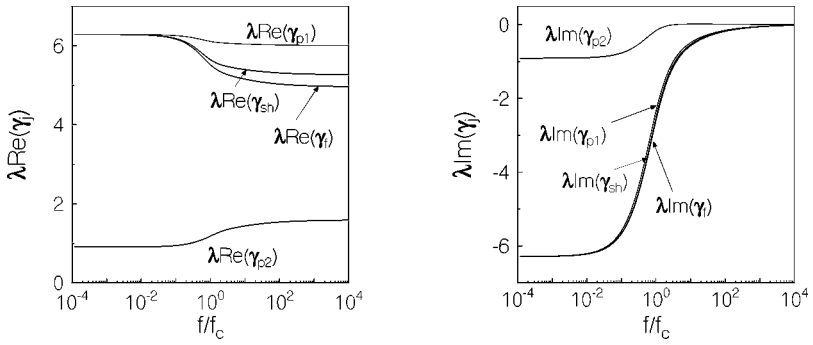


FIG. 3. Real (left) and imaginary parts (right) of the normal decrements for the true Stoneley wave (wavelength λ) under sealed-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I.

where A_1, A_2, B, and A_f are the amplitude factors, and $\gamma_j \equiv (k_x^2 - k_j^2)^{1/2}$ for $j = p1$, $p2$, sh, f. A true surface mode has the property that for a frictionless porous medium [all four (k_j)'s are now real valued] its velocity is real and all (γ_j)'s are real, and thus its velocity is less than all four bulk (V_j)'s. When the porous medium becomes lossy [k_j becomes complex-valued with $\mathrm{Re}(k_j)>0$ and $\mathrm{Im}(k_j)<0$ for $j = p1$, $p2$, sh], this velocity constraint no longer holds (as we will see).

The pseudo-Stoneley wave has a frictionless velocity that is higher than the slowest bulk mode but less than all other bulk modes. The pseudo-Rayleigh wave is, under frictionless conditions, faster than both the slowest bulk mode and the fluid bulk mode. These velocity relations imply that the pseudo-Stoneley wave radiates into the slowest bulk mode and that the pseudo-Rayleigh wave radiates into both the slowest bulk mode and the fluid bulk mode. Because of energy radiation, these pseudomodes are damped as they propagate along the interface, i.e., $\mathrm{Im}(k_x)<0$. From the definition $k_x^2 - k_j^2 = \gamma_j^2$, we know that $\mathrm{Re}(k_x)\cdot\mathrm{Im}(k_x) = \mathrm{Re}(\gamma_j)\cdot\mathrm{Im}(\gamma_j)$, so that the radiation and attenuation conditions imply that $\mathrm{Re}(\gamma_j)<0$ and $\mathrm{Im}(\gamma_j)>0$ for the radiating bulk modes. Usually, also the condition (for the radiating bulk modes) $|\mathrm{Re}(\gamma_j)| \ll |\mathrm{Im}(\gamma_j)|$ is satisfied, corresponding to "slightly" leaky pseudosurface waves. Under frictionless conditions, the (γ_j)'s of the radiating bulk modes are therefore in the second quadrant of the Argand diagram. We will show that as we go down in frequency for lossy porous media, the pseudomodes no longer necessarily radiate energy and also become damped because of the viscous effects. As k_j becomes complex valued, the relation $\mathrm{Re}(k_x)\cdot\mathrm{Im}(k_x) = \mathrm{Re}(\gamma_j)\cdot\mathrm{Im}(\gamma_j)$ no longer holds. We will show that the (γ_j)'s of the modes that radiate under frictionless conditions pass from the second to the first quadrant, and thus cease to radiate, as we go down in frequency.

IV. BOUNDARY CONDITIONS AND SECULAR EQUATIONS

For the open-pore boundary conditions at the interface $z=0$, we find that the porous material is stress free, and that there is continuity of pressure and volume flux over the interface. Defining the pressure in the fluid half-space p_f, we may write accordingly

$$\sigma_{zz} = \sigma_{zx} = 0, \tag{29}$$

$$p = p_f, \tag{30}$$

$$(1-\phi)u_z + \phi U_z = U_{fz}. \tag{31}$$

From the definitions of stresses and pressure (1), (2), and from the stress–strain relations (3), (4), we find that

$$-\sigma_{ij} = N\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + (A'\boldsymbol{\nabla}\cdot\mathbf{u} + Q'\boldsymbol{\nabla}\cdot\mathbf{U})\,\delta_{ij}, \tag{32}$$

$$-\phi p = Q\boldsymbol{\nabla}\cdot\mathbf{u} + R\boldsymbol{\nabla}\cdot\mathbf{U}, \tag{33}$$

where $A' = A - (1-\phi)Q/\phi$ and $Q' = Q - (1-\phi)R/\phi$. Now substituting (22), (23) into (32), (33), we find that

$$-\sigma_{zz} = 2N\left[\frac{\partial^2(\varphi_{p1}+\varphi_{p2})}{\partial z^2} + \frac{\partial^2\psi_{\mathrm{sh}}}{\partial x\,\partial z}\right] + (A' + Q'\beta_{p1})\boldsymbol{\nabla}^2\varphi_{p1} + (A' + Q'\beta_{p2})\boldsymbol{\nabla}^2\varphi_{p2}, \tag{34}$$

$$-\sigma_{zx} = N\left[2\frac{\partial^2(\varphi_{p1}+\varphi_{p2})}{\partial x\,\partial z} + \frac{\partial^2\psi_{\mathrm{sh}}}{\partial x^2} - \frac{\partial^2\psi_{\mathrm{sh}}}{\partial z^2}\right], \tag{35}$$

$$-\phi p = (Q + R\beta_{p1})\nabla^2\varphi_{p1} + (Q + R\beta_{p2})\nabla^2\varphi_{p2}. \tag{36}$$

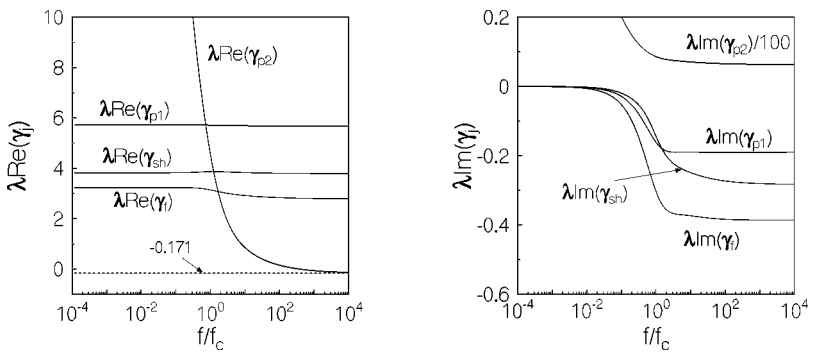


FIG. 4. Real (left) and imaginary parts (right) of the normal decrements for the pseudo-Stoneley wave (wavelength λ) under sealed-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I. The dotted line (left, −0.171) refers to the frictionless γ_{p2}.

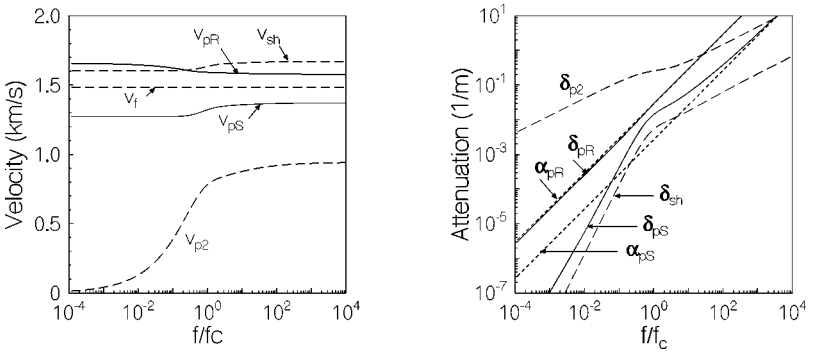


FIG. 5. Surface wave velocities (left, solid lines) and attenuation coefficients (right, solid lines) for the open-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I. Dashed lines refer to bulk modes. Dotted lines α_{pR} and α_{pS} (right) refer to frictionless attenuation (pure radiation).

The z components of the fluid and solid velocities u_z, U_z, and U_{fz} easily follow from (22)–(24). The fluid pressure in the upper half-space is related to the potential φ_f,

$$p_f = \omega^2 \rho_f \varphi_f . \tag{37}$$

Now substituting (25)–(28) into (34)–(37), the stresses, pressures, and displacements are expressed in terms of the unknown amplitude factors A_1, A_2, B, A_f. Subsequent substitution into the boundary conditions (29)–(31) yields the equations

$$\underline{\mathbf{M}}_o \cdot (A_1, A_2, B, A_f)^T = 0, \tag{38}$$

where the elements of the matrix $\underline{\mathbf{M}}_o$ are given in the appendix.

For the sealed interface, the following conditions apply:

$$\sigma_{zz} + p = p_f , \tag{39}$$

$$\sigma_{zx} = 0, \tag{40}$$

$$u_z = U_z = U_{fz} . \tag{41}$$

From these boundary conditions the equations

$$\underline{\mathbf{M}}_s \cdot (A_1, A_2, B, A_f)^T = 0 \tag{42}$$

are obtained, where the elements of the matrix $\underline{\mathbf{M}}_s$ are given in the appendix as well. The zero's of the determinants of the matrices $\underline{\mathbf{M}}_o$ and $\underline{\mathbf{M}}_s$ are determined numerically, subject to the branch cut restrictions on the square roots discussed in the preceding section.

V. SEALED-PORE BOUNDARY CONDITIONS

We take the parameters for the porous solid and the fluid from Table I. An overview of the computed high-frequency bulk and surface velocities is given in Table II, and will be discussed in the forthcoming. For the sealed-pore conditions, we find the true Stoneley (tS) wave and the pseudo-Stoneley (pS) wave. The pseudo-Rayleigh wave is not found as it only exists for higher values of K_p.[11] Velocities and damping coefficients of the surface modes are plotted in Fig. 2 as a function of frequency. The total attenuation coefficients are denoted δ whereas the frictionless (radiation) damping is denoted α. The true surface mode is slower than all the bulk modes. We find that it closely resembles the bulk slow wave, both in velocity and attenuation. At high frequencies it is propagatory with a limiting velocity of 911.0 m/s, which is slightly lower than the limiting V_{p2} (943.1 m/s). At low frequencies the true mode becomes diffusive with a $f^{1/2}$ frequency dependence of velocity and attenuation. Real and imaginary parts of the corresponding (γ_j)'s are plotted in Fig. 3. As the true mode does not radiate, $\mathrm{Re}(\gamma_j) > 0$. In the high-frequency limit, we must therefore have that $\mathrm{Im}(\gamma_j) < 0$. We find that in fact $\mathrm{Im}(\gamma_j)$ is always smaller than zero, except for $\mathrm{Im}(\gamma_{p2})$ that attains a (small) positive maximum before falling back to the fourth quadrant in an Argand diagram.

The pseudo-Stoneley wave has a velocity that is always higher than the bulk slow wave and lower than all other bulk modes. We find that it has little velocity dispersion (see Fig. 2, left). The attenuation of the pseudo-Stoneley mode displays f^2 frequency dependence at low frequencies, catching up to frictionless f^1 behavior for higher frequencies (see Fig. 2, right). We find that at high frequencies the pseudo-Stoneley wave is attenuated by radiating into the bulk slow wave, but for lower frequencies the radiation ceases, and it becomes purely viscosity damped. This can be seen in Fig. 4, where the real and imaginary parts of the (γ_j)'s are plotted. We notice that for decreasing frequency, γ_{p2} passes from the second, radiative, quadrant to the first, nonradiative, quadrant

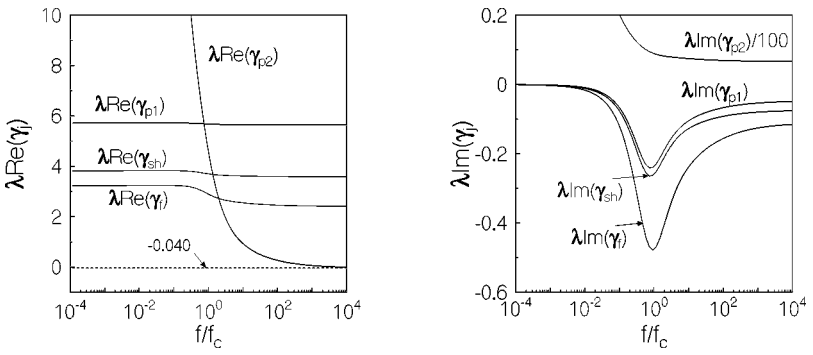


FIG. 6. Real (left) and imaginary parts (right) of the normal decrements for the pseudo-Stoneley wave (wavelength λ) under open-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I. The dotted line (left, −0.040) refers to the frictionless γ_{p2}.

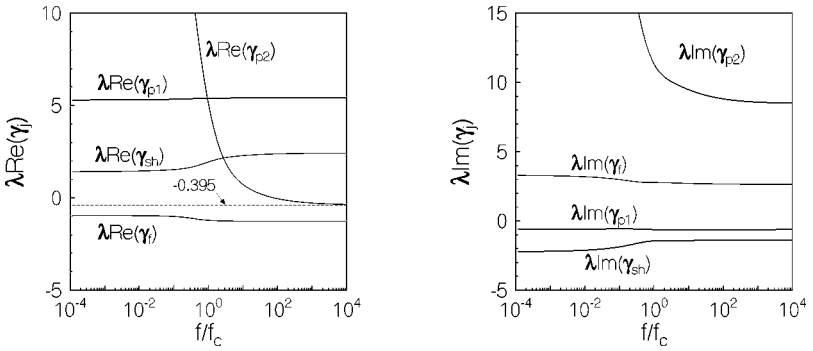


FIG. 7. Real (left) and imaginary parts (right) of the normal decrements for the pseudo-Rayleigh wave (wavelength λ) under open-pore boundary conditions, as a function of reduced frequency. The parameters are taken from Table I. The dotted line (left, −0.395) refers to the frictionless γ_{p2}.

in an Argand diagram. All other (γ_j)'s are restricted to the fourth, nonradiative, quadrant.

VI. OPEN-PORE BOUNDARY CONDITIONS

For the open-pore boundary conditions, we find a pseudo-Stoneley wave and a pseudo-Rayleigh wave. We do not find the true mode, which only exists for a finite range of relatively low K_p values.[11] Velocities and damping coefficients are plotted in Fig. 5. Under frictionless conditions, the pseudo-Stoneley wave is slightly faster (1371.9 m/s) than it was for the sealed-pore conditions (1333.4 m/s). We also find that it radiates less than under sealed-pore conditions, as α_{pS} is much lower than in Fig. 2 (right). For lower frequencies on the other hand, the attenuation δ_{pS} is higher than under sealed-pore conditions because of the free movement of the fluid in and out of the pores which causes high viscous interaction.

In Fig. 6 the real and imaginary parts of the corresponding (γ_j)'s are plotted. The radiation is less (0.040) indeed than under sealed-pore conditions (0.171, see Fig. 4, left). We find that also under open-pore conditions, γ_{p2} passes from the second to the first quadrant, and thus ceases to radiate, as we go down in frequency. All other (γ_j)'s are restricted to the fourth quadrant and have distinct minima.

The pseudo-Rayleigh (pR) wave is under frictionless conditions faster than the fluid mode and the bulk slow wave (see Fig. 5, left). Remarkably, as we go down in frequency, the velocity of the pseudo-Rayleigh wave becomes higher than that of the bulk shear wave, which would have been impossible for a frictionless medium (real-valued wave numbers $k_j=\omega/V_j$). Yet at lower frequencies when the medium becomes lossy $[\mathrm{Im}(k_j)<0]$, the condition $\gamma_j\equiv(k_x^2-k_j^2)^{1/2}$ does allow such outcome. The attenuation δ_{pR} is well predicted by the frictionless (f^1) conditions over the entire frequency regime. This is an indication that radiation remains important even at lower frequencies. To investigate this, we plotted in Figs. 7 the real and imaginary parts of the corresponding (γ_j)'s. We find that γ_f is always in the radiative second quadrant in an Argand diagram, whereas γ_{p2} passes from the second to the first quadrant, and thus ceases to radiate as we go down in frequency. The other two modes dwell in the nonradiative fourth quadrant. This means that for high frequencies the pseudo-Rayleigh wave is radiating both into the slow wave and the fluid wave (and not only into the bulk slow wave as for the pseudo-Stoneley wave), and for lower frequencies it is still radiating into the fluid mode (in contrast with the pseudo-Stoneley wave that does not radiate any more). As a consequence, not only the radiation damping α_{pR} is higher than α_{pS} of the pseudo-Stoneley wave (see Fig. 5, right), but also the total damping δ_{pR} is higher than δ_{pS} over the entire frequency regime.

VII. CONCLUSIONS

The true surface mode, the pseudo-Stoneley mode, and the pseudo-Rayleigh mode are crucially dependent on frequency. For the sealed-pore conditions, the true surface mode is always slower than all the bulk modes and it is damped by the viscous fluid–solid interaction just like all the bulk modes. The pseudo-Stoneley wave radiates into the slow wave at high frequencies and it ceases to radiate as frequency goes down. The same pseudo-Stoneley wave is also found for the open-pore conditions, where it radiates less than for the sealed-pore surface conditions. When it ceases to radiate for lower frequencies, however, its damping is higher than in the sealed-pore case because of the free fluid flow in and out of the pores. The pseudo-Rayleigh wave radiates both into the slow wave and the fluid bulk mode at high frequencies. When frequency goes down, radiation into the slow wave ceases, radiation into the fluid mode remains. This means that this surface mode is always damped by a combination of radiation and viscous fluid–solid interaction and thus displays f^1 dependence over the entire frequency domain. For lower frequencies, the velocity of the pseudo-Rayleigh wave may even become higher than the shear wave velocity, which would be impossible for frictionless media.

ACKNOWLEDGMENTS

This work was supported by INTAS (project 2000-50). The first two authors were also supported by the President of the Russian Federation for State Support of Leading Scientific Schools (Grant No. NSh.2059.2003.1).

APPENDIX: ELEMENTS OF $\underline{\underline{\mathbf{M}}}_o$ AND $\underline{\underline{\mathbf{M}}}_s$

Defining $\phi_j=1-\phi+\phi\beta_j$ for $j=p1$, $p2$, sh, the matrix $\underline{\underline{\mathbf{M}}}_o$ contains the following elements:

$$\underline{\underline{\mathbf{M}}}_o(1,1)=2N\gamma_{p1}^2-(A'+Q'\beta_{p1})k_{p1}^2,$$

$$\underline{\underline{\mathbf{M}}}_o(1,2)=2N\gamma_{p2}^2-(A'+Q'\beta_{p2})k_{p2}^2,$$

$$\underline{\underline{\mathbf{M}}}_o(1,3)=-2Nk_x\gamma_{\mathrm{sh}},$$

$$
\begin{aligned}
&\underline{\underline{\mathbf{M}}}_o(1,4)=0,\\
&\underline{\underline{\mathbf{M}}}_o(2,1)=2k_x\gamma_{p1},\\
&\underline{\underline{\mathbf{M}}}_o(2,2)=2k_x\gamma_{p2},\\
&\underline{\underline{\mathbf{M}}}_o(2,3)=(k_{\mathrm{sh}}^2-2k_x^2),\\
&\underline{\underline{\mathbf{M}}}_o(2,4)=0,\\
&\underline{\underline{\mathbf{M}}}_o(3,1)=(Q+R\beta_{p1})k_{p1}^2/\phi,\\
&\underline{\underline{\mathbf{M}}}_o(3,2)=(Q+R\beta_{p2})k_{p2}^2/\phi,\\
&\underline{\underline{\mathbf{M}}}_o(3,3)=0,\\
&\underline{\underline{\mathbf{M}}}_o(3,4)=-\omega^2\rho_f,\\
&\underline{\underline{\mathbf{M}}}_o(4,1)=\phi_{p1}\gamma_{p1},\\
&\underline{\underline{\mathbf{M}}}_o(4,2)=\phi_{p2}\gamma_{p2},\\
&\underline{\underline{\mathbf{M}}}_o(4,3)=-\phi_{\mathrm{sh}}k_x,\\
&\underline{\underline{\mathbf{M}}}_o(4,4)=\gamma_f.
\end{aligned}
\tag{A1}
$$

The matrix $\underline{\underline{\mathbf{M}}}_s$ contains the following elements:

$$
\begin{aligned}
&\underline{\underline{\mathbf{M}}}_s(1,1)=-2N\gamma_{p1}^2+[A+Q+(Q+R)\beta_{p1}]k_{p1}^2,\\
&\underline{\underline{\mathbf{M}}}_s(1,2)=-2N\gamma_{p2}^2+[A+Q+(Q+R)\beta_{p2}]k_{p2}^2,\\
&\underline{\underline{\mathbf{M}}}_s(1,3)=2Nk_x\gamma_{\mathrm{sh}},\\
&\underline{\underline{\mathbf{M}}}_s(1,4)=-\omega^2\rho_f,\\
&\underline{\underline{\mathbf{M}}}_s(2,1)=2k_x\gamma_{p1},\\
&\underline{\underline{\mathbf{M}}}_s(2,2)=2k_x\gamma_{p2},\\
&\underline{\underline{\mathbf{M}}}_s(2,3)=(k_{\mathrm{sh}}^2-2k_x^2),\\
&\underline{\underline{\mathbf{M}}}_s(2,4)=0,\\
&\underline{\underline{\mathbf{M}}}_s(3,1)=\gamma_{p1},\\
&\underline{\underline{\mathbf{M}}}_s(3,2)=\gamma_{p2},\\
&\underline{\underline{\mathbf{M}}}_s(3,3)=-k_x,\\
&\underline{\underline{\mathbf{M}}}_s(3,4)=\gamma_f,\\
&\underline{\underline{\mathbf{M}}}_s(4,1)=\beta_{p1}\gamma_{p1},\\
&\underline{\underline{\mathbf{M}}}_s(4,2)=\beta_{p2}\gamma_{p2},\\
&\underline{\underline{\mathbf{M}}}_s(4,3)=-\beta_{\mathrm{sh}}k_x,\\
&\underline{\underline{\mathbf{M}}}_s(4,4)=\gamma_f.
\end{aligned}
\tag{A2}
$$

[1] F. L. Paillet and C. H. Cheng, *Acoustic Waves in Boreholes* (CRC Press, Boca Raton, FL, 1991).
[2] X. M. Tang and C. H. Cheng, "Fast inversion of formation permeability from Stoneley wave logs using a simplified Biot–Rosenbaum model," Geophysics **61**, 639–645 (1996).
[3] J. H. Rosenbaum, "Synthetic microseismograms: Logging in porous formations," Geophysics **39**, 14–32 (1974).
[4] M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated porous solid," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[5] D. P. Schmitt, M. Bouchon, and G. Bonnet, "Full-wave synthetic acoustic logs in radially semiinfinite saturated porous media," Geophysics **53**, 807–823 (1988).
[6] H. L. Liu, "Borehole modes in a cylindrical fluid-saturated permeable medium," J. Acoust. Soc. Am. **84**, 424–431 (1988).
[7] K. W. Winkler, H. S. Liu, and D. L. Johnson, "Permeability and borehole Stoneley waves: Comparison between experiment and theory," J. Acoust. Soc. Am. **54**, 66–75 (1989).
[8] C. J. Wisse, D. M. J. Smeulders, M. E. H. van Dongen, and G. Chao, "Guided wave modes in porous cylinders: Experimental results," J. Acoust. Soc. Am. **112**, 890–895 (2002).
[9] H. Deresiewicz and R. Skalak, "On uniqueness in dynamic poroelasticity," Bull. Seismol. Soc. Am. **53**, 783–788 (1963).
[10] H. Deresiewicz, "The effect of boundaries on wave propagation in a liquid-filled porous solid: IV. Surface waves in a half-space," Bull. Seismol. Soc. Am. **52**, 627–638 (1962).
[11] 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).
[12] P. B. Nagy, "Observation of a new surface mode on a fluid-saturated permeable solid," Appl. Phys. Lett. **60**, 2735–2737 (1992).
[13] L. Adler and P. B. Nagy, "Measurements of acoustic surface waves on fluid-filled porous rocks," J. Geophys. Res., [Oceans] **99**, 17863–17869 (1994).
[14] J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier, New York, 1975).
[15] E. Strick, "Propagation of elastic wave motion from an impulsive source along a fluid/solid interface. Parts II and III," Philos. Trans. R. Soc. London, Ser. A **251**, 465–523 (1959).
[16] R. A. Phinney, "Propagation of leaking interface waves," Bull. Seismol. Soc. Am. **51**, 527–555 (1961).
[17] J. H. Ansell, "The roots of the Stoneley wave equation for solid–liquid interfaces," Pure Appl. Geophys. **94**, 172–188 (1972).
[18] J. H. M. T. van der Hijden, "Quantitative analysis of the pseudo-Rayleigh phenomenon," J. Acoust. Soc. Am. **75**, 1041–1047 (1984).
[19] W. L. Roever and T. F. Vining, "Propagation of elastic wave motion from an impulsive source along a fluid/solid interface. I. Experimental Pressure Response," Philos. Trans. R. Soc. London, Ser. A **251**, 455–465 (1959).
[20] J. G. J. Scholte, "The range of existence of Rayleigh and Stoneley waves," Mon. Not. R. Astron. Soc. **5**, 120–126 (1947).
[21] J. Miklowitz, *The Theory of Elastic Waves and Waveguides* (North-Holland, Amsterdam, 1978).
[22] A. T. de Hoop and J. H. M. T. van der Hijden, "Generation of acoustic waves by an impulsive point source in a fluid/solid configuration with a plane boundary," J. Acoust. Soc. Am. **75**, 1709–1715 (1984).
[23] 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).
[24] L. Kelders, W. Lauriks, and J. F. Allard, "Surface waves above thin porous layers saturated by air at ultrasonic frequencies," J. Acoust. Soc. Am. **104**, 882–889 (1998).
[25] 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).
[26] J. F. Allard, M. Henry, C. Glorieux, S. Petillon, and W. Lauriks, "Laser-induced surface modes at an air-porous medium interface," J. Appl. Phys. **93**, 1298–1304 (2003).
[27] M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).
[28] J. G. Berryman, "Elastic wave propagation in fluid-saturated porous media," J. Acoust. Soc. Am. **69**, 416–424 (1981).

Simplifications for the calculation of surface wave pulses generated by laser-irradiation

J. D. Achenbach[a)]
Northwestern University, McCormick School of Engineering and Applied Science, Center for Quality Engineering and Failure Prevention, 2137 N. Sheridan Rd., Catalysis Bldg., Rm. 327, Evanston, Illinois 60208-3020

(Received 27 January 2004; revised 16 June 2004; accepted 17 June 2004)

In this paper a simplified approach is presented to determine the ultrasonic surface wave motion generated by laser irradiation of a linearly elastic solid. It combines three novel ideas. The first is a way to directly determine the mechanical loading equivalent to laser irradiation by considering local thermoelastic effects in the area that is heated by the laser. The second idea is the use of a new formulation for surface waves which considers a surface wave as a carrier wave that satisfies Helmholtz' equation in surface coordinates and that supports depth-dependent motions that are the same for any kind of carrier wave. The third idea is to directly determine the surface wave motion generated by the equivalent mechanical loading by applying the elastodynamic reciprocity relation for the generated surface wave motion and an appropriately selected ''virtual'' wave motion. The analysis based on the reciprocity relations is carried out for the time-harmonic case, but Fourier superposition is used to obtain the surface wave pulses corresponding to pulsed laser irradiation. The approach based on these ideas is taken through several steps for an isotropic and a transversely isotropic solid whose mechanical properties may depend on the distance from the surface. For a homogeneous isotropic solid the details have been worked out for the surface wave pulses generated by the laser irradiation of arbitrary time dependence at a point, an infinitely long line, and a line of finite length. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779752]

PACS numbers: 43.35.Pt [JJM] Pages: 1481–1487

I. INTRODUCTION

The generation of ultrasound by heat deposition due to laser irradiation has several advantages over conventional generation by piezoelectric transducers. Some of these advantages are as follows: a small area of noncontact generation, the applicability of fiber optics, narrow-band and broadband generation, easy scanning and ability to generate ultrasound on curved and rough surfaces and at hard-to-access locations.

The analysis of ultrasound generation by laser irradiation is a problem of thermo-elastodynamics. It is intuitively quite obvious that, by neglecting thermal diffusion, a point source of heat in the interior of a linearly elastic body is mechanically equivalent to a center of dilatation. Since a center of dilatation can be represented by three mutually orthogonal force-dipoles of equal magnitudes, it is tempting to assume that, when a point source of heat is applied at the free surface of a solid body, the dipole normal to the free surface is simply zero, while in the orthogonal plane, i.e., in the free surface, the two remaining dipoles have the same value as for a point source in the body's interior. It was, however, shown that this assumption yields the wrong result for such a surface center of expansion.

A mathematical analysis of ultrasound generated by point source heating of the surface of an elastic half-space was given by Rose,[1] who presented a detailed discussion based on the use of the Hankel transform technique. The analogous two-dimensional case of line-source excitation has been analyzed by Bernstein and Spicer.[2] The surface center of the expansion model predicts the major features of the waveform and agrees well with experiments for highly focused (small spot size) and short (nanosecond duration) laser pulses; see Aussel *et al.*[3] and Hutchins.[4]

In this paper we use simple thermoelastic arguments to determine the mechanical loading equivalent to point heating by laser irradiation. The geometry is shown in Fig. 1. The approach presented in this paper has the advantage that the surface center of expansion can be determined for a general class of elastic materials without the necessity of using integral transform techniques. As an example, we consider laser irradiation of a transversely isotropic material whose axis of symmetry is perpendicular to the free surface of the body. Moreover, the material may be functionally graded, i.e., its mechanical and thermal properties depend on z, the distance from the free surface.

General calculations by Arias and Achenbach[5] of ultrasound generated by laser irradiation have shown that by far the most significant part of the generated surface motion is a pulse propagating with the velocity of Rayleigh surface waves. This result begs the question of whether the surface wave pulse can be obtained directly, without consideration of the complete wave field. In the conventional approach for point-irradiation, a Hankel transform is usually applied with respect to the radial coordinate in the free surface, and the resulting system of equations is solved, to yield an inverse transform in the form of an infinite integral. The inverse integral cannot necessarily be evaluated in closed form, but

a)Electronic mail: achenbach@northwestern.edu

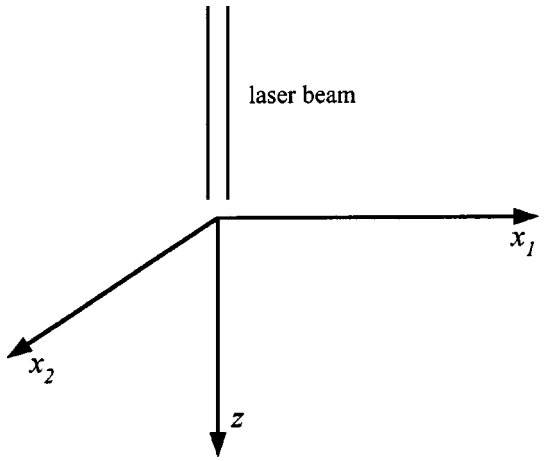


FIG. 1. Geometry of irradiated half-space.

expressions for the surface waves can be obtained as contributions from poles due to the zeros in the integrand's denominator. Clearly this is still an arduous procedure, with some treacherous possibilities for error.

In this paper, the surface pulse is directly determined. First, a novel analysis of surface waves in a linearly elastic solid, based on the use of a single potential for a carrier wave on the free surface, is employed. Next, the actual surface wave generated by laser irradiation is determined for time-harmonic excitation by use of the reciprocity theorem, whereby the two reciprocal solutions are the actually generated surface waves and a ''virtual'' surface wave. The technique has general applicability for anisotropic and inhomogeneous materials, as long as the general form of the surface wave pulse is known. Here we will explore the cases of an isotropic, a transversely isotropic and a functionally graded material.

In the last part of the paper, Fourier superposition of ultrasonic surface wave harmonics for point irradiation is used to obtain the surface wave pulse for irradiation along an infinitely long line, and for irradiation along a line of finite length.

II. EQUIVALENT MECHANICAL LOADING

In terms of a Cartesian coordinate system, (x_1,x_2,z), the stress–strain relations for a transversely isotropic solid may be written as

$$\tau_{11}=C_{11}\frac{\partial u_1}{\partial x_1}+C_{12}\frac{\partial u_2}{\partial x_2}+C_{13}\frac{\partial w}{\partial z}, \tag{1}$$

$$\tau_{22}=C_{12}\frac{\partial u_1}{\partial x_1}+C_{11}\frac{\partial u_2}{\partial x_2}+C_{13}\frac{\partial w}{\partial z}, \tag{2}$$

$$\tau_{zz}=C_{13}\frac{\partial u_1}{\partial x_1}+C_{13}\frac{\partial u_2}{\partial x_2}+C_{33}\frac{\partial w}{\partial z}, \tag{3}$$

$$\tau_{2z}=C_{44}\left(\frac{\partial u_2}{\partial z}+\frac{\partial w}{\partial x_2}\right), \tag{4}$$

$$\tau_{1z}=C_{44}\left(\frac{\partial u_1}{\partial z}+\frac{\partial w}{\partial x_1}\right), \tag{5}$$

$$\tau_{12}=C_{66}\left(\frac{\partial u_1}{\partial x_2}+\frac{\partial u_2}{\partial x_1}\right). \tag{6}$$

Here the x_1x_2-plane is the plane of transverse isotropy. The solid may be functionally graded, which implies that the elastic constants may depend on z.

First let us consider the equation that governs the temperature distribution due to heat deposition near the surface of the body. Taking into account that the thermal conductivity, K, may be transversely isotropic, and may also depend on z, we have

$$K_1(z)\left(\frac{\partial^2 T}{\partial x_1^2}+\frac{\partial^2 T}{\partial x_2^2}\right)+\frac{\partial}{\partial z}\left[K_2(z)\frac{\partial T}{\partial z}\right]=\rho(z)c_V(z)\frac{\partial T}{\partial t}-q, \tag{7}$$

where $\rho(z)$ and $c_V(z)$ are the mass density and the specific heat at constant deformation. In Eq. (7), q represents the heat deposition per unit volume and per unit time. The rate of heat production by mechanical deformation has been neglected.

A suitable expression for q is often taken as

$$q=E(1-R_i)h(z)f(r)g(t), \tag{8}$$

where

$$h(z)=\gamma e^{-\gamma z}, \tag{9}$$

$$f(r)=\frac{2}{\pi}\frac{1}{R_G^2}e^{-2r^2/R_G^2}, \tag{10}$$

and

$$g(t)=\frac{8t^3}{\nu^4}e^{-2t^2/\nu^2}. \tag{11}$$

In these expressions, E is the energy of the laser pulse, R_i is the surface reflectivity, R_G is the Gaussian beam radius, ν is a laser pulse rise time parameter and γ is the extinction coefficient. For both the temporal and the spatial profile, the functional dependence of q has been constructed so that in the limits $\nu\to 0$, $R_G\to 0$, and $\gamma\to\infty$, an equivalent concentrated point source of the form

$$q=E(1-R_i)\,\delta(z)\,\frac{\delta(r)}{2\pi r}\,\delta(t) \tag{12}$$

is obtained.

In the next step it is assumed that the energy deposition is so fast that the generated wave motion is not affected by heat conduction effects. Equation (7) then reduces to

$$\frac{dT}{dt}=\frac{E}{\rho(z)c_V(z)}(1-R_i)h(z)f(r)g(t), \tag{13}$$

and the increment of temperature, ΔT, becomes

$$\Delta T=\frac{E}{\rho(z)c_V(z)}(1-R_i)h(z)f(r)\int_0^t g(s)ds. \tag{14}$$

To proceed further, we do a thought experiment on the mechanism of ultrasound generation by laser irradiation. Suppose the heat energy of the laser beam is deposited in a surface element of the illuminated body, and let us hypothetically remove this element from the body. The removed element has radius r_0 and thickness d. The element, which is

now free on all sides, is free to expand due to the temperature rise ΔT. The strains due to thermal expansion are

$$\varepsilon_r = \varepsilon_\theta = \alpha \overline{\Delta T}, \tag{15}$$

where $\overline{\Delta T}$ is the averaged temperature increment,

$$\overline{\Delta T} = \frac{E}{\rho^0 c_V^0}(1-R_i)\frac{H(d,r_0)}{\pi r_0^2 d}\int_0^t g(s)ds, \tag{16}$$

and

$$H(d,r_0) = 2\pi \int_0^d h(z)dz \int_0^{r_0} f(r) r dr. \tag{17}$$

It has been assumed that d is small enough so that ρ^0 and c_V^0, the values of $\rho(z)$ and $c_V(z)$ at $z=0$, can be used in Eq. (16).

The removal of the element from the body has created a cavity. To put the expanded element back into the cavity, it has to be compressed isothermally to its original radius. The compressive radial stresses are provided by the material surrounding the element, and their reactions on the walls of the cavity generate the ultrasound in the irradiated body.

The compression of the element must produce strains equal and opposite to the ones given by Eq. (15), while the normal stress must remain zero, i.e., $\tau_z = 0$. We use the isothermal stress–strain relations for axial symmetry in polar coordinates,

$$\tau_r = C_{11}\varepsilon_r + C_{12}\varepsilon_\theta + C_{13}\varepsilon_z, \tag{18}$$

$$\tau_\theta = C_{12}\varepsilon_r + C_{11}\varepsilon_\theta + C_{13}\varepsilon_z, \tag{19}$$

$$\tau_z = C_{13}\varepsilon_r + C_{13}\varepsilon_\theta + C_{33}\varepsilon_z. \tag{20}$$

From $\varepsilon_r = \varepsilon_\theta = -\alpha \overline{\nabla T}$ and Eq. (20), the condition $\tau_z = 0$ yields

$$\varepsilon_z = \frac{2C_{13}\alpha \overline{\Delta T}}{C_{33}}. \tag{21}$$

The corresponding stresses τ_r and τ_θ follow from (18), (19), and (21) as

$$\tau_r = \tau_\theta = \frac{2C_{13}^0 C_{13}^0 - (C_{11}^0 + C_{12}^0)C_{33}^0}{C_{33}^0}\alpha \overline{\Delta T}. \tag{22}$$

For a very small value of d, the values of the elastic constants at $z=0$ may be used. The radial tractions on the surface of the cavity (d times the radial stress), which are the reactions to the compression of the disk, are equivalent to radial tractions on the surface $z=0$, defined by

$$F_r = \frac{D}{\pi r_0^2}\int_0^t g(s)ds, \tag{23}$$

acting along the circle of radius r_0 on the surface of the body; see Fig. 2. Here D follows from Eq. (22) as

$$D = -\frac{2C_{13}^0 C_{13}^0 - (C_{11}^0 + C_{12}^0)C_{13}^0}{C_{33}^0}\frac{\alpha E}{\rho^0 c_V^0}(1-R_i)H(d,r_0). \tag{24}$$

It has been shown by Arias and Achenbach[5] that, in the limit $r_0 \to 0$ the circumferential distribution of radial surface tractions is equivalent to a set of orthogonal dipoles of magnitude D acting on the surface of the half space.

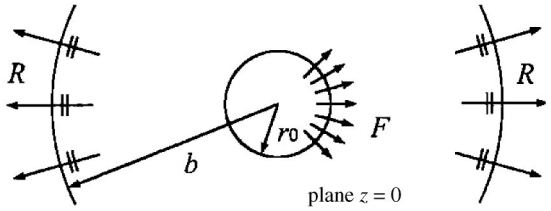


FIG. 2. Equivalent axially-symmetric surface tractions for point-focus laser-irradiation of a half-space.

For a homogenous isotropic material we have

$$C_{44} = C_{66} = \tfrac{1}{2}(C_{11} - C_{12}), \tag{25}$$

and, in terms of the Lamé elastic constants, we find

$$C_{11} = C_{33} = \lambda + 2\mu, \quad C_{12} = C_{13} = \lambda, \quad C_{44} = C_{66} = \mu. \tag{26}$$

For that case D reduces to

$$D = \frac{2\mu}{\lambda + 2\mu}(3\lambda + 2\mu)\frac{\alpha E}{\rho c_V}(1-R_i)H(d,r_0). \tag{27}$$

It is noteworthy that D, as given above, differs from the one for an unbounded solid by the factor $2\mu/(\lambda+2\mu)$. Equation (27) agrees with the corresponding result that can be derived from the paper by Rose.[1]

For future reference we also write Eq. (23) in terms of an exponential Fourier integral,

$$F_r = \frac{D}{\pi r_0^2}\int_{-\infty}^{\infty}\frac{\hat{g}(\omega)}{(-i\omega)}e^{-i\omega t}d\omega, \tag{28}$$

where

$$\hat{g}(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(s)e^{i\omega s}ds. \tag{29}$$

III. SURFACE WAVES

In this section we establish an approach to consider the propagation of surface waves along the free surface of a transversely isotropic functionally graded solid. The axis of symmetry is normal to the free surface and the mechanical properties depend on the distance from that surface. Following an earlier approach by the author,[6] we consider a surface wave as a carrier wave that propagates over the free surface as it would over a membrane, while it carries with it a system of depth dependent superimposed motions. For a time-harmonic wave motion the carrier wave satisfies a reduced wave equation in the coordinates that span the free surface. The functions representing the superimposed motions are solutions of ordinary differential equations in the normal coordinate z. This formulation unequivocally shows that the dependence on the normal coordinate is the same for any time-harmonic carrier wave that satisfies Helmholtz' equation.

For the formulation described above, we seek solutions for the displacement components in the general forms

$$u_\alpha(x_1,x_2,z,t) = \frac{1}{k}V(z)\frac{\partial \varphi}{\partial x_\alpha}(x_1,x_2)e^{-i\omega t}, \tag{30}$$

$$u_z(x_1,x_2,z,t)=W(z)\varphi(x_1,x_2)e^{-i\omega t}, \tag{31}$$

where x_1 and x_2 are Cartesian coordinates in the free surface, $\alpha=1$, 2, and k is a wave-number-like quantity,

$$k=\omega/c. \tag{32}$$

The dimensionless function $\varphi(x_1,x_2)$ is the solution of Helmholtz' equation,

$$\frac{\partial^2\varphi}{\partial x_1^2}+\frac{\partial^2\varphi}{\partial x_2^2}=k^2\varphi. \tag{33}$$

In what follows the time-harmonic time factor is omitted.

Substitution of the displacement expressions in the stress–strain relations given by Eqs. (1)–(6), and subsequent substitution in the stress-equations of motion yields, after some rearranging,

$$\frac{\partial}{\partial x_\alpha}\left\{\left[-kC_{11}V(z)+(C_{13}+C_{44})\frac{dW}{dz}+\frac{1}{k}C_{44}\frac{d^2V}{dz^2}+\frac{\rho\omega^2}{k}V(z)+\frac{dC_{44}}{dz}\left(\frac{1}{k}\frac{dV}{dz}+W\right)\right]\varphi\right\}=0, \tag{34}$$

where $\alpha=1$ and $\alpha=2$. Clearly this equation is satisfied for

$$-kC_{11}V(z)+(C_{13}+C_{44})\frac{dW}{dz}+\frac{1}{k}C_{44}\frac{d^2V}{dz^2}+\frac{\rho\omega^2}{k}V(z)+\frac{dC_{44}}{dz}\left[\frac{1}{k}\frac{dV}{dz}+W(z)\right]=0. \tag{35}$$

The third equation yields, in a similar way,

$$-k^2\left[\frac{1}{k}(C_{44}+C_{13})\frac{dV}{dz}+C_{44}W(z)\right]+C_{33}\frac{d^2W}{dz^2}+\rho\omega^2W(z)-k\frac{dC_{13}}{dz}V(z)+\frac{dC_{33}}{dz}\frac{dW}{dz}=0. \tag{36}$$

Equations (35) and (36) define a system of coupled ordinary differential equations for $V(z)$ and $W(z)$, with coefficients that depend on z. These equations are very difficult if not impossible to solve analytically. However, for the very special case that all elastic constants, C_{ij}, as well as the mass density, ρ, are taken as having the same exponential dependence on z, say

$$C_{ij}(z)=C_{ij}^0e^{\beta z} \quad \text{and} \quad \rho=\rho^0e^{\beta z}, \tag{37}$$

a solution, in principle, can be obtained. For a solid whose mechanical constants increase with depth, this special case may give an acceptable approximation for small z, i.e., in the region where we consider the propagation of surface waves. By the substitution of Eqs. (37) into Eqs. (35) and (36), these equations reduce to a coupled system of ordinary differential equations with constants coefficients,

$$-kC_{11}^0V(z)+\frac{\rho^0\omega^2}{k}V(z)+\frac{\beta}{k}C_{44}^0\frac{dV}{dz}+\frac{1}{k}C_{44}^0\frac{d^2V}{dz^2}+\beta C_{44}^0W(z)+(C_{13}^0+C_{44}^0)\frac{dW}{dz}=0, \tag{38}$$

$$-k\beta C_{13}^0V(z)-k(C_{13}^0+C_{44}^0)\frac{dV}{dz}-k^2C_{44}^0W(z)+\rho^0\omega^2W(z)+\beta C_{33}^0\frac{dW}{dz}+C_{33}^0\frac{d^2W}{dz^2}=0. \tag{39}$$

For surface waves we seek solutions of Eqs. (38) and (39) of the general forms

$$V(z)=\overline{V}e^{-pz} \quad \text{and} \quad W(z)=\overline{W}e^{-pz}. \tag{40}$$

Substitution in Eqs. (38) and (39) yields

$$\left[-kC_{11}^0+\frac{\rho^0\omega^2}{k}-\beta\frac{p}{k}C_{44}^0+\frac{p^2}{k}C_{44}^0\right]\overline{V}+[\beta C_{44}^0-p(C_{13}^0+C_{44}^0)]\overline{W}=0, \tag{41}$$

$$-[\beta kC_{13}^0-kp(C_{13}^0+C_{44}^0)]\overline{V}-[k^2C_{44}^0-\rho^0\omega^2+\beta pC_{33}^0-p^2C_{33}^0]\overline{W}=0. \tag{42}$$

This is a system of homogeneous algebraic equations for $\overline{V}$ and $\overline{W}$. The required vanishing of the determinant yields a bi-quadratic equation for p, which can, in principle, be solved. *Here we will, however, not further explore this rather special case*. To illustrate the general procedure we will turn our attention to the case of a *homogeneous isotropic solid*, by setting $\beta\equiv0$, and by introducing the constants given by Eq. (26). After a little manipulation, Eqs. (41) and (42) then reduce to

$$\left(-\frac{c_L^2}{c_T^2}+\frac{\omega^2}{k^2c_T^2}+\frac{p^2}{k^2}\right)\overline{V}-\frac{p}{k}\frac{c_L^2-c_T^2}{c_T^2}\overline{W}=0, \tag{43}$$

$$-\frac{p}{k}\frac{c_L^2-c_T^2}{c_L^2}\overline{V}-\left(\frac{c_T^2}{c_L^2}-\frac{\omega^2}{k^2c_L^2}-\frac{p^2}{k^2}\right)\overline{W}=0, \tag{44}$$

where

$$c_L^2=\frac{\lambda+2\mu}{\rho}, \quad \text{and} \quad c_T^2=\frac{\mu}{\rho}. \tag{45}$$

The required vanishing of the determinant yields, for this case,

$$\left(\frac{p}{k}\right)^4+\left(\frac{\omega^2}{k^2c_L^2}+\frac{\omega^2}{k^2c_T^2}-2\right)\left(\frac{p}{k}\right)^2+\left(\frac{\omega^2}{k^2c_L^2}-1\right)\left(\frac{\omega^2}{k^2c_T^2}-1\right)=0. \tag{46}$$

The two solutions of the above equation are

$$\left(\frac{p}{k}\right)^2=1-\frac{\omega^2}{k^2c_L^2} \quad \text{and} \quad \left(\frac{q}{k}\right)^2=1-\frac{\omega^2}{k^2c_T^2}. \tag{47}$$

The corresponding solutions of (43) and (44) can be written as

$$V(z)=Ae^{-pz}+Be^{-qz}, \tag{48}$$

$$W(z)=-\frac{p}{k}Ae^{-pz}-\frac{k}{q}Be^{-qz}. \tag{49}$$

We still have to satisfy the boundary conditions at $z=0$:

$$\tau_{z\beta}(x_1,x_2,0)=0 \quad \text{and} \quad \tau_{zz}(x_1,x_2,0)=0, \tag{50}$$

where $\beta=1, 2$. These equations yield a system of homogeneous algebraic equations for the constants A and B. As shown in Ref. 7, the condition that the determinant of the coefficients must vanish yields the well known equation for the phase velocity, $c=c_R$, of Rayleigh surface waves. In a slight rewrite of Eqs. (30) and (31), the final results for the isotropic case are

$$u_\alpha = A\frac{1}{k_R}V^R(z)\frac{\partial\varphi}{\partial x_\alpha}(x_1,x_2), \tag{51}$$

$$u_z = AW^R(z)\varphi(x_1,x_2), \tag{52}$$

where

$$k_R=\omega/c_R. \tag{53}$$

Also,

$$V^R(z)=d_1e^{-pz}+d_2e^{-qz}, \tag{54}$$

$$W^R(z)=d_3e^{-pz}-e^{-qz}, \tag{55}$$

where d_1, d_2, and d_3 are given in Ref. 7, p. 121, but by using Eq. (53) they are rewritten here as the following constants:

$$d_1=-\frac{1}{2}\frac{k_R^2+q^2}{k_Rp}=-\frac{1}{2}\frac{2-c_R^2/c_T^2}{(1-c_R^2/c_L^2)^{1/2}}, \tag{56}$$

$$d_2=\frac{q}{k_R}=(1-c_R^2/c_T^2)^{1/2}, \tag{57}$$

$$d_3=\frac{1}{2}\frac{k_R^2+q^2}{k_R^2}=\frac{1}{2}(2-c_R^2/c_T^2). \tag{58}$$

The analogous expressions for the transversely isotropic solid can be found in Ref. 8.

It is to be expected that just as for the case of a homogeneous solid, the bi-quadratic equation will yield two suitable expressions for p, i.e., expressions such that $\mathrm{Re}(p)\geqslant 0$. Then, once $V(z)$ and $W(z)$ have been obtained in the forms (40), the corresponding stresses τ_{zz}, τ_{z1}, and τ_{z2} can also be determined. Next, the boundary conditions at $z=0$ of vanishing surface stresses will again yield two homogeneous algebraic equations from which the equation for the surface wave velocity can be obtained.

As an example we will consider the case of axial symmetry. In polar coordinates the function $\varphi(r)$ satisfies the equation

$$\frac{d^2\varphi}{dr^2}+\frac{1}{r}\frac{d\varphi}{dr}+k^2\varphi=0. \tag{59}$$

For an outgoing wave compatible with $\exp(-i\omega t)$, the solution of Eq. (59) is the Hankel function:[8]

$$\varphi(r)=H_0^{(1)}(kr). \tag{60}$$

Equations (30) and (31) yield

$$u_r=-AV^R(z)H_1^{(1)}(kr), \tag{61}$$

$$u_z=AW^R(z)H_0^{(1)}(kr). \tag{62}$$

The relevant corresponding stresses are of the general forms

$$\tau_{rr}=A\left[T_{rr}(z)H_0^{(1)}(kr)+\overline{T}_{rr}(z)\frac{1}{kr}H_1^{(1)}(kr)\right], \tag{63}$$

$$\tau_{rz}=-AT_{rz}(z)H_1^{(1)}(kr). \tag{64}$$

For the isotropic case, expressions for $T_{rr}(z), \overline{T}_{rr}(z)$ and $T_{rz}(z)$ can be found in Achenbach,[9] Eqs. (56)–(57).

IV. SURFACE WAVES GENERATED BY LASER IRRADIATION

As discussed in Sec. II, the wave motion generated by laser irradiation of a surface can be approximated by the waves generated by an equivalent mechanical loading. The analysis of the generated wave motion then becomes a purely isothermal problem of elastodynamics. The application of a surface disturbance generates a set of body waves and a surface wave. Due to the different rates of geometrical decay, only the surface wave needs to be considered sufficiently far from the area of application of a surface disturbance.

To directly calculate the surface wave, we use the elastodynamic reciprocity theorem. For a region V with boundary S, and for two time-harmonic states defined by superscripts A and B, this theorem may be written as

$$\int_V (f_j^Au_j^B-f_j^Bu_j^A)dV=\int_S(\tau_{ij}^Bu_j^A-\tau_{ij}^Au_j^B)n_i\,dS, \tag{65}$$

where n is the outward normal.

From Eq. (23) we find that the relevant harmonic component of the traction on the surface may be written as

$$t_r=-\frac{D}{\pi r_0^2}\delta(r-r_0). \tag{66}$$

Following earlier work by the author,[9] the surface wave that is generated by this condition can be obtained by an application of the reciprocity theorem over a cylindrical body, $z\geqslant 0, 0\leqslant r\leqslant b, 0\leqslant\theta\leqslant 2\pi$, and bounded by $z=0$ and $r=b$. For State A, we select surface waves that propagate away from the circular area of application of the axially symmetric surface tractions. These surface waves may be written in the general form given by Eqs. (61) and (62), where $A(\omega)$ is an unknown amplitude factor. The relevant corresponding stresses are given by Eqs. (63) and (64) where, for the isotropic case $T_{rr}(z)$, $\overline{T}_{rr}(z)$, and $T_{rz}(z)$ can be found in Ref. 9, Eqs. (56)–(57).

For State B, the virtual wave, we select a combination of outgoing and incoming surface waves so that the displacements are bounded at $r=0$:

$$u_r^B=-\tfrac{1}{2}BV^R(z)[H_1^{(1)}(k_Rr)+H_1^{(2)}(k_Rr)], \tag{67}$$

$$u_z^B=\tfrac{1}{2}BW^R(z)[H_0^{(1)}(k_Rr)+H_0^{(2)}(k_Rr)]. \tag{68}$$

The corresponding stresses follow from Hooke's law; they are explicitly given elsewhere.[9] Since there are no internal body forces, the left hand side of Eq. (65) vanishes. We obtain

$$-2\frac{D}{r_0}BV^R(0)J_1(k_Rr_0)$$

$$=b\int_0^{2\pi}\int_0^{\infty}\{[u_r^A\tau_{rr}^B-u_r^B\tau_{rr}^A]+[u_z^A\tau_{rz}^B-u_z^B\tau_{rz}^A]\}d\theta dz. \tag{69}$$

It is very useful that the products of terms representing counter-propagating waves on the right-hand side of Eq. (69) cancel each other, and the remaining terms reduce to a very simple result:

$$-2\frac{D}{r_0}V^R(0)J_1(k_Rr_0)=4i\frac{\mu}{k_R}A(\omega)J, \tag{70}$$

where

$$J=\int_0^{\infty}[T_{rr}^R(z)V^R(z)-T_{rz}^R(z)W^R(z)]dz. \tag{71}$$

This integral has a constant value that has been calculated for an isotropic solid[7] (p. 140). It is noted that $A(\omega)$ as solved from Eq. (70) would contain the length parameter r_0, which can be taken as R_G, the Gaussian beam radius, i.e., a parameter defining the irradiated area. A still simpler result is obtained by considering the Bessel functions $J_1(k_Rr_0)$ at very small values of k_Rr_0. Then we have $(1/r_0)J_1(k_Rr_0)\sim k_R/2$, and Eq. (70) yields

$$A(\omega)=-\frac{i}{2}C(-i\omega)^2, \tag{72}$$

where $k_R=\omega/c_R$ has been used, and the constant C is

$$C=\frac{D}{2}\frac{1}{c_R^2}\frac{V^R(0)}{\mu J}. \tag{73}$$

In the frequency domain, the vertical displacement at $z=0$ follows from Eq. (62), after restoring the time factor, as

$$\hat{u}_z(r,\omega)=-\frac{i}{2}CW^R(0)(-i\omega)\hat{g}(\omega)H_0^{(1)}(k_Rr)e^{-i\omega t}. \tag{74}$$

This expression is the exponential Fourier transform of $u_z(r,t)$. Compatible with the definition of $\hat{g}(\omega)$ given by Eq. (29), we write

$$u_z(r,t)=-\frac{i}{2}CW^R(0)\int_{-\infty}^{\infty}(-i\omega)\hat{g}(\omega)H_0^{(1)}(\omega\bar{r})e^{-i\omega t}d\omega, \tag{75}$$

where $\bar{r}=r/c_R$. It is emphasized that C and $W_R(0)$ are constants that depend on the velocity of surface waves.

The integral in Eq. (75) can be simplified considerably. The final result has been derived elsewhere,[9]

$$u_z(r,t)=-\frac{1}{\pi}\frac{D}{2}\frac{1}{c_R^2}\frac{V^R(0)W^R(0)}{\mu J}I, \tag{76}$$

where

$$I=\int_0^{t-r/c_R}\frac{\dot{g}(s)ds}{[(t-s)^2-(r/c_R)^2]^{1/2}}. \tag{77}$$

Based on the work of Rose,[1] Berthelot[10] presented a further simplification of the point-focus pulse referred to as the half-order derivative of $g(t)$. This simplification can be obtained from Eq. (77) by noting that the term $[t-s+r/c_R]^{1/2}$ gets most of its contribution from the vicinity of the singularity of the integrand at $s=t-r/c_R$. Hence we can substitute this value for s in $[t-s+r/c_R]^{1/2}$. The result is given in Ref. 10.

Equation (76) can be interpreted as the field generated by a source of time dependence $\dot{g}(t)$ applied at the origin as a spatial delta function, $\delta(x_1)\delta(x_2)$. In the time-space domain the field at position (t,x_1,x_2) has a domain of dependence defined by the cone

$$(t-s)-[(\bar{x}_1-\bar{\zeta})^2+(\bar{x}_2-\bar{\xi})^2]^{1/2}\geqslant 0,\quad t\geqslant s\geqslant 0, \tag{78}$$

where $\bar{x}_1=x_1/c_R$ and $\bar{x}_2=x_2/c_R$. Now suppose a uniform distribution of sources is applied along the complete length of the x_2-axis. For the resulting two-dimensional problem, instead of Eq. (76), we then obtain an integration over the intersection of the cone with the tx_2-plane,

$$I_1=c_R\int_0^{t-\bar{x}_1}\dot{g}(s)ds$$

$$\times\int_{\bar{x}_2-[(t-s)^2-\bar{x}_1^2]^{1/2}}^{\bar{x}_2+[(t-s)^2-\bar{x}_1^2]^{1/2}}\frac{d\bar{\xi}}{[(t-s)^2-\bar{x}_1^2-(\bar{x}_2-\bar{\xi})^2]^{1/2}}. \tag{79}$$

The second integral in Eq. (79) can be evaluated to yield π. It follows that Eq. (76) becomes

$$u_z(r,t)=-\frac{\bar{D}}{2}\frac{1}{c_R}\frac{V^R(0)W^R(0)}{\mu J}g\left(t\pm\frac{x_1}{c_R}\right). \tag{80}$$

Here we have placed a bar over the D to indicate that the $\bar{D}$ should now be interpreted as the dipole per unit length.

A comparison of Eqs. (76) and (80) shows that the surface wave pulses for line-focus and point-focus laser-irradiation are quite different in shape. For line-focus irradiation the pulse shape is proportional to the laser pulse, while for the point-focus case an initial pulse is followed by a smaller pulse with a somewhat extended tail.

Irradiation of a line segment $-a\leqslant x_2\leqslant a$ is more interesting than either point-or infinite-line irradiation. An efficient way to obtain the ultrasonic field generated by illumination of a line segment is to use three building blocks: (1) "positive" irradiation over $-\infty<x_2<\infty$; (2) "negative" irradiation over $a\leqslant x_2<\infty$; and (3) "negative" irradiation over $-\infty<x_2\leqslant -a$. The superposition of the three results yields "positive" irradiation over the strip $-a\leqslant x_2\leqslant a$. The ultrasonic field for case (1) follows from Eqs. (76) and (79). For case (2) we will now evaluate the relevant superposition integral corresponding to Eq. (76) by integration over the appropriate area of dependence in the x_1,s-plane. We obtain for a point (x_1,x_2) where $x_2<a$,

$$I_2=c_R\int_0^{t-[\bar{x}_1^2+(\bar{a}-\bar{x}_2)^2]^{1/2}}\dot{g}(s)ds$$

$$\times\int_{\bar{a}}^{\bar{x}_2+[(t-s)^2-\bar{x}_1^2]^{1/2}}\frac{d\bar{\xi}}{[(t-s)^2-\bar{x}_1^2-(\bar{x}_2-\bar{\xi})^2]^{1/2}}, \tag{81}$$

where $\bar{a}=a/c_R$. The integral can be evaluated further to yield

$$I_2=\frac{\pi}{2}c_R g[t-\sqrt{\bar{x}_1^2+(\bar{a}-\bar{x}_2)^2}]+c_R\int_0^{t-[\bar{x}_1^2+(\bar{a}-\bar{x}_2)^2]^{1/2}}\dot{g}(s)\times\sin^{-1}\frac{\bar{x}_2-\bar{a}}{[(t-s)^2-\bar{x}_1^2]^{1/2}}\,ds. \tag{82}$$

Similarly for the third problem we get

$$I_3=\frac{\pi}{2}c_R g[t-\sqrt{\bar{x}_1^2+(\bar{a}+\bar{x}_2)^2}]-c_R\int_0^{t-[\bar{x}_1^2+(\bar{a}+\bar{x}_2)^2]^{1/2}}\dot{g}(s)\times\sin^{-1}\frac{\bar{x}_2+\bar{a}}{[(t-s)^2-\bar{x}_1^2]^{1/2}}\,ds. \tag{83}$$

The vertical displacement at $(x_1,x_2),|x_2|\leqslant a$ for line irradiation along the line segment $x_1=0,|x_2|\leqslant a$ may then be obtained by replacing I in Eq. (76) by $I_1-I_2-I_3$.

V. CONCLUDING COMMENTS

Our purpose in this paper has been to introduce a few ideas to simplify the calculation of the surface wave pulse that is generated by laser irradiation of a free surface. These simplifications provide an alternative method to the use of integral transform techniques. They offer a direct determination of the mechanical loading equivalent to laser irradiation, a more general formulation of free surface waves, and a direct way of determining the surface waves generated by the equivalent mechanical loading. The applicability of the approach was demonstrated for an isotropic and a transversely isotropic solid whose axis of symmetry is perpendicular to the free surface. For both cases the mechanical properties were allowed to depend on the distance from the free surface. Complete details were worked out for the case of a *homogeneous, isotropic elastic* solid, for which the surface wave pulses were determined for laser irradiation of arbitrary time dependence at a point, an infinitely long line and a line of finite length.

It should be quite apparent that at least parts of the approach can be extended to more general materials and configurations. Thus the direct determination of the equivalent mechanical loading can be generalized to solids of more general anisotropy. The formulation of surface waves as carrier waves of subsurface motions has already been extended to guided waves in a more general sense, such as Lamb waves for a layer,[6] and it can further be extended to guided waves in a multiple-layered solid, and to interface waves. Since the reciprocity relation is valid for inhomogeneous solids of general anisotropy, the guided waves generated by mechanical loading can be determined provided that the analytical form of the guided wave motion is known. For cases of continuous material inhomogeneity, such as functionally graded materials, it may be necessary to use numerical techniques to determine the depth dependence of the guided wave motion.

As a final note, and in response to comments by a reviewer, I will admit that the methods presented in this paper have not been justified by rigorous mathematical considerations. I would be delighted if a reader of this paper would provide a rigorous methodology to show that the approach is mathematically justifiable, or, alternatively, represents an approximation for which error estimates can be given. Until such a time, I will not totally discard the possibility, perish the thought, that there is no rigorous mathematical justification, even though for some specific cases it has been shown by direct comparison that the correct results are obtained.

[1] L. R. F. Rose, "Point-source representation for laser-generated ultrasound," J. Acoust. Soc. Am. **75**, 723–732 (1984).
[2] J. Bernstein and J. Spicer, "Line source representation for laser-generated ultrasound in aluminum," J. Acoust. Soc. Am. **107**, 1352–1357 (2000).
[3] J. Aussel, A. L. Brun, and J. Badoux, "Generating acoustic waves by laser: theoretical and experimental study of the emission source," Ultrasonics **25**, 245–255 (1988).
[4] D. Hutchins, "Ultrasonic generation by pulsed laser," *Physical Acoustics* (Academic, New York 1988), Vol. XVIII.
[5] I. Arias and J. D. Achenbach, "Thermoelastic generation of ultrasound by laser line-source irradiation," Int. J. Solids Struct. **40**, 6917–6835 (2003).
[6] J. D. Achenbach, "Lamb waves as thickness vibrations superimposed on a membrane carrier wave," J. Acoust. Soc. Am. **103**, 2283–2286 (1998).
[7] J. D. Achenbach, *Reciprocity in Elastodynamics* (Cambridge University Press, Cambridge, UK, 2004).
[8] J. D. Achenbach, "Explicit solutions for carrier waves supporting surface waves and plate waves," Wave Motion **28**, 89–97 (1998).
[9] J. D. Achenbach, "Laser excitation of surface wave motion," J. Mech. Phys. Solids **51**, 1885–1902 (2003).
[10] Y. H. Berthelot, "Half-order derivative formulation for the analysis of laser-generated Rayleigh waves," Ultrasonics **32**, 153–154 (1994).

Electromagnetoacoustic surface waves on dispersive piezoelectric layered media

Maurizio Romeo[a)]
D.I.B.E., Università di Genova, via Opera Pia 11/a

(Received 23 April 2004; revised 18 June 2004; accepted 18 June 2004)

A full electromagnetic approach is employed to study the propagation of SH surface waves on a piezoelectric solid layer superimposed on a piezoelectric solid substrate. The materials of the layered structure allow for dielectric dispersion and two different boundary conditions are considered at the free surface. The dispersion laws for both boundary conditions are derived in a general form and some special cases are discussed. The solutions of the dispersion equations reveal that besides the expected quasi-acoustic dispersive waves, new quasi-electromagnetic modes exist which propagate at high speeds and depths within the layered structure. These dispersive modes occur also in nondispersive materials. Moreover, in addition to subsonic quasi-acoustic modes, accounting for dielectric dispersion gives rise to quasi-acoustic surface modes whose speed exceeds the bulk wave speed of the layer. Special electroelastic surface waves are also discussed which are characterized by the absence of mechanical perturbation in the substrate. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1780574]

PACS numbers: 43.35.Pt, 68.35.Iv, 77.65.Dq [ANN] Pages: 1488–1497

I. INTRODUCTION

In dealing with wave propagation in nondispersive electroelastic media, it is often assumed that the influence of rapidly varying electromagnetic fields is negligible. Thus, a quasi-static assumption is made which consists in writing the electric field as the gradient of a scalar potential. In almost all the practical applications of surface waves on piezoelectric half-spaces this hypothesis allows us to obtain a proper qualitative and quantitative description of quasi-acoustic surface modes such as piezoelectric Rayleigh waves or Bleustein–Gulyaev waves.[1,2] Many related problems on surface and interfacial waves propagating in anisotropic or layered piezoelectric media are based on this assumption (see, for example, Refs. 3, 4 and references therein). Actually, it has been shown that a full electromagnetic theory, without the shortcut of the quasi-static assumption, slightly modifies the wave speed and the propagation depth of quasi-acoustic shear horizontal (SH) surface modes in piezoelectric nondispersive media.[5–7] Besides, although it accounts for possible optical effects, such a theory allows for a unique solution to the dispersion equation for admissible SH surface waves and this solution is always a quasi-acoustic mode.[6]

If the dielectric response of the electroelastic medium is nonlocal in time, so that dispersive properties of the material are not negligible, the dispersion law and the wave amplitude are expected to change significantly with respect to nondispersive media. In this case a full electromagnetic theory can reveal the existence of quasi-electromagnetic modes which do not appear in nondispersive media.[6] We remark that a relevant consequence of accounting for dielectric dispersion is that the wave speed of SH surface modes on a piezoelectric half-space becomes frequency dependent, i.e., the wave is dispersive.

Dispersive surface waves also arise from the presence of a characteristic length in the material structure, independently of the occurrence of dielectric dispersion. This is the case of an unbounded plate of given thickness or of a layered half-space, and one can be reasonably questioned about the existence of quasi-electromagnetic modes on such piezoelectric structures. This motivation is at the basis of the present work in which SH surface waves are considered on a piezoelectric layer superimposed to a piezoelectric unbounded substrate, within a full electromagnetic approach. To the author's knowledge, no similar investigations appear in the current literature on nondispersive and dispersive piezoelectric media.

Our main task consists in deriving the dispersion equations for electromagnetoacoustic waves under two different boundary conditions at the free surface of the layer, assuming that the materials of the layered structure are dispersive. In this respect we adopt a linear theory of continuum electromechanics with memory-dependent dielectric properties. The constitutive model and the governing equations are stated in Sec. II. The form of the solution for monochromatic plane waves is derived in Sec. III where the pertinent boundary conditions are imposed together with the restrictions on the wave numbers to allow for surface waves. The dispersion equations are then obtained in Sec. IV for the cases of electrically grounded free surfaces and surfaces matched with an external potential. The equations here obtained generalize some results derived in Ref. 8 for a simplified version of the nondispersive problem. Special solutions are also worked out in Sec. V for a thin layer where, in particular, an approximated dispersion law is given for grounded free surfaces. There, we also show the occurrence of special modes which are characterized by the absence of mechanical perturbation in the substrate (see also Ref. 8). Among the main results obtained from the numerical solution of the dispersion equations, we remark on the existence of quasi-electromagnetic

a)Electronic mail: romeo@dibe.unige.it

 © 2004 Acoustical Society of America

modes in both dispersive and nondispersive materials. These waves propagate with speeds higher than the speed of light in the layer and lower than the speed of light in the substrate and their penetration depth within the substrate is much higher than the layer's thickness. In the case of dispersive materials we also obtain quasi-acoustic modes whose speed exceeds the slowest bulk wave speed of the structure.

II. ELECTROMAGNETOELASTIC DISPERSIVE LAYERS

We consider an unbounded solid plane layer $\mathcal{L}$ of thickness d superimposed to a solid half-space $\mathcal{B}$. Both $\mathcal{L}$ and $\mathcal{B}$ consist of homogeneous electroelastic, nonmagnetizable, dispersive media and are in welded contact at the common interface $\mathcal{I}$. We assume that a common sixfold axis of material symmetry exists in $\mathcal{L}$ and $\mathcal{B}$, which is parallel to $\mathcal{I}$. A reference coordinate system $\{O,\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ is chosen in such a way that $\mathbf{e}_3$ lies along the axis of symmetry, $\mathbf{e}_2$ is orthogonal to $\mathcal{I}$ and $x_2=d$, $x_2=0$ represent, respectively, the interface $\mathcal{I}$ and the free surface $\mathcal{S}$ of the layer. The half-space $x_2<0$ is supposed to be empty and will be denoted by $\mathcal{V}$.

Accounting for dispersive effects due to the noninstantaneous dielectric polarization within $\mathcal{L}$ and $\mathcal{B}$, we adopt a theory of memory-dependent electromagnetoelastic continua (see, for example, Refs. 9, 10). In particular we assume that the electric displacement $\mathbf{D}$ depends on the past history of the electric field $\mathbf{e}$. Having in mind the high frequency behavior of the layered structure, both viscoelastic and electroelastic dispersive coupling will be neglected. Denoting by $\mathbf{T}$ and $\mathbf{h}=\frac{1}{2}[\nabla\mathbf{u}+(\nabla\mathbf{u})^T]$, respectively, the Cauchy stress tensor and the infinitesimal strain tensor, we assume the following linear constitutive equations to hold in $\mathcal{L}$ and $\mathcal{B}$:

$$\mathbf{D}(\mathbf{x},t)=\boldsymbol{\varepsilon}[\mathbf{e}(\mathbf{x},\cdot)](t)+\mathbf{E}\mathbf{h}(\mathbf{x},t), \tag{1}$$

$$\mathbf{T}(\mathbf{x},t)=\mathbf{G}\mathbf{h}(\mathbf{x},t)-\mathbf{e}(\mathbf{x},t)\mathbf{E}, \tag{2}$$

where $\mathbf{G}$ and $\mathbf{E}$ are, respectively, the elasticity (fourth-order and positive definite) tensor and the piezoelectric (third-order) tensor. According to our dispersive model the dielectric tensor operator $\boldsymbol{\varepsilon}$ is defined as

$$\boldsymbol{\varepsilon}[\mathbf{f}(\mathbf{x},\cdot)](t)=\boldsymbol{\epsilon}(0)\mathbf{f}(\mathbf{x},t)+\int_0^\infty \boldsymbol{\epsilon}'(\tau)\mathbf{f}(\mathbf{x},t-\tau)\,\mathrm{d}\tau, \tag{3}$$

for any field $\mathbf{f}(t)$ integrable on $\mathbb{R}^+$. The dielectric kernel $\boldsymbol{\epsilon}(t)$ is such that $\boldsymbol{\epsilon}'(t)\to 0$ as $t\to+\infty$ and the instantaneous permittivity $\boldsymbol{\epsilon}(0)$ is taken to be positive definite. In addition, according to thermodynamics, the half-range sine transform of the kernel derivative $\boldsymbol{\epsilon}'(t)$ is taken to be positive semidefinite on $\mathbb{R}^+$ (see Ref. 11).

In view of a full electromagnetic model we introduce the scalar and vector potentials ϕ, $\mathbf{A}$ to express the electric field and the magnetic induction as

$$\mathbf{e}=-\nabla\phi-\frac{\partial\mathbf{A}}{\partial t},\quad \mathbf{B}=\nabla\times\mathbf{A}. \tag{4}$$

Since we ignore magnetization, the magnetic field will be given by $\mathbf{H}=(1/\mu_0)\,\mathbf{B}$, where μ_0 is the magnetic permeability in vacuum. According to Eqs. (1), (2) and (4), the Ampère law and the equation of motion take the form

$$\nabla(\nabla\cdot\mathbf{A})-\Delta\mathbf{A}=-\mu_0\frac{\partial}{\partial t}\left\{\boldsymbol{\varepsilon}\left[\nabla\phi+\frac{\partial\mathbf{A}}{\partial t}\right]-\mathbf{E}\nabla\mathbf{u}\right\}, \tag{5}$$

$$\rho\ddot{\mathbf{u}}=\nabla\cdot(\mathbf{G}\nabla\mathbf{u})+\nabla\cdot\left[\left(\nabla\phi+\frac{\partial\mathbf{A}}{\partial t}\right)\mathbf{E}\right], \tag{6}$$

where ρ is the mass density and where we have exploited the known symmetries of the tensors $\mathbf{G}$ and $\mathbf{E}$.

In the following we shall deal with waves propagating along $\mathbf{e}_1$ and assume that $\mathbf{u}$, ϕ, $\mathbf{A}$ do not depend on x_3. As shown in a previous work,[6] the occurrence of the sixfold axis of symmetry allows Eqs. (5) and (6) to decouple into two sets of three integrodifferential equations which involve, respectively, the fields' components (A_3,u_1,u_2) and (A_1,A_2,ϕ,u_3). Since we are interested in shear waves we consider only the second set which can be closed by introducing the following Lorentz gauge constraint:

$$\nabla\cdot\mathbf{A}_s+\mu_0\varepsilon\left[\frac{\partial\phi}{\partial t}\right]=0, \tag{7}$$

where $\mathbf{A}_s=A_1\mathbf{e}_1+A_2\mathbf{e}_2$ is the *sagittal* part of the vector potential and ε denotes the entry ε_{11} of the operator $\boldsymbol{\varepsilon}$. The resulting equations can be rewritten in the following form:[5,6]

$$\rho\frac{\partial^2 u_3}{\partial t^2}=\overline{G}[\Delta u_3], \tag{8}$$

$$\frac{\partial^2\psi}{\partial t^2}=\frac{1}{\mu_0}\varepsilon^{-1}[\Delta\psi], \tag{9}$$

$$\Delta\mathbf{A}_s-\mu_0\varepsilon\left[\frac{\partial^2\mathbf{A}_s}{\partial t^2}\right]+\mu_0 E\frac{\partial}{\partial t}\nabla u_3=0, \tag{10}$$

where $E=E_{15}$, $\overline{G}=G_{55}+E^2\varepsilon^{-1}$ is a scalar integral operator and where ψ is a suitable scalar potential given by

$$\psi=\phi-E(\varepsilon^{-1}\circ R)[u_3],$$

with R defined as

$$R^{-1}[f](t)=f(t)-\frac{\mu_0}{\rho}\overline{G}\varepsilon[f](t),$$

for any $f(t)$ integrable on $\mathbb{R}^+$.

In the vacuum half-space $x_2<0$ the governing equations for $\phi\equiv\psi$ and $\mathbf{A}_s$ reduce to

$$\Delta\phi-\mu_0\epsilon_0\frac{\partial^2\phi}{\partial t^2}=0,\quad \Delta\mathbf{A}_s-\mu_0\epsilon_0\frac{\partial^2\mathbf{A}_s}{\partial t^2}=0, \tag{11}$$

together with the Lorentz condition

$$\nabla\cdot\mathbf{A}_s+\mu_0\epsilon_0\frac{\partial\phi}{\partial t}=0, \tag{12}$$

where ϵ_0 is the dielectric permittivity in vacuum.

III. SH SURFACE WAVES

We look for monochromatic plane wave solutions to Eqs. (7)–(12) propagating along $\mathbf{e}_1$ in the regions $\mathcal{L}$, $\mathcal{B}$, and $\mathcal{V}$. To this end, material parameters and fields pertaining to $\mathcal{V}$, $\mathcal{L}$, and $\mathcal{B}$ will be denoted, respectively, by superimposed indexes 0, 1, and 2. Posing $u_3=u$, according to the present model we write

$$\begin{pmatrix}\phi^{(0)}\\ \mathbf{A}_s^{(0)}\end{pmatrix}(x_1,x_2,t)=\begin{pmatrix}\Phi^{(0)}\\ \mathbf{a}_s^{(0)}\end{pmatrix}\exp(-i\kappa^{(0)}x_2)\times\exp(-i\omega t+ikx_1),\quad x_2<0, \tag{13}$$

$$\begin{pmatrix}u^{(1)}\\ \mathbf{A}_s^{(1)}\end{pmatrix}(x_1,x_2,t)=\begin{pmatrix}U^{(1)}\\ \mathbf{a}_s^{(1)}\end{pmatrix}(x_2)\exp(-i\omega t+ikx_1),\quad 0\leqslant x_2<d, \tag{14}$$

$$\psi^{(1)}(x_1,x_2,t)=\Psi^{(1)}(x_2)\exp(-i\omega t+ikx_1),$$

$$\begin{pmatrix}u^{(2)}\\ \mathbf{A}_s^{(2)}\end{pmatrix}(x_1,x_2,t)=\begin{pmatrix}U^{(2)}\\ \mathbf{a}_s^{(2)}\end{pmatrix}(x_2)\exp(-i\omega t+ikx_1),\quad x_2\geqslant d, \tag{15}$$

$$\psi^{(2)}(x_1,x_2,t)=\Psi^{(2)}(x_2)\exp(-i\omega t+ikx_1),$$

where $\omega\in\mathbb{R}^{++}$. Without loss in generality we assume that the real part of the wave numer k be strictly positive. In addition, to allow for bounded solutions in $\mathcal{V}$, we require

$$\Im(\kappa^{(0)})\geqslant 0. \tag{16}$$

In view of the form of the integral operators $\varepsilon,\overline{G}$ and R, and owing to the time dependence of the fields in $\mathcal{L}$ and $\mathcal{B}$, we can write

$$\varepsilon^{(j)}[f(t)]=\hat{\boldsymbol{\epsilon}}^{(j)}(\omega)f(t),\quad \overline{G}^{(j)}[f(t)]=\hat{G}^{(j)}(\omega)f(t),$$

$$R^{(j)}[f(t)]=\hat{R}^{(j)}(\omega)f(t),\quad j=1,2,$$

for any $f(t)=F\exp(-i\omega t)$, where

$$\hat{\boldsymbol{\epsilon}}^{(j)}(\omega)=\boldsymbol{\epsilon}^{(j)}(0)+\int_0^\infty(\boldsymbol{\epsilon}^{(j)})'(\tau)\exp(i\omega\tau)\,\mathrm{d}\tau,\quad j=1,2,$$

$$\hat{G}^{(j)}(\omega)=G^{(j)}+\frac{E^{(j)2}}{\hat{\boldsymbol{\epsilon}}^{(j)}(\omega)},$$

$$\hat{R}^{(j)}=\left(1-\frac{\mu_0}{\rho^{(j)}}\hat{G}^{(j)}(\omega)\hat{\boldsymbol{\epsilon}}^{(j)}(\omega)\right)^{-1},\quad j=1,2,$$

are complex-valued functions of ω.

In view of Eqs. (15), the governing equations (8) and (9) in $\mathcal{B}$ give

$$U^{(2)}(x_2)=\overline{U}^{(2)}\exp(i\kappa_u^{(2)}x_2),\qquad \Psi^{(2)}(x_2)=\overline{\Psi}^{(2)}\exp(i\kappa_\psi^{(2)}x_2), \tag{17}$$

where

$$\kappa_u^{(2)^2}=\frac{\rho^{(2)}\omega^2}{\hat{G}^{(2)}}-k^2,\quad \kappa_\psi^{(2)^2}=\mu_0\hat{\boldsymbol{\epsilon}}^{(2)}\omega^2-k^2. \tag{18}$$

Substituting into Eq. (10) we obtain

$$\begin{pmatrix}a_1^{(2)}\\ a_2^{(2)}\end{pmatrix}(x_2)=\begin{pmatrix}\overline{a}_1^{(2)}\\ \overline{a}_2^{(2)}\end{pmatrix}\exp(i\kappa_\psi^{(2)}x_2)+\frac{\mu_0E^{(2)}}{\rho^{(2)}\omega}\times\hat{G}^{(2)}\hat{R}^{(2)}\begin{pmatrix}k\\ \kappa_u^{(2)}\end{pmatrix}\overline{U}^{(2)}\exp(i\kappa_u^{(2)}x_2). \tag{19}$$

Analogously, in view of Eqs. (14), from Eqs. (8)–(10), the electromagnetoelastic field in $\mathcal{L}$ is given by

$$U^{(1)}(x_2)=U_+^{(1)}\exp(i\kappa_u^{(1)}x_2)+U_-^{(1)}\exp(-i\kappa_u^{(1)}x_2),$$

$$\Psi^{(1)}(x_2)=\Psi_+^{(1)}\exp(i\kappa_\psi^{(1)}x_2)+\Psi_-^{(1)}\exp(-i\kappa_\psi^{(1)}x_2),$$

$$\begin{pmatrix}a_1^{(1)}\\ a_2^{(1)}\end{pmatrix}(x_2)=\begin{pmatrix}a_{1+}^{(1)}\\ a_{2+}^{(1)}\end{pmatrix}\exp(i\kappa_\psi^{(1)}x_2)+\begin{pmatrix}a_{1-}^{(1)}\\ a_{2-}^{(1)}\end{pmatrix}\times\exp(-i\kappa_\psi^{(1)}x_2)+\frac{\mu_0E^{(1)}}{\rho^{(1)}\omega}\hat{G}^{(1)}\hat{R}^{(1)}\times\left[\begin{pmatrix}k\\ \kappa_u^{(1)}\end{pmatrix}U_+^{(1)}\exp(i\kappa_u^{(1)}x_2)+\begin{pmatrix}k\\ -\kappa_u^{(1)}\end{pmatrix}U_-^{(1)}\exp(-i\kappa_u^{(1)}x_2)\right], \tag{20}$$

where

$$\kappa_u^{(1)^2}=\frac{\rho^{(1)}\omega^2}{\hat{G}^{(1)}}-k^2,\quad \kappa_\psi^{(1)^2}=\mu_0\hat{\boldsymbol{\epsilon}}^{(1)}\omega^2-k^2. \tag{21}$$

Finally, from Eqs. (11), the solution in $\mathcal{V}$ is given by Eq. (13) with

$$\kappa^{(0)^2}=\mu_0\boldsymbol{\epsilon}_0\omega^2-k^2. \tag{22}$$

Equations (17)–(21) represent effective surface waves if the following restrictions are satisfied:

$$\Im(\kappa_u^{(2)})>0,\quad \Im(\kappa_\psi^{(2)})>0, \tag{23}$$

while no restrictions are required for $\kappa_u^{(1)}$ and $\kappa_\psi^{(1)}$.

In order to obtain the pertinent dispersion equations for $k(\omega)$ we must impose boundary conditions on $\mathcal{I}$ and $\mathcal{S}$. Since $\mathcal{L}$ and $\mathcal{B}$ are in welded contact, posing $\mathbf{n}=\mathbf{e}_2$, we require the following continuity conditions at $\mathcal{I}$:

$$[\![\mathbf{u}]\!]_d=\mathbf{0},\quad [\![\mathbf{Tn}]\!]_d=\mathbf{0},\quad [\![\mathbf{n}\times\mathbf{e}]\!]_d=\mathbf{0},$$

$$[\![\mathbf{n}\times\mathbf{H}]\!]_d=\mathbf{0},\quad [\![\mathbf{n}\cdot\mathbf{D}]\!]_d=0,$$

where $[\![f]\!]_d=f|_{x_2=d^+}-f|_{x_2=d^-}$. Explicitly we get

$$U_+^{(1)}e_{u+}^{(1)}+U_-^{(1)}e_{u-}^{(1)}=\overline{U}^{(2)}e_u^{(2)}, \tag{24}$$

$$\hat{G}^{(1)}\kappa_u^{(1)}[U_+^{(1)}e_{u+}^{(1)}-U_-^{(1)}e_{u-}^{(1)}]+\kappa_\psi^{(1)}E^{(1)}[\Psi_+^{(1)}e_{\psi+}^{(1)}-\Psi_-^{(1)}e_{\psi-}^{(1)}]-\omega E^{(1)}[a_{2+}^{(1)}e_{\psi+}^{(1)}+a_{2-}^{(1)}e_{\psi-}^{(1)}]=\hat{G}^{(2)}\kappa_u^{(2)}\overline{U}^{(2)}e_u^{(2)}+\kappa_\psi^{(2)}E^{(2)}\overline{\Psi}^{(2)}e_\psi^{(2)}-\omega E^{(2)}\overline{a}_2^{(2)}e_\psi^{(2)}, \tag{25}$$

$$-k[\Psi_+^{(1)}e_{\psi+}^{(1)}+\Psi_-^{(1)}e_{\psi-}^{(1)}]-kE^{(1)}(\hat{\boldsymbol{\epsilon}}^{(1)})^{-1}[U_+^{(1)}e_{u+}^{(1)}+U_-^{(1)}e_{u-}^{(1)}]+\omega[a_{1+}^{(1)}e_{\psi+}^{(1)}+a_{1-}^{(1)}e_{\psi-}^{(1)}]=-k\overline{\Psi}^{(2)}e_\psi^{(2)}-kE^{(2)}(\hat{\boldsymbol{\epsilon}}^{(2)})^{-1}\overline{U}^{(2)}e_u^{(2)}+\omega\overline{a}_1^{(2)}e_\psi^{(2)}, \tag{26}$$

$$k[a_{2+}^{(1)}e_{\psi+}^{(1)}+a_{2-}^{(1)}e_{\psi-}^{(1)}]-\kappa_\psi^{(1)}[a_{1+}^{(1)}e_{\psi+}^{(1)}-a_{1-}^{(1)}e_{\psi-}^{(1)}]=k\overline{a}_2^{(2)}e_\psi^{(2)}-\kappa_\psi^{(2)}\overline{a}_1^{(2)}e_\psi^{(2)}, \tag{27}$$

$$\hat{\boldsymbol{\epsilon}}^{(1)}[\kappa_\psi^{(1)}(\Psi_+^{(1)}e_{\psi+}^{(1)}-\Psi_-^{(1)}e_{\psi-}^{(1)})-\omega(a_{2+}^{(1)}e_{\psi+}^{(1)}+a_{2-}^{(1)}e_{\psi-}^{(1)})]=\hat{\boldsymbol{\epsilon}}^{(2)}[\kappa_\psi^{(2)}\overline{\Psi}^{(2)}-\omega\overline{a}_2^{(2)}]e_\psi^{(2)}, \tag{28}$$

where

$$e_{u\pm}^{(j)}=\exp[\pm i\kappa_u^{(j)}d],\quad e_{\psi\pm}^{(j)}=\exp[\pm i\kappa_\psi^{(j)}d]\quad (j=1,2).$$

We assume that the surface $\mathcal{S}$ is free from mechanical traction and distinguish two different electrical boundary conditions. In a first case we suppose that $\mathcal{S}$ is interfaced with an external field and in the second case we allow it to be electrically grounded. In both cases we impose

$$\mathbf{Tn}|_{\mathcal{S}}=\mathbf{0},$$

which amounts to $T_{32}|_{x_2=0}=0$. More explicitly,

$$\hat{G}^{(1)}\kappa_u^{(1)}(U_+^{(1)}-U_-^{(1)})+\kappa_\psi^{(1)}E^{(1)}(\Psi_+^{(1)}-\Psi_-^{(1)})-\omega E^{(1)}(a_{2+}^{(1)}+a_{2-}^{(1)})=0. \tag{29}$$

In addition, if $\mathcal{S}$ is matched with an external field we pose

$$[\![\mathbf{n}\times\mathbf{e}]\!]_0=\mathbf{0},\quad [\![\mathbf{n}\times\mathbf{H}]\!]_0=\mathbf{0},\quad [\![\mathbf{n}\cdot\mathbf{D}]\!]_0=0,$$

which yields

$$-k(\Psi_+^{(1)}+\Psi_-^{(1)})-kE^{(1)}(\hat{\epsilon}^{(1)})^{-1}(U_+^{(1)}+U_-^{(1)})+\omega(a_{1+}^{(1)}+a_{1-}^{(1)})=-k\Phi^{(0)}+\omega a_1^{(0)}, \tag{30}$$

$$k(a_{2+}^{(1)}+a_{2-}^{(1)})-\kappa_\psi^{(1)}(a_{1+}^{(1)}-a_{1-}^{(1)})=ka_2^{(0)}+\kappa^{(0)}a_1^{(0)}, \tag{31}$$

$$-\hat{\epsilon}^{(1)}\kappa_\psi^{(1)}(\Psi_+^{(1)}-\Psi_-^{(1)})+\omega\hat{\epsilon}^{(1)}(a_{2+}^{(1)}+a_{2-}^{(1)})=\epsilon_0(\kappa^{(0)}\Phi^{(0)}+\omega a_2^{(0)}). \tag{32}$$

If $\mathcal{S}$ is grounded we require the scalar potential ϕ to vanish at $\mathcal{S}$, that is,

$$\Psi_+^{(1)}+\Psi_-^{(1)}+E^{(1)}(\hat{\epsilon}^{(1)})^{-1}\hat{R}^{(1)}(U_+^{(1)}+U_-^{(1)})=0. \tag{33}$$

Equations (24)–(33) must be supplemented with the gauge conditions (7) and (12) which hold, respectively, in the material structure and in vacuum. In $\mathcal{L}$ they read as

$$\begin{aligned}\omega\mu_0\hat{\epsilon}^{(1)}\Psi_+^{(1)}&=ka_{1+}^{(1)}+\kappa_\psi^{(1)}a_{2+}^{(1)},\\ \omega\mu_0\hat{\epsilon}^{(1)}\Psi_-^{(1)}&=ka_{1-}^{(1)}-\kappa_\psi^{(1)}a_{2-}^{(1)},\end{aligned} \tag{34}$$

while in $\mathcal{B}$ and in $\mathcal{V}$ we get, respectively,

$$\omega\mu_0\hat{\epsilon}^{(2)}\overline{\Psi}^{(2)}=k\overline{a}_1^{(2)}+\kappa_\psi^{(2)}\overline{a}_2^{(2)}, \tag{35}$$

and

$$\omega\mu_0\epsilon_0\Phi^{(0)}=ka_1^{(0)}-\kappa^{(0)}a_2^{(0)}. \tag{36}$$

IV. DISPERSION EQUATIONS

The boundary conditions derived in the previous section can be conveniently rearranged to obtain a set of equations involving only the fields in $\mathcal{L}$. We discuss separately the case of a matched surface $\mathcal{S}$ and the case of $\mathcal{S}$ grounded.

A. Matched surface

Owing to Eqs. (34)–(36) the conditions (31) and (27) turn out to be equivalent, respectively, to Eqs. (32) and (28). Hence Eqs. (29)–(31) and (24)–(27) can be reduced to a set of equations involving $U_+^{(1)}, U_-^{(1)}$ and the quantities

$$b_+^{(1)}=\kappa_\psi^{(1)}a_{1+}^{(1)}-ka_{2+}^{(1)},\quad b_-^{(1)}=\kappa_\psi^{(1)}a_{1-}^{(1)}+ka_{2-}^{(1)}.$$

We obtain the homogeneous system,

$$\mathbf{Q}_m\mathbf{z}_m=\mathbf{0}, \tag{37}$$

where $\mathbf{z}_m=(U_+^{(1)},U_-^{(1)},b_+^{(1)},b_-^{(1)})^T$,

$$\mathbf{Q}_m=\begin{pmatrix} \hat{G}^{(1)}\kappa_u^{(1)} & -\hat{G}^{(1)}\kappa_u^{(1)} & \dfrac{kE^{(1)}}{\omega\mu_0\hat{\epsilon}^{(1)}} & -\dfrac{kE^{(1)}}{\omega\mu_0\hat{\epsilon}^{(1)}} \\ -\mu_0\omega k\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}} & -\mu_0\omega k\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}} & \dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\dfrac{\kappa^{(0)}}{\epsilon_0} & \dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}-\dfrac{\kappa^{(0)}}{\epsilon_0} \\ (\hat{G}^{(1)}\kappa_u^{(1)}-\hat{G}^{(2)}\kappa_u^{(2)})e_{u+}^{(1)} & -(\hat{G}^{(1)}\kappa_u^{(1)}+\hat{G}^{(2)}\kappa_u^{(2)})e_{u-}^{(1)} & \dfrac{kE^{(1)}}{\mu_0\omega\hat{\epsilon}^{(1)}}\mathcal{E}e_{\psi+}^{(1)} & -\dfrac{kE^{(1)}}{\mu_0\omega\hat{\epsilon}^{(1)}}\mathcal{E}e_{\psi-}^{(1)} \\ -k\mu_0\omega\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}}\mathcal{E}e_{u+}^{(1)} & -k\mu_0\omega\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}}\mathcal{E}e_{u-}^{(1)} & \left(\dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}-\dfrac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi+}^{(1)} & \left(\dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\dfrac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi-}^{(1)} \end{pmatrix},$$

and where

$$\mathcal{E}=1-\frac{E^{(2)}\hat{\epsilon}^{(1)}}{E^{(1)}\hat{\epsilon}^{(2)}}.$$

In order to get nontrivial solutions we impose

$$\det\mathbf{Q}_m=0. \tag{38}$$

We observe that, in particular, Eq. (38) provides the dispersion laws for the corresponding surface wave problem in a half-space and the interfacial wave problem at the separation boundary between two half-spaces. In fact, if $d\to\infty$, we require that inequalities (23) also hold for $\kappa_u^{(1)}$ and $\kappa_\psi^{(1)}$. It follows that $e_{\psi+}^{(1)}\to 0$, $e_{u+}^{(1)}\to 0$, as $d\to\infty$. Then, Eq. (38) simplifies to

$$F_m^{(1)}:=\hat{G}^{(1)}\kappa_u^{(1)}\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\frac{\kappa^{(0)}}{\epsilon_0}\right)+k^2\left(\frac{E^{(1)}}{\hat{\epsilon}^{(1)}}\right)^2=0, \tag{39}$$

and

$$(\hat{G}^{(1)}\kappa_u^{(1)}+\hat{G}^{(2)}\kappa_u^{(2)})\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\frac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)+k^2\left(\frac{E^{(1)}}{\hat{\epsilon}^{(1)}}\right)^2\mathcal{E}^2=0. \tag{40}$$

The first equation is the dispersion law for surface waves on the medium 1 pertaining to a matched boundary. The second

one is the dispersion equation for interfacial waves between the media 1 and 2. These dispersion equations are equivalent to those derived in Refs. 5, 6, and 7, and based on a different representation of the vector potentials $\mathbf{A}_s^{(j)}$ $(j=1,2)$. The dispersion law for waves at the free surface of the medium 2 can be obtained by letting $d\to 0$ into Eq. (38). As is expected, the result is $F_m^{(2)}=0$, which is the counterpart in 2 of Eq. (39).

Going back to the general case, we pose

$$c_u := \tfrac{1}{2}(e_{u+}^{(1)}+e_{u-}^{(1)})=\cos(\kappa_u^{(1)}d),$$

$$s_u := \tfrac{1}{2}(e_{u+}^{(1)}-e_{u-}^{(1)})=i\sin(\kappa_u^{(1)}d),$$

$$c_\psi := \tfrac{1}{2}(e_{\psi+}^{(1)}+e_{\psi-}^{(1)})=\cos(\kappa_\psi^{(1)}d),$$

$$s_\psi := \tfrac{1}{2}(e_{\psi+}^{(1)}-e_{\psi-}^{(1)})=i\sin(\kappa_\psi^{(1)}d),$$

and introduce the following nondimensional quantities:

$$\Gamma_c = c_u - \frac{\hat{G}^{(2)}\kappa_u^{(2)}}{\hat{G}^{(1)}\kappa_u^{(1)}}s_u, \quad \chi_c = c_\psi - \frac{\kappa_\psi^{(2)}\hat{\epsilon}^{(1)}}{\kappa_\psi^{(1)}\hat{\epsilon}^{(2)}}s_\psi,$$

$$\chi = c_\psi - \frac{\kappa^{(0)}\hat{\epsilon}^{(1)}}{\kappa_\psi^{(1)}\epsilon_0}s_\psi, \quad \eta^{(1)} = \frac{(E^{(1)})^2}{\hat{G}^{(1)}\hat{\epsilon}^{(1)}}.$$

Then, assuming $d\neq 0$, Eq. (38) can be cast in the form

$$\mathcal{E}^2 z^2 + (\Gamma_c\chi_c + \mathcal{E}^2\chi c_u - 2\mathcal{E})z + (\Gamma_c c_u - 1)(\chi_c\chi - 1) = 0, \tag{41}$$

where

$$z = \eta^{(1)}\frac{k^2}{\kappa_u^{(1)}\kappa_\psi^{(1)}}s_u s_\psi.$$

B. Grounded surface

The pertinent boundary conditions at $\mathcal{S}$ are Eqs. (29) and (33). We note that, in this case, the field's representation (20) is redundant since Eqs. (8)–(10) can be satisfied, compatibly with the condition $\phi=0$ at $\mathcal{S}$, assuming one of the pairs $a_{1\pm}^{(1)}$ or $a_{2\pm}^{(1)}$ to vanish. This amounts to the corresponding component of the vector potential $\mathbf{A}_s^{(1)}$ be proportional to the mechanical displacement u_3. If we take $a_{1\pm}^{(1)}=0$, use Eqs. (34) and rearrange the boundary conditions at $\mathcal{I}$ exactly in the same way as in the matched case, we obtain the following homogeneous system for the fields $a_{2+}^{(1)}$, $a_{2-}^{(1)}$, $U_+^{(1)}$ and $U_-^{(1)}$,

$$\mathbf{Q}_g\mathbf{z}_g = \mathbf{0}, \tag{42}$$

where $\mathbf{z}_g=(U_+^{(1)},U_-^{(1)},a_{2+}^{(1)},a_{2-}^{(1)})^T$,

$$\mathbf{Q}_g = \begin{pmatrix} \hat{G}^{(1)}\kappa_u^{(1)} & -\hat{G}^{(1)}\kappa_u^{(1)} & -\dfrac{k^2E^{(1)}}{\omega\mu_0\hat{\epsilon}^{(1)}} & -\dfrac{k^2E^{(1)}}{\omega\mu_0\hat{\epsilon}^{(1)}} \\ \mu_0\omega E^{(1)}\hat{R}^{(1)} & \mu_0\omega E^{(1)}\hat{R}^{(1)} & \kappa_\psi^{(1)} & -\kappa_\psi^{(1)} \\ (\hat{G}^{(1)}\kappa_u^{(1)}-\hat{G}^{(2)}\kappa_u^{(2)})e_{u+}^{(1)} & -(\hat{G}^{(1)}\kappa_u^{(1)}+\hat{G}^{(2)}\kappa_u^{(2)})e_{u-}^{(1)} & -\dfrac{k^2E^{(1)}}{\mu_0\omega\hat{\epsilon}^{(1)}}\mathcal{E}e_{\psi+}^{(1)} & -\dfrac{k^2E^{(1)}}{\mu_0\omega\hat{\epsilon}^{(1)}}\mathcal{E}e_{\psi-}^{(1)} \\ -\mu_0\omega\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}}\mathcal{E}e_{u+}^{(1)} & -\mu_0\omega\dfrac{E^{(1)}}{\hat{\epsilon}^{(1)}}\mathcal{E}e_{u-}^{(1)} & -\left(\dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}-\dfrac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi+}^{(1)} & \left(\dfrac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\dfrac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi-}^{(1)} \end{pmatrix}.$$

The compatibility condition of system (42) reads as

$$\det\mathbf{Q}_g = 0. \tag{43}$$

As in the previous case we observe that, as $d\to\infty$, we get $e_{u+}^{(1)}\to 0$, $e_{\psi+}^{(1)}\to 0$ and Eq. (43) factorizes into Eq. (40) and

$$\kappa_u^{(1)}\kappa_\psi^{(1)} + k^2\hat{R}^{(1)}\eta^{(1)} = 0. \tag{44}$$

The last equation is the dispersion law for SH surface waves on the electrically grounded boundary of the medium 1. It is equivalent to that derived in Refs. 5, 6 and based on a different representation of the vector potential $\mathbf{A}_s^{(1)}$. We also note that, in contrast to the matched case, Eq. (44) cannot be obtained as the compatibility condition for $d\to 0$, owing to the mismatch of the boundary conditions at the coincident surfaces $\mathcal{S}$ and $\mathcal{I}$.

In the general case Eq. (43) can be cast in the form

$$\hat{R}^{(1)}\mathcal{E}^2z^2 + [\mathcal{E}^2c_uc_\psi - \mathcal{E}(1+\hat{R}^{(1)}) + \hat{R}^{(1)}\Gamma_c\chi_c]z + \Gamma_s\chi_s s_u s_\psi = 0, \tag{45}$$

where

$$\Gamma_s = s_u - \frac{\hat{G}^{(2)}\kappa_u^{(2)}}{\hat{G}^{(1)}\kappa_u^{(1)}}c_u, \quad \chi_s = s_\psi - \frac{\kappa_\psi^{(2)}\hat{\epsilon}^{(1)}}{\kappa_\psi^{(1)}\hat{\epsilon}^{(2)}}c_\psi.$$

C. Final formulas

The dispersion equations obtained in Secs. IV A and IV B can be conveniently rewritten in terms of the variables $s=k/\omega$ and $v=\omega d$. By substituting $\kappa_u^{(j)}$, $\kappa_\psi^{(j)}$, $\kappa^{(0)}$, k, and d, respectively, by $K_u^{(j)}(s)=\kappa_u^{(j)}/\omega$, $K_\psi^{(j)}(s)=\kappa_\psi^{(j)}/\omega$, $K^{(0)}(s)=\kappa^{(0)}/\omega$, s and v, all the quantities c_u, c_ψ, s_u, s_ψ, Γ_c, Γ_s, χ_c, χ_s, χ, and z become functions of only s and v. As a result, Eqs. (41) and (45) take the form

$$z^2(s,v) + 2B_m(s,v)z(s,v) + C_m(s,v) = 0,$$

$$z^2(s,v) + 2B_g(s,v)z(s,v) + C_g(s,v) = 0, \tag{46}$$

where

$$B_m(s,v) = \frac{1}{2\mathcal{E}^2}(\Gamma_c\chi_c + \mathcal{E}^2\chi c_u - 2\mathcal{E}),$$

$$C_m(s,v)=\frac{1}{\mathcal{E}^2}(\Gamma_c c_u-1)(\chi_c\chi-1),$$

$$B_g(s,v)=\frac{1}{2\hat{R}^{(1)}\mathcal{E}^2}[\mathcal{E}^2 c_u c_\psi-\mathcal{E}(1+\hat{R}^{(1)})+\hat{R}^{(1)}\Gamma_c\chi_c],$$

$$C_g(s,v)=\frac{1}{\hat{R}^{(1)}\mathcal{E}^2}\Gamma_s\chi_s s_u s_\psi.$$

Each of Eqs. (46) decouples into the pair ($\alpha=m$, or $\alpha=g$),

$$z(s,v)+B_\alpha(s,v)-[B_\alpha^2(s,v)-C_\alpha(s,v)]^{1/2}=0, \tag{47}$$

$$z(s,v)+B_\alpha(s,v)+[B_\alpha^2(s,v)-C_\alpha(s,v)]^{1/2}=0, \tag{48}$$

and, since $K_u^{(j)}$, $K_\psi^{(j)}$, and $K^{(0)}$ are irrational functions of s, these last equations correspond to several branches in the plane v,s. However, in view of inequalities (16) and (23), not all the branches are expected to describe admissible SH surface modes.

V. THIN LAYERS AND SPECIAL SURFACE MODES

In this section we consider some special solutions to the problem of SH surface wave propagation on a layered structure.

We first derive simplified dispersion equations for the case in which the thickness of the layer is relatively small. More precisely we assume that d is small in comparison to the propagation depth within the layer $\mathcal{L}$. This condition is accomplished by letting $|\kappa_u^{(1)}|d\ll 1$ and $|\kappa_\psi^{(1)}|d\ll 1$, for any $\omega\in\mathbb{R}^{++}$ or, equivalently,

$$|K_u^{(1)}|v\ll 1, \quad |K_\psi^{(1)}|v\ll 1.$$

Expanding c_u, s_u, C_ψ, and s_ψ in powers of $(K_u^{(1)}v)$ and $(K_\psi^{(1)}v)$ we obtain

$$c_u=1-\tfrac{1}{2}{K_u^{(1)}}^2v^2+O((K_u^{(1)}v)^4),$$

$$s_u=iK_u^{(1)}v+O((K_u^{(1)}v)^3),$$

and analogous expressions for c_ψ and s_ψ. Substituting into Eq. (41) and retaining only first-order terms in $K_u^{(1)}v$ and $K_\psi^{(1)}v$, we obtain the thin-layer approximated dispersion law for the matched surface:

$$\mathcal{F}_m^{(2)}(s)-iv\left[\frac{\hat{G}^{(2)}K_u^{(2)}}{\hat{\epsilon}^{(1)}}({K_\psi^{(1)}}^2+\eta^{(1)}s^2)+\frac{\hat{G}^{(1)}K_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right.$$
$$\left.\times(K_u^{(1)}K_\psi^{(1)}+\eta^{(1)}s^2)+\frac{\hat{G}^{(1)}K_e}{\epsilon_0}(K_u^{(1)}K_\psi^{(1)}+\eta^{(1)}\mathcal{E}^2s^2)\right]$$
$$=0, \tag{49}$$

where

$$\mathcal{F}_m^{(2)}(s)=\hat{G}^{(2)}K_u^{(2)}\left(\frac{K_\psi^{(2)}}{\hat{\epsilon}^{(2)}}+\frac{K_e}{\epsilon_0}\right)+s^2\left(\frac{E^{(2)}}{\hat{\epsilon}^{(2)}}\right)^2.$$

By the same token, from Eq. (45), we arrive at the dispersion law for the grounded surface in the thin-layer approximation,

$$\mathcal{F}_g(s)-iv\left[(\eta^{(1)}s^2\hat{R}^{(1)}+{K_\psi^{(1)}}^2)\frac{\hat{G}^{(2)}}{\hat{G}^{(1)}}K_u^{(2)}\right.$$
$$\left.+(\eta^{(1)}s^2\hat{R}^{(1)}+{K_u^{(1)}}^2)\frac{\hat{\epsilon}^{(1)}}{\hat{\epsilon}^{(2)}}K_\psi^{(2)}\right]=0, \tag{50}$$

where

$$\mathcal{F}_g(s)=(\mathcal{E}-\hat{R}^{(1)})(\mathcal{E}-1)\eta^{(1)}s^2+K_u^{(2)}K_\psi^{(2)}\frac{\hat{G}^{(2)}\hat{\epsilon}^{(1)}}{\hat{G}^{(1)}\hat{\epsilon}^{(2)}}.$$

We remark that the zeroth-order dispersion law $\mathcal{F}_g(s)=0$ corresponds to the limit of Eq. (43) as $d\to 0$. It can adequately describe SH surface waves propagating on a piezoelectric film superimposed to a piezoelectric substrate when the free surface is electrically shorted by a thin metallization.

Further special solutions can be derived under additional restrictions on the compatibility conditions for surface waves. In particular we consider here the occurrence of surface modes which carry no mechanical perturbations in $\mathcal{B}$. The existence of such waves has been discussed in Ref. 8, where the interesting property of mechanical energy localization within the layer has been pointed out.

We start from the case of a matched surface $\mathcal{S}$. If the constraint $U^{(2)}=0$ is imposed on Eqs. (24)–(27) and (29)–(31), we obtain, in place of Eq. (37),

$$\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\frac{\kappa^{(0)}}{\epsilon_0}-\frac{k^2{E^{(1)}}^2s_u}{\hat{G}^{(1)}(\hat{\epsilon}^{(1)})^2\kappa_u^{(1)}c_u}\right)b_+^{(1)}$$
$$+\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}-\frac{\kappa^{(0)}}{\epsilon_0}+\frac{k^2{E^{(1)}}^2s_u}{\hat{G}^{(1)}(\hat{\epsilon}^{(1)})^2\kappa_u^{(1)}c_u}\right)b_-^{(1)}=0,$$

$$\left(\mathcal{E}e_{\psi+}^{(1)}-\frac{1}{c_u}\right)b_+^{(1)}-\left(\mathcal{E}e_{\psi-}^{(1)}-\frac{1}{c_u}\right)b_-^{(1)}=0,$$

$$\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}-\frac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi+}^{(1)}b_+^{(1)}+\left(\frac{\kappa_\psi^{(1)}}{\hat{\epsilon}^{(1)}}+\frac{\kappa_\psi^{(2)}}{\hat{\epsilon}^{(2)}}\right)e_{\psi-}^{(1)}b_-^{(1)}=0.$$

The compatibility conditions in order to get nontrivial solutions are

$$\chi_c-\mathcal{E}c_u=0, \quad \mathcal{E}z+\chi\chi_c-1=0. \tag{51}$$

Equation (51) must be solved for s and v. If any, a solution (s,v), compatible with inequalities (16) and (23), will exist only for a suitable matching of material parameters in $\mathcal{L}$ and $\mathcal{B}$.

A similar procedure can be adopted in the case of grounded surfaces. Here, owing to the condition $U^{(2)}=0$, Eqs. (29), (33) and (24)–(27) can be reduced to a system of three equations for $a_+^{(1)}$ and $a_-^{(1)}$, whose compatibility conditions read as

$$\chi_c-\mathcal{E}c_u=0, \quad \mathcal{E}\hat{R}^{(1)}z+\chi_c c_\psi-1=0. \tag{52}$$

Again, the existence of solutions to these equations implies an additional constraint on the material parameters in $\mathcal{L}$ and $\mathcal{B}$.

TABLE I. Material parameters for the piezoelectric substrate $\mathcal{B}$ (PTZ-4) and the piezoelectric layer $\mathcal{L}$ (PTZ-5).

	PTZ-4 ($\mathcal{B}$)	PTZ-5 ($\mathcal{L}$)
$\rho(10^3)$ Kg/m^3	7.50	7.75
$G(10^{10}$ Pa)	2.56	2.11
E(C/m^2)	12.7	12.3
$\epsilon(10^{-9}$ F/m)	6.4634	8.1103
	$c_{\mathcal{B}}=2596$ m/s	$c_{\mathcal{L}}=2265$ m/s
	$v_{\mathcal{B}}=1.109\times10^7$ m/s	$v_{\mathcal{L}}=9.905\times10^6$ m/s

VI. DISPERSIVE MODEL AND NUMERICAL RESULTS

In order to deal with high frequency surface modes we assume that the kernel of the dielectric integral operator $\boldsymbol{\epsilon}(t)$ allows for a Taylor expansion up to the desired order at $t=0$. As a consequence, we write

$$\hat{\boldsymbol{\epsilon}}^{(j)}(\omega)=\boldsymbol{\epsilon}_0^{(j)}\left[1+\sum_{h=1}^{N} i^h\frac{\boldsymbol{\epsilon}_h^{(j)}}{\omega^h}\right]+R_{\boldsymbol{\epsilon}}^{(j)}(\omega),$$

$$\omega\in\mathbb{R}^{++},\quad j=1,2, \tag{53}$$

where

$$\boldsymbol{\epsilon}_0^{(j)}=\boldsymbol{\epsilon}^{(j)}(0),\quad \boldsymbol{\epsilon}_h^{(j)}=\frac{\mathrm{d}^h}{\mathrm{d}t^h}\boldsymbol{\epsilon}^{(j)}\bigg|_{t=0},\quad R_{\boldsymbol{\epsilon}}^{(j)}(\omega)=O(\omega^{-N-1}).$$

In view of the semidefiniteness of $\boldsymbol{\epsilon}'(t)$ we shall take $\boldsymbol{\epsilon}_1^{(j)}\geqslant 0$ $(j=1,2)$. In terms of the variable v we rewrite Eq. (53) as

$$\hat{\boldsymbol{\epsilon}}^{(j)}(v)=\boldsymbol{\epsilon}_0^{(j)}\left[1+\sum_{h=1}^{N} i^h\frac{\boldsymbol{\epsilon}_h^{(j)}d^h}{v^h}\right]+R_{\boldsymbol{\epsilon}}^{(j)}(v),$$

$$v\in\mathbb{R}^{++},\quad j=1,2. \tag{54}$$

Owing to their definitions, the quantities $\hat{G}^{(j)}$ and $\hat{R}^{(j)}$ will admit analogous expansions:

$$\hat{G}^{(j)}(v)=G_0^{(j)}+\sum_{h=1}^{N}\frac{G_h^{(j)}}{v^h}+R_G^{(j)}(v),$$

$$\hat{R}^{(j)}(v)=R_0^{(j)}+\sum_{h=1}^{N}\frac{R_h^{(j)}}{v^h}+R_R^{(j)}(v),$$

where $G_1^{(j)}$ and $R_1^{(j)}$ turn out to be nonpositive and where the remainder terms $R_G^{(j)}$ and $R_R^{(j)}$ are $O(v^{-N-1})$. For nondispersive media, each of the quantities $\hat{\boldsymbol{\epsilon}}^{(j)},\hat{G}^{(j)},\hat{R}^{(j)}$ reduces to the zeroth-order term in its respective expansion. The quasistatic approximation of the present model can be achieved by letting $\mu_0=0$, and consequently $\hat{R}^{(j)}=1$ $(j=1,2)$.

In the following we discuss numerical solutions for nondispersive and dispersive problems stated in the previous sections, assuming that the materials in $\mathcal{L}$ and $\mathcal{B}$ are, respectively, PTZ-5 and PTZ-4. As in the case of Love's waves in elastic layers the wave speed of SH bulk waves in the layer is lower than that of SH bulk waves in the substrate. Material parameters are given in Table I pertaining to nondispersive media. The quantities $c_{\mathcal{B}}$ and $c_{\mathcal{L}}$ represent the speed of acoustic bulk shear waves in the piezoelectrically stiffened materials 2 and 1, respectively. The quantities $v_{\mathcal{B}}$ and $v_{\mathcal{L}}$ represent the speed of light in materials 2 and 1, respectively.

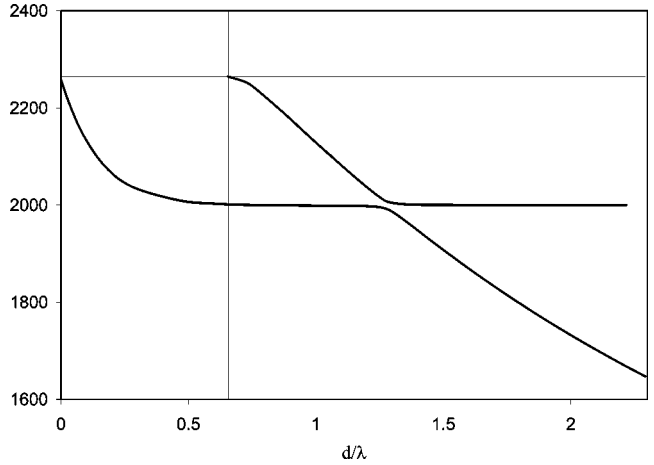


FIG. 1. Speed of two surface modes for a nondipersive layered structure when the free surface is electrically grounded. The horizontal line corresponds to the bulk wave speed in the layer.

Dispersion equations (47) and (48) have been solved with respect to s for given values of v. In doing so we have taken into account that $K_u^{(j)}$, $K_\psi^{(j)}$, and $K^{(0)}$ are defined by Eqs. (21) and (22) except for the choice of sign. Hence we considered all the resulting dispersion equations according to the possible choices of signs. Each solution to these equations has been checked to meet inequalities (16) and (23) in order to represent effective surface waves. In all cases considered here, Eqs. (47) and (48) admit at least one solution which is an effective surface wave. Both quasi-acoustic and quasi-electromagnetic modes have been found which are discussed separately in the following.

A. Quasi-acoustic modes

Looking through the effects of the dispersive model outlined in Eq. (54), we start to examine frequency-independent parameters for piezoelectric materials in $\mathcal{L}$ and $\mathcal{B}$, corresponding to the nondispersive, or zeroth-order, approximation. Two real solutions of Eqs. (47) and (48) are obtained for both grounded and matched free surfaces $\mathcal{S}$. The results are shown in Figs. 1 and 2 where the wave speed $c=1/s$ is

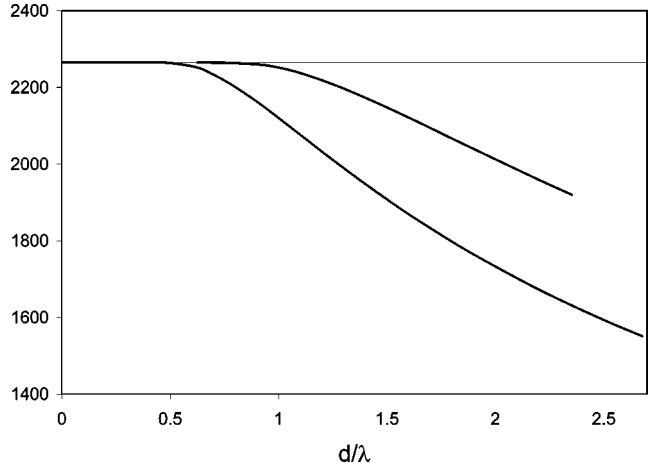


FIG. 2. Speed of two surface modes for a nondipersive layered structure when the free surface is matched with an external potential. The horizontal line corresponds to the bulk wave speed in the layer.

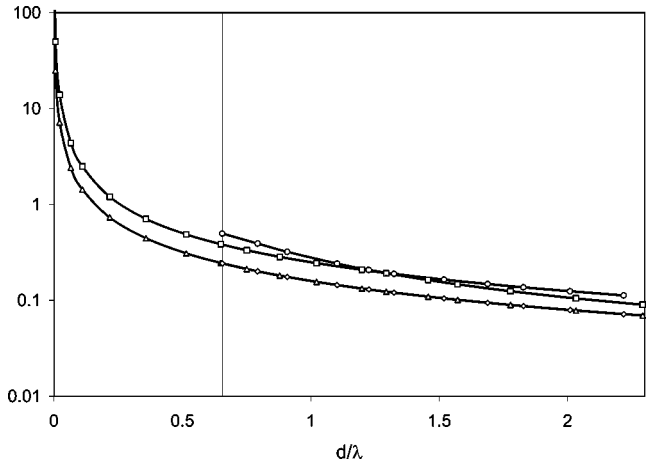


FIG. 3. Relative propagation depths p_u and p_ψ of the two modes represented in Fig. 1. Squares and circles denote p_u while triangles and diamonds denote p_ψ.

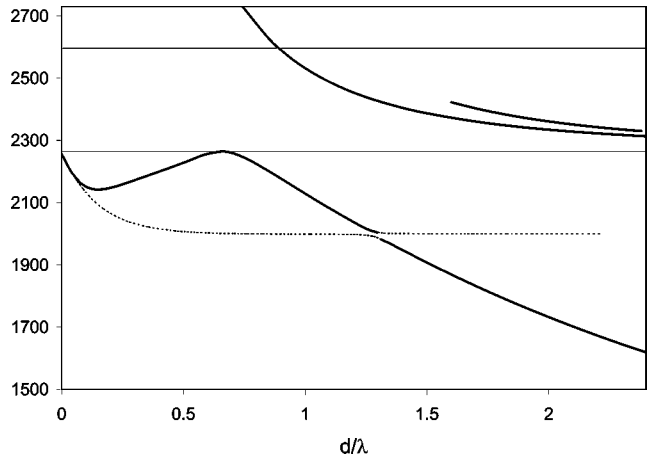


FIG. 4. Speed of subsonic and supersonic surface modes for a grounded free surface in the dispersive case. Lower and higher horizontal lines denote $c_{\mathcal{L}}$ and $c_{\mathcal{B}}$, respectively.

plotted versus the rate $d/\lambda = sv/2\pi$. Both waves are subsonic since their speeds are lower than the smallest speed $c_{\mathcal{L}}$ and one mode shows a cut off value of d/λ. The typical behavior of a resonance coupling appears in the grounded case between two modes near a critical value of d/λ. Such a kind of critical points has been found in surface waves propagating in magnetoelastic or electroelastic media owing to the presence of a small coupling factor.[12,13] In that context, turning the small parameter to zero, two branches of the dispersion curve intersect at the critical point. However, in the present case, a detailed numerical analysis shows that no real values of the material parameters make our solution decouple into two intersecting branches so that no independent modes exist to be coupled if the media 1 and 2 are nondispersive. In both grounded and matched cases the wave speed decreases with the thickness of the layer $\mathcal{L}$ and increases with the wave length, but in the matched case the wave speed remains very close to $c_{\mathcal{L}}$ for small values of d/λ.

The relative propagation depths p_u and p_ψ are given in Fig. 3 for the case of a grounded free surface. These quantities are a measure of the penetration distance of the surface wave within the substrate and are defined as

$$p_u = \frac{1}{\Im(\kappa_u^{(2)})d} = \frac{1}{\Im(K_u^{(2)})v}, \quad p_\psi = \frac{1}{\Im(\kappa_\psi^{(2)})d} = \frac{1}{\Im(K_\psi^{(2)})v}.$$

The behavior of p_u and p_ψ for the matched free surface is qualitatively similar to that of the grounded case and is not given here for brevity. Figure 3 shows that the propagation depth p_u of the purely mechanical field is always greater than that of the electromagnetic field. In addition, small values of d/λ imply the high surface wave's penetration.

Now we account for dielectric dispersion by assuming a first-order approximation of $\hat{\epsilon}^{(j)}$ in Eq. (54). In particular we pose $\epsilon_1^{(1)}d=1$ and $\epsilon_1^{(2)}d=2$. The parameters $\hat{\epsilon}^{(j)}$, $\hat{G}^{(j)}$, and $\hat{R}^{(j)}$ turn out to be complex valued and so are the solutions for s. They consist of a sequence of admissible surface modes which are subsonic or supersonic. In Figs. 4 and 5 the real part of $1/s$ versus the real part of d/λ is shown. Two supersonic modes are given for both grounded and matched cases. Their speeds are decreasing functions of d/λ with values greater than $c_{\mathcal{L}}$. The subsonic modes remain practically unchanged with respect to the nondispersive layer in the matched case (see Fig. 5). In the grounded case the subsonic mode partially follows the pattern of the two branches in the solutions of the nondispersive layer and shows that dispersive effects switch off the resonance coupling (see Fig. 4).

All the quasi-acoustic modes presented in this subsection hold with very slight quantitative differences if the quasi-static assumption is made.

B. Quasi-electromagnetic modes

Accounting for a full electromagnetic theory implies the existence of fast-propagating quasi-electromagnetic surface waves in the piezoelectric layered structure. For both grounded and matched free surfaces we have obtained solutions to Eqs. (47) and (48) which correspond to admissible surface modes. Their speed is always greater than the speed of light $v_{\mathcal{L}}$ and lower than the speed of light $v_{\mathcal{B}}$. The propagation speeds of the first three modes are shown in Fig. 6 for grounded free surfaces. The relative propagation depths p_u

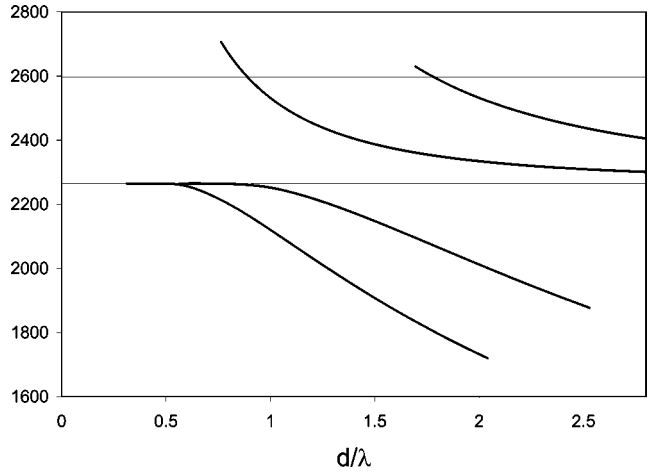


FIG. 5. Speed of subsonic and supersonic surface modes for a matched free surface in the dispersive case. Lower and higher horizontal lines denote $c_{\mathcal{L}}$ and $c_{\mathcal{B}}$, respectively.

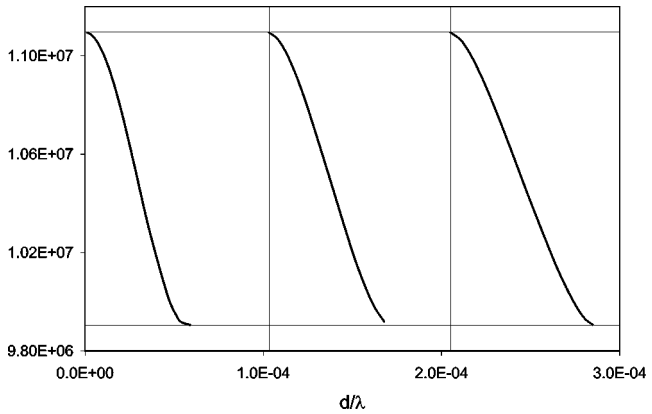


FIG. 6. Speed of quasi-electromagnetic modes for a grounded free surface in the dispersive case. Lower and higher horizontal lines denote $v_{\mathcal{L}}$ and $v_{\mathcal{B}}$, respectively.

and p_{ψ} are shown in Fig. 7. Similar results are obtained in the case of matched free surfaces. Quasi-electromagnetic modes propagate with great depth within the substrate and, as in the quasi-acoustic solutions, the wave speed and the depth are decreasing functions of the layer's thickness and increasing functions of the wave length.

C. Special surface modes

We finally inquire into the existence of solutions to systems (51) and (52), which represent admissible surface waves. As a critical material parameter to describe such solutions we have chosen $\alpha = -E^{(2)}/E^{(1)}$. We have found one admissible solution in terms of the couple (s,v) for both systems (51) and (52) when $0<\alpha<1$. This restriction implies that, at least in the present material setting, SH surface waves which carry no mechanical perturbation within $\mathcal{B}$ are allowed when the polarization axis of the layer $\mathcal{L}$ is opposite to that of the substrate $\mathcal{B}$ and when the absolute value of the piezoelectric coupling constant $E^{(2)}$ of the substrate is smaller than its counterpart $|E^{(1)}|$ in the layer. The first condition has been also pointed out in the results of Jin *et al.*,[8] where two identical materials with opposite polarization axes

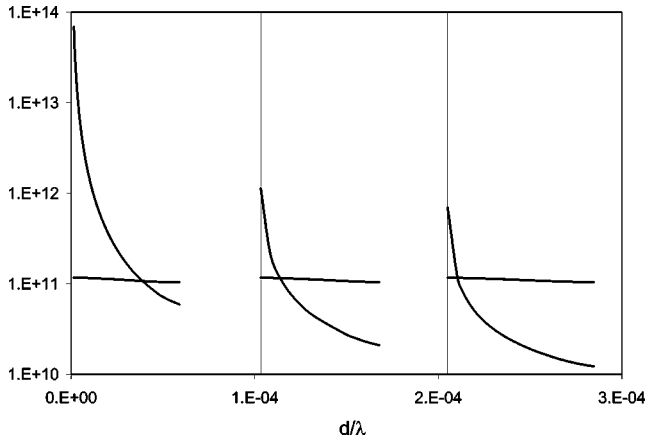


FIG. 7. Relative propagation depths p_u and p_{ψ} of the modes represented in Fig. 6. The steepest curves represent p_u.

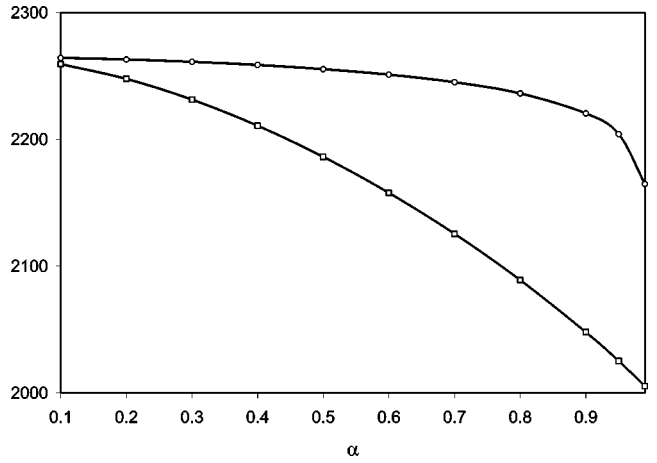


FIG. 8. Speed of special surface waves which do not carry mechanical perturbation in $\mathcal{B}$ versus $\alpha = -E^{(2)}/E^{(1)}$. Grounded free surface (squares) and matched free surface (circles).

are considered in the place of $\mathcal{L}$ and $\mathcal{B}$. In Fig. 8 we give the wave speeds versus the parameter α for both matched and grounded cases.

VII. CONCLUDING REMARKS

The results obtained in this work point out two relevant features of SH surface waves propagating on a layered piezoelectric structure.

The first noticeable aspect is that by accounting for a full electromagnetic theory it is possible to show the existence of quasi-electromagnetic modes which propagate along the surface with a high penetration depth in comparison to the layer's thickness. Such waves are characterized by a wave speed which lies between the slow-layer speed of light, $v_{\mathcal{L}}$, and the speed of light in the substrate, $v_{\mathcal{B}}$. They exist independently of the occurrence of dispersive effects due to the material response of the dielectric media in $\mathcal{L}$ and $\mathcal{B}$.

The second important feature concerns the assumption of dispersive materials. In contrast to the nondispersive case, quasi-acoustic modes exist which propagate at a speed greater than the slow-layer speed $c_{\mathcal{L}}$. In addition, if the free surface $\mathcal{S}$ is electrically grounded, the presence of dispersive effects changes considerably the pattern of subsonic modes reducing two monotone branches of the dispersion curve to one single nonmonotone branch.

We note that all the admissible solutions obtained for dispersive dielectrics are such that $\Im(k)>0$. In other words, the wave amplitude for such modes exponentially decays along the propagation direction.

Finally it is worth remarking that special modes with null mechanical displacement in the substrate $\mathcal{B}$ can always occur when $\mathcal{L}$ and $\mathcal{B}$ are polarized in opposite directions. These modes are found in dispersive and nondispersive media and under the quasi-static hypothesis as well as within the full electromagnetic assumption.

ACKNOWLEDGMENTS

The research leading to this work has been performed under the auspices of GNFM and supported by MIUR within the COFIN.2002 Project ''Modelli Matematici per la Scienza dei Materiali.''

[1] A. A. Maradudin and G. I. Stegeman, ''Surface acoustic waves,'' in *Surface Phonons*, edited by W. Kress and F. W. de Wette (Springer-Verlag, Berlin, 1991).
[2] G. A. Maugin, ''Elastic surface waves with transverse horizontal polarization,'' Adv. Appl. Mech. **23**, 373–434 (1983).
[3] A. N. Darinskii and V. N. Lyubimov, ''Shear interfacial waves in piezoelectrics,'' J. Acoust. Soc. Am. **106**, 3296–3304 (1999).
[4] Z. Qian, F. Jin, Z. Wang, and K. Kishimoto, ''Dispersion relations for SH-wave propagation in periodic piezoelectric composite layered structures,'' Int. J. Eng. Sci. **42**, 673–689 (2004).
[5] S. Li, ''The electromagneto-acoustic surface wave in a piezoelectric medium: The Bleustein–Gulyaev mode,'' J. Appl. Phys. **80**, 5264–5269 (1996).
[6] M. Romeo, ''Non-dispersive and dispersive electromagnetoacoustic SH surface modes in piezoelectric media,'' Wave Motion **39**, 93–110 (2004).
[7] M. Romeo, ''Electromagnetoelastic waves at piezoelectric interfaces,'' Int. J. Eng. Sci. **42**, 753–768 (2004).
[8] F. Jin, Z. Wang, and T. Wang, ''The Bleustein Gulyaev (B–G) wave in a piezoelectric layered half-space,'' Int. J. Eng. Sci. **39**, 1271–1285 (2001).
[9] G. A. Maugin, *Continuum Mechanics of Electromagnetic Solids* (North-Holland, Amsterdam, 1988).
[10] A. C. Eringen and G. A. Maugin, *Electrodynamics of Continua* (Springer-Verlag, New York, 1990), Vol. II.
[11] M. Fabrizio and A. Morro, ''Thermodynamics of electromagnetic isothermal systems with memory,'' J. Non-Equilib. Thermodyn. **22**, 110–128 (1997).
[12] J. Pouget and G. A. Maugin, ''Bleustein–Gulyaev surface modes in elastic ferroelectrics,'' J. Acoust. Soc. Am. **69**, 1304–1318 (1981).
[13] J. Pouget and G. A. Maugin, ''Piezoelectric Rayleigh waves in elastic ferroelectrics,'' J. Acoust. Soc. Am. **69**, 1319–1325 (1981).

Focused array transducer for two-dimensional optoacoustic tomography

Victor Kozhushko, Tanya Khokhlova,[a] Alexei Zharinov, Ivan Pelivanov, Vladimir Solomatin, and Alexander Karabutov
International Laser Center of Moscow State University, 119992, Moscow, Russia

(Received 18 December 2003; revised 22 June 2004; accepted 22 June 2004)

Optoacoustic (OA) imaging utilizes short laser pulses to create acoustic sources in tissue and time resolved detection of generated pressure profiles for image reconstruction. The ultrasonic transients provide information on the distribution of optical absorption coefficient that can be useful for early cancer diagnostics. In this work a new design of wide-band array transducer is developed and tested. The array consists of 32 focused piezo-elements made of PVDF slabs imposed on a cylindrical surface. A single array element response to an OA signal coming from arbitrarily located point source is investigated theoretically and experimentally. The measured signals correspond well to numerically calculated ones. Focal zone maps of the elements with aperture angles 30° and 60° are presented and discussed; the resolution in direction perpendicular to the imaging plane is determined. Point spread function of the whole array is calculated using experimentally obtained signals from the sources located at different distances from the array. Backprojection algorithm is employed for reconstruction of the optoacoustic images. It is shown that the spatial resolution of the images yielded by the proposed array increases significantly compared to previous transducer designs. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781710]

PACS numbers: 43.35.Sx, 43.60.Fg, 43.64.Jb [YHB] Pages: 1498–1506

I. INTRODUCTION

The optoacoustic (OA) effect,[1–3] which refers to the generation of ultrasonic transients through absorption of laser radiation, can be utilized for imaging of biological tissue.[4–11] The method was applied for various medical problems such as early cancer detection,[4–7] teeth structure characterization,[10] and measuring of tissue properties.[11]

One of the most important applications of optoacoustic imaging is early breast cancer diagnostics. The main advantage of the method is high contrast of light absorption rate between cancerous and healthy tissues,[12] whereas the contrast in acoustic impedance of tumors and surrounding tissues utilized in ultrasonic imaging is relatively low. Therefore, optoacoustic tomography can become a new effective method for early cancer diagnostics in addition to ultrasonic and x-ray imaging.[5]

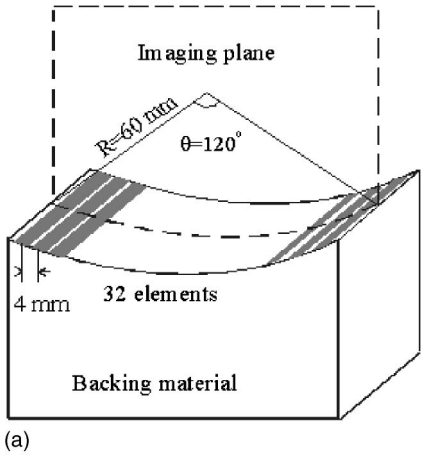
For registration of laser induced ultrasonic transients various detection systems were proposed. Beard *et al.*[7] utilized optical interferometry for detection of OA transients. Unfortunately, the sensitivity level was two orders lower than that in case of piezodetection. Piezoelectric array transducer for three-dimensional (3D) optoacoustic tomography was developed by Kruger *et al.*[13] Hoelen *et al.* used a phased array concept for 3D imaging of blood vessels in a scattering medium.[14] The 3D imaging, however, is not capable of real-time operation. For two-dimensional (2D) optoacoustic tomography Oraevsky *et al.* designed a piezoelectric arc-array transducer which is represented in Fig. 1(a).[5] The transducer consisted of 32 linear piezo-detectors imposed on a cylindrical surface parallel to the cylinder axis. The aperture angle of the transducer was 120°, the element spacing −4 mm. The *ex vivo* and *in vivo* studies revealed high contrast of images and fairly high in-depth resolution (~1 mm). However, resolution in the direction perpendicular to the imaging plane, determined by the length of piezo-detectors and depth of object location, was relatively low, about 15–20 mm.[5] As in 2D tomography the images are obtained by scanning the imaging plane through the object under study, poor resolution in direction perpendicular to the imaging plane means that the ''thickness'' of the imaging plane is large.

In the present paper another design of an arc-array transducer is proposed. As seen in Fig. 1(b), it consists of the same number of piezo-detectors, however, all the elements are focused, as they are imposed on a cylindrical surface perpendicular to the cylinder axis. The advantage of this design is enhanced resolution in the direction perpendicular to the imaging plane due to the narrow focal zone of each piezo-detector. It means that the imaging plane of the array becomes ''thinner.'' The goal of this paper is to conduct theoretical and experimental investigations of the array properties with relation to its spatial resolution and sensitivity.

Resolution yielded by the array transducer in different directions δX, δY, and δZ, as represented in Fig. 2, is discussed in detail. The in-depth resolution δX defined by the detection bandwidth of the piezo-elements[5] is discussed in Sec. II. The resolution in the direction perpendicular to the imaging plane, δY, can be obtained from the focal zone map of the single focused piezo-detector. The transient response of the detector is studied theoretically and experimentally in Secs. II and IV, correspondingly, and the focal zone is examined in Sec. V. Lateral resolution, δZ, in the imaging plane, which can be obtained from the point spread function of the whole array, is discussed in Sec. V.

[a]Electronic mail: t_khokhlova@mail.ru

 © 2004 Acoustical Society of America



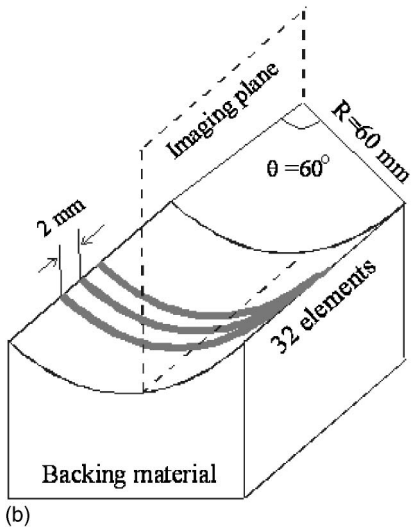


FIG. 1. Previous (a) and innovated (b) design of the arc-array transducer.

II. SIMULATION OF THE TRANSIENT RESPONSE OF THE SINGLE FOCUSED PIEZO-DETECTOR

The parameters of a single wideband element of the array influence the temporal shape profiles of detected OA signals and, consequently, determine the shape and size of the focal region. In order to estimate this influence the transient response of a single focused detector was calculated.

The calculation of the transient response of a shaped piezo-transducer of finite thickness is generally a complicated problem. In our case of a thin film piezo-detector a simplified approach was applied.[15] The approach was based on the assumption that the transformation of the longitudinal wave into the shear waves was negligible. As the detector under study is focused its sensitivity area is localized (see Sec. V). Only sources placed close to the detector focal region yield, therefore, significant level of detected signal. Consequently, in our studies the angle between the normal to the OA wave front and the normal to the detector surface did not exceed 12°. This angle was sufficiently smaller than the critical angle for a longitudinal wave at water–PVDF boundary and resulted in less than 3% variations of the reflection coefficient. Coupling into the shear modes was less than 5%.[16] The results of this approach were in a good agreement with the experiment.

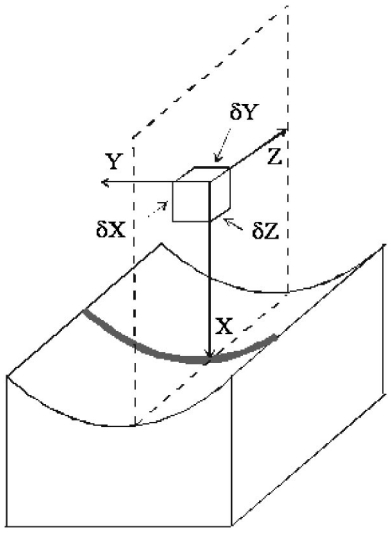


FIG. 2. Coordinate system relative to the transducer. δX, δY, δZ are the resolutions along X, Y, and Z axes correspondingly. The zero point is at the focus line between 16th and 17th detectors.

The simplified approach included two steps. At the first step, the detector bandwidth was considered infinite and the ''impulse response'' [17] was found. This function characterizes the distortion of the pressure profile induced by the finite and curved shape of the detector. It was calculated as the signal temporal profile induced by a point source located arbitrarily in the plane of the detector referred to as the (XY) plane (see Fig. 2). In experiment the source is, of course, not point-like, but distributed. However, it is shown in Sec. II C that OA source of a certain shape can be considered as a point source.

At the second step the spectral sensitivity of the detector was calculated. This characteristic is determined by the piezo-detector thickness, acoustic impedance of the backing, and other specific features of detector design discussed in Sec. II B. Inverse Fourier transform of the spectral sensitivity yields the detector response to a δ pulse propagating perpendicularly to its surface. The spectral sensitivity thus does not include the wave form distortion effect related to the geometry of the detector.

The resulting transient response was then calculated as a convolution of the impulse response and Fourier transform of the spectral sensitivity. The convolution of the transient response with a probe OA signal, calculated theoretically in Sec. II C, is presented in Sec. IV. The resulting output signal is then compared with experimental data.

A. Impulse response of the detector

The calculation of impulse response (IR) in the case of plane polygonal apertures is studied in detail in a series of publications;[17–19] computation optimizing algorithms are proposed in Ref. 19, however, in our case of a cylindrical surface, the IR can be calculated only numerically. The surface of our detector is divided into small elementary surfaces, that can be considered plane within the discretization time interval, therefore, the Rayleigh integral for calculation of impulse response IR(r,t) can be utilized here as in[17]

$$\mathrm{IR}(r,t)=\int_S \frac{\delta\left(t-\dfrac{|r-r'|}{v}\right)}{2\pi|r-r'|}dS', \tag{1}$$

where S is the surface of the transducer, r the radius-vector of the source, and r' the radius-vector of an element of the surface S.

As our detector has a cylindrical symmetry, the calculations of IR are performed in cylindrical coordinate system with the center at the focus of the detector. The radius-vector of the point source and the surface element have coordinates (r_0,φ_0,z_0) and (R_0,φ,z) correspondingly. Here R_0 is the curvature radius of the element, φ varies within the aperture angle φ_0 of the element, and z varies within the width of the element $h/2$. The Rayleigh integral equation (1) can be rewritten in these notations as

$$\begin{aligned}&\mathrm{IR}(t,r_0,\varphi_0,z_0)\\&=\int_{-\varphi_0}^{\varphi_0}d\varphi\int_{-h/2}^{h/2}dz\,R_0\\&\quad\times\frac{\delta\left(t-\dfrac{\sqrt{r_0^2+R_0^2+(z-z_0)^2-2R_0r_0\cos(\varphi-\varphi_0)}}{v}\right)}{2\pi\sqrt{r_0^2+R_0^2+(z-z_0)^2-2R_0^2r_0\cos(\varphi-\varphi_0)}}.\end{aligned} \tag{2}$$

The integral was calculated numerically.

B. Spectral sensitivity of the detector

Relative spectral sensitivity of a film piezo-detector operating in open circuit mode at the frequency f is proportional[6] to the displacement of the particles at the boundaries of the detector:

$$S(f)=A\int_0^d\frac{\partial\eta}{\partial x}dx. \tag{3}$$

Here d is the thickness of the piezo-detector, η is the displacement of particles, and A is a constant depending on longitudinal piezo-electric modulus and permitivity of the film.

Spectral sensitivity $S(f)$ was calculated using modified matrix method.[20] This method utilizes the continuity conditions for all harmonics of oscillation velocity and pressure at the boundaries of the media and allows one to obtain the amplitudes of oscillation velocity and pressure within each layer in a layered structure. In our case the piezo-detector was considered a three-layered structure: a PVDF film placed between infinitely thick layers of immersion liquid and damping material (PMMA). The set of equations for amplitudes of oscillation velocity and pressure is

$$\begin{aligned}&\rho_1c_1(v_1+v_1')=\rho_{2I}c_2(v_2+v_2'),\\&v_1-v_1'=v_2-v_2',\\&\rho_2c_2(v_2\exp(-ikd)+v_2'\exp(ikd))=\rho_{3I}c_3v_3,\\&v_2\exp(-ikd)-v_2'\exp(ikd)=v_3,\end{aligned} \tag{4}$$

where $\rho_{1,2,3}$ are the densities of water, PVDF, and PMMA correspondingly, $c_{1,2,3}$ the sound wave velocities, d is the thickness of PVDF layer, and $v_{1,2,3}$, $v'_{1,2,3}$ are oscillation velocities of a given harmonic f in the waves propagating in opposite directions. Spectral sensitivity, thus, can be rewritten as follows:

$$S(f)=A\frac{i2z_2c_2/2\pi f(2z_2-i2z_2\sin(2\pi fd/c_2)-2z_3\cos(2\pi fd/c_2))}{(z_1+z_2)(z_2+z_3)\exp(i2\pi fd/c_2)+(z_1-z_2)(z_2-z_3)\exp(-i2\pi fd/c_2)}, \tag{5}$$

where z_1, z_2, and z_3 are acoustic impedances of water, PVDF film, and PMMA correspondingly, and c_2 is the acoustic wave velocity in PVDF. The resulting transient response of the detector was obtained as a convolution of inverse Fourier transform of the spectral sensitivity and impulse response of the detector.

The bandwidth of the detector determines the in-depth resolution[6] $\delta X\sim v/f_{\mathrm{res}}$, where f_{res} is the upper limit of the bandwidth: the wider the bandwidth, the better the resolution. On the other hand, the increase of the bandwidth leads to the decrease of the sensitivity[6] and, consequently, to the decrease of the diagnostics depth. Moreover, at higher ultrasound frequencies sound wave attenuation within the tissue is stronger.[14] Taking into account all the preceding arguments, the value of δX value was chosen 0.5 mm. The parameters of the detector were chosen so that the corresponding value of $f_{\mathrm{res}}=3$ MHz was provided.

C. Calculation of the probe optoacoustic signal

The generation of OA transients is governed by[1]

$$\frac{\partial^2p(r,t)}{\partial t^2}-v^2\Delta p(r,t)=\Gamma\frac{\partial}{\partial t}S(r,t), \tag{6}$$

where p and v are the acoustic pressure and the acoustic wave velocity disturbances, correspondingly, Γ is the efficiency of OA excitation,[1] $S(r,t)=f(t)Q(r)$ is the heat source function, $f(t)$ is the temporal profile of laser pulse intensity envelope, and $Q(r)$ is the distribution of the OA sources.

The laser pulse was considered infinitely short, therefore, OA sources could be considered phased-in.[21] The OA transient thus is determined by superposition of coherent signals, generated by the sources distributed in medium. The wave front of the produced signal was considered spherical,

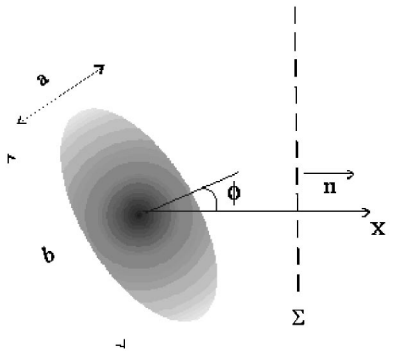


FIG. 3. Excitation of optoacoustic transient by an object with Gaussian distribution of OA sources.

as the distance at which the signals were detected in experiments was much longer than the diffraction length. The acoustic pressure then can be represented as

$$p(r,t)=\frac{1}{4\pi}\frac{\partial}{\partial t}\int_V \frac{\Gamma Q(r')}{|r-r'|}\,\delta\left(t-\frac{|r-r'|}{v}\right)dV'. \tag{7}$$

Here $r-r'=\xi n$ is a vector in the direction from the source to the point of observation. Elementary volume dV may be rewritten as $dV=d\xi d\Sigma$, where $d\Sigma$ is an elementary surface orthogonal to the vector n. Integration of Eq. (7) over ξ yields

$$p(r,t)=\frac{\Gamma}{4\pi t}\frac{\partial}{\partial t}\int Q(r-vtn)d\Sigma. \tag{8}$$

As is seen from Eq. (8), the OA signal is proportional to the area of cross section of OA source distribution and the sphere of radius vt centered at the observation point. At a large distance from the observation point the spheres Σ can be replaced by planes[13] as shown in Fig. 3.

In order to integrate Eq. (8) analytically the in-depth distribution of the OA sources was assumed to be a Gaussian function: $Q(x)=\exp(-x^2/a^2)$, where parameter a characterizes the size of the distribution. The exact shape of the in-depth distribution in the source with $\mu_a=10^3$ cm^{-1} that was used in the experiments could not be resolved by our detection system. As stated in Sec. II B, the detector bandwidth was limited by 3 MHz, which corresponds to the in-depth resolution $\delta X\sim v/f_{\mathrm{res}}\sim 0.5$ mm.

In the plane (ZY) that is perpendicular to the laser beam the OA sources distribution was determined by power distribution across the laser beam: $Q=Q_0\exp(-(z^2+y^2)/b^2)$. Considering all the preceding arguments, $Q(r)$ could be represented as follows:

$$Q=Q_0\exp\left[-\frac{x^2}{a^2}-\frac{y^2+z^2}{b^2}\right]. \tag{9}$$

The size of the in-depth distribution is assumed to be smaller than the radius of the beam, therefore, $a<b$.

The resulting OA signal can be expressed as follows:

$$p=\left(\frac{Q_0\Gamma v^3}{2}\right)\frac{\varepsilon}{(s^2+\varepsilon^2c^2)}\frac{\Theta}{X}\exp\left[-\frac{\Theta^2}{\Theta_0^2}\right]. \tag{10}$$

Here $\Theta=(vt-x)/a$ is the normalized retarded time, $X=x/a$ is the normalized coordinate, $\varepsilon=a/b\ll 1$, $s=\sin\phi$, $c=\cos\phi$, where ϕ is the azimuthal angle shown in Fig. 3,

$$\Theta_0=\sqrt{\frac{s^2+\varepsilon^2c^2}{\varepsilon^2}}$$

characterizes the duration of OA transient.

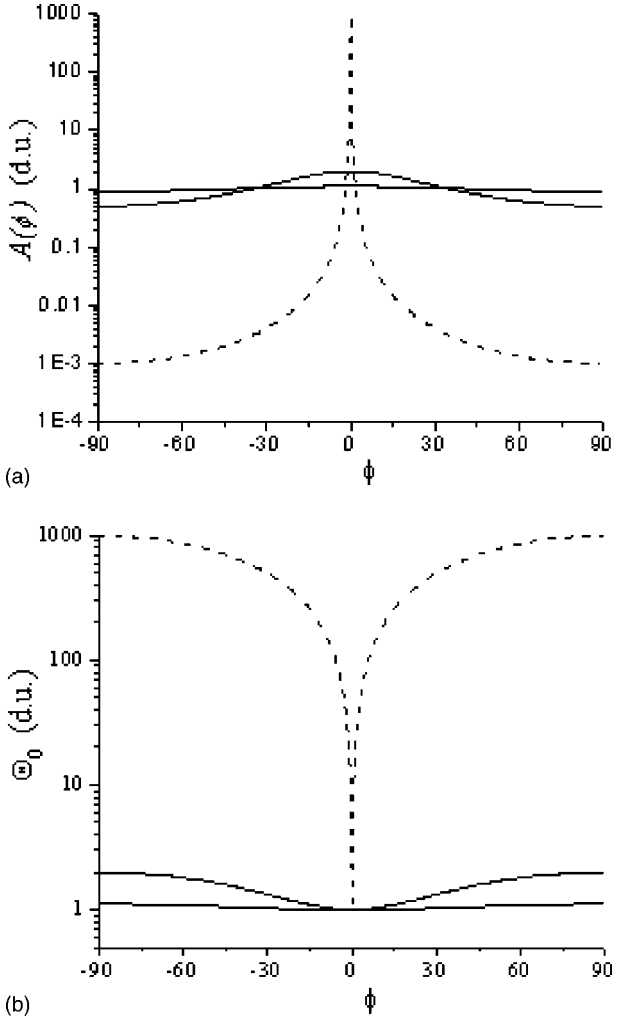


FIG. 4. Theoretically calculated angular distribution of amplitude A (a) and duration Θ_0 (b) of the OA transient produced by an object with Gaussian distribution of the OA sources.

The temporal profile of the OA transient is the time derivative of the Gaussian function with the amplitude proportional to parameter A: $A=\varepsilon/(s^2+\varepsilon^2c^2)$, and duration proportional to parameter Θ_0. These functions depend on the angle ϕ and determine the directivity pattern of the OA source. Figures 4(a) and 4(b) represent normalized amplitude A and duration Θ_0 for various values of object sizes ratio ε. If $\varepsilon\approx 1$, which corresponds to the spherical source distribution, the represented dependencies become uniform, which means that a spherical wave is produced. For $\varepsilon<1$ the amplitude angular distribution reaches its maximum, i.e., the produced wave is close to a plane wave. It is seen that the smaller the ε value, the sharper the maximum.

The width of the radiation pattern defined at 50% level can be determined from Eq. (10):

$$\Delta\varphi=2\arccos\left(\sqrt{\frac{2\varepsilon^2-1}{\varepsilon^2-1}}\right). \tag{11}$$

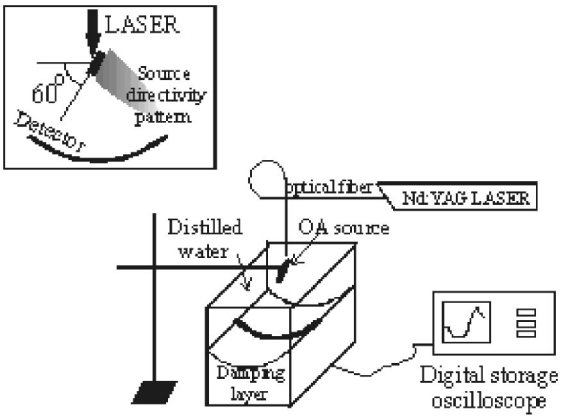


FIG. 5. Diagram of the experimental setup.

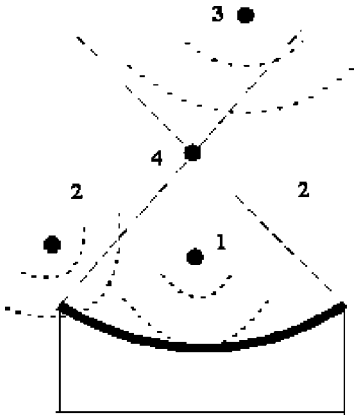


FIG. 6. Areas of possible source positions that determine characteristic features of detected signal.

III. EXPERIMENTAL SETUP

Diagram of the experimental setup is shown in Fig. 5. A diode-pumped *Q*-switched Nd:YAG laser (LCS-DTL-324Q, LaserCompact Co Ltd, Russia) with pulse duration $\tau_L = 8$ ns and pulse repetition rate 500 Hz was employed for excitation of OA signals. Laser pulse was delivered to the OA source by an optical fiber. As for the source itself, it should be as close to a point as possible to justify the theoretical approach discussed in Sec. II. The point source is characterized by a uniform directivity pattern. A small absorbing sphere would look ideal for this purpose. However, it would be hard to provide uniform irradiation conditions for the sphere and it would be difficult to handle. On the other hand, the directivity pattern of a thin flat source can be considered quasiuniform over a wide interval of angles as seen in Fig. 4 ($\varepsilon = 1 \times 10^{-3}$, angle range 30°–90°). Therefore, we used a 20-μm-thick black polyethylene film with absorption coefficient $\mu_a \approx 1000\ \mathrm{cm}^{-1}$ and inclined it at the angle 60 relative to the plane of the detector as shown in Fig. 5. Laser beam at the surface of the polyethylene film was nearly 1 mm in diameter and the flux of laser radiation was 7–8 $\mathrm{mJ/cm}^2$. The source was placed into the immersion liquid of distilled water.

The OA signal was registered by the array consisting of 32 PVDF piezo-detectors that were imposed on a cylindrical surface of 120 mm in diameter. Two arrays with different aperture angles of 30° and 60° were investigated. The angular step between two neighboring detectors was ≈2. The detector bandwidth defined from Eq. (5) at 1/*e* level yields $f_{\mathrm{res}} \approx 3{,}5$ MHz.

Signals from the channels were read consecutively. The received electrical signal was amplified and registered (with 128 samples average) with a digital storage oscilloscope (Tektronix TDS220) on-line with a PC. All detected signals were normalized to the amplitude of the signal from the source placed at the detector focus.

The low frequency sensitivity of the piezo-detector was measured using the acoustic nonlinearity method,[22] the source being located at the focus. The result was 8 μV/Pa. The noise voltage was calculated according to Nyquist formula: it was 13 and 18 μV for 60° and 30° detector correspondingly. Therefore, minimally detectable pressures were $p_{\min 60°} = 3.3$ Pa, $p_{\min 30°} = 4.5$ Pa.[6]

IV. RESULTS

Figure 7 represents the temporal profiles of experimentally obtained OA signals for different source locations in (*XY*) plane (solid lines) and theoretically calculated signals (dotted line). The impulse responses that correspond to the same source locations are represented in the incuts in order to analyze how the curved geometry influences the shape of the OA signal.

According to the characteristic features of impulse responses obtained for various source locations, the (*XY*) plane can be divided into three areas shown in Fig. 6. Areas 1 and 3 are bounded by the piezo-detector aperture angle and lie before and behind the focus, correspondingly. Area 2 lies outside the aperture angle of the detector, and area 4 is the detector focus.

The measured OA signal coming from the detector focus, referred to as area 4, is represented in Fig. 7(a). In this case the signal duration is minimal, as the wave front reaches all the points of the detector simultaneously. Therefore, IR is close to a δ function and its influence on the OA signal temporal profile is minimal. The signal itself consists of a compression phase followed by a rarefaction phase. The latter is ''stretched'' due to the limited bandwidth of the detector.

The OA signal arriving from area 1 is represented in Fig. 7(b). In this case the amplitude of the compression phase is higher than that of the rarefaction phase. This can be explained by the IR behavior. The IR has a steep leading edge corresponding to the wave front arrival to the detector area nearest to the source. Then the wave front reaches more distant parts of the detector that corresponds to the lower trailing edge of the IR.

For area 3 the situation is reversed. The amplitude of the rarefaction phase exceeds the one of the compression phase shown in Fig. 7(c). In this case the wave reaches the edges of the detector first and then propagates along its surface toward the center.

Figure 7(d) shows, that if the OA source is located in area 2, the signal amplitude is lower than that for areas 1 and 3. The reason for this is that the signals coming from areas 1 and 3 reach symmetrical parts of the aperture simultaneously unlike the signals coming from area 2. In the latter case the

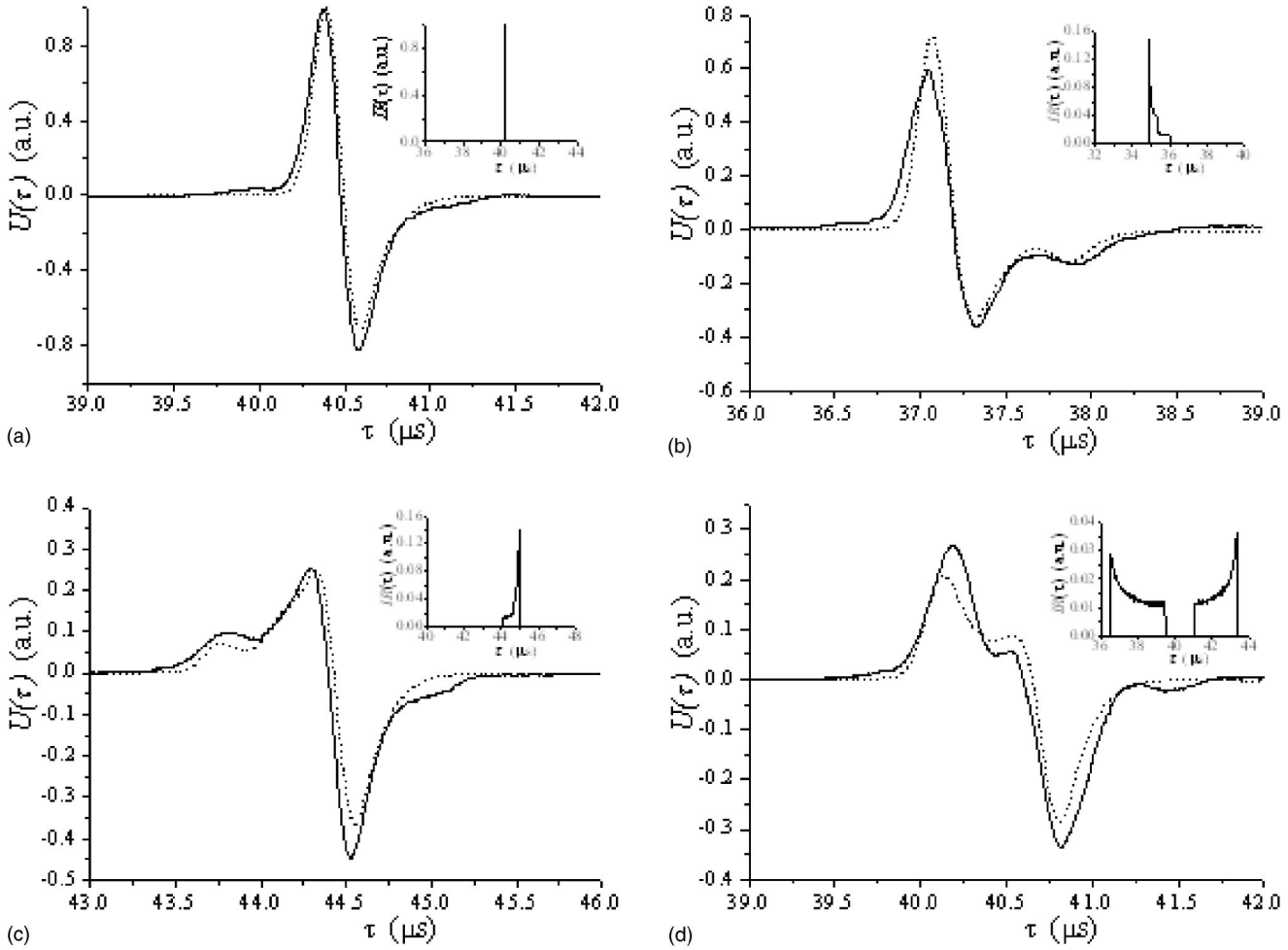


FIG. 7. OA signals registered by a single focused piezo-detector for various positions of the source: (a) the source is placed at the focus (point 4), (b) in area 1, (c) in area 3, (d) in area 2. Solid lines correspond to experimental signals, dotted lines to theoretically calculated ones. The corresponding impulse responses are given in the upper corners.

leading edge of the IR corresponds to the wave front arrival to the nearest edge of the detector; the falling edge occurs when the wave reaches the farthest edge of the detector. Relative levels of the leading and falling edges depend on whether the source is located behind or before the focus.

According to numerically simulated OA signals the ''focal zone maps'' of the detectors with 30° and 60° aperture angles were plotted and are represented in Figs. 8(a) and 9(a). A focal zone map represents amplitudes of the OA signals versus geometric coordinates of the source the signal comes from. The sources are located in (XY) plane, the point $\{x=0,y=0\}$ in Figs. 8(a) and 9(a) corresponds to the focus of the detector. The detected OA transients coming from each point of the focal zone were calculated as the convolution of the impulse response, transient response, and the probe OA signal derived in Sec. II. The map was assumed symmetrical relative to the (XZ) and the (XY) planes. Before plotting the focal zone map all the signals were integrated over time in order to reconstruct the OA source distribution, according to Eq. (10). After that the signal amplitudes were normalized to the maximum amplitude.

The resolution δY (shown in Fig. 2) the imaging system can provide can be determined from the focal zone map. This statement needs further explanation. In case of a plane detector the incident OA signals are integrated over the detector aperture in both lateral directions (Z and Y). Therefore, it is not possible to state at which point of the sensitivity area the OA source is located. In our case of a focused transducer the sensitivity area is localized, therefore, the Y size of the region of maximum detected amplitude can be considered as δY.

The region of maximum detected amplitude is stretched

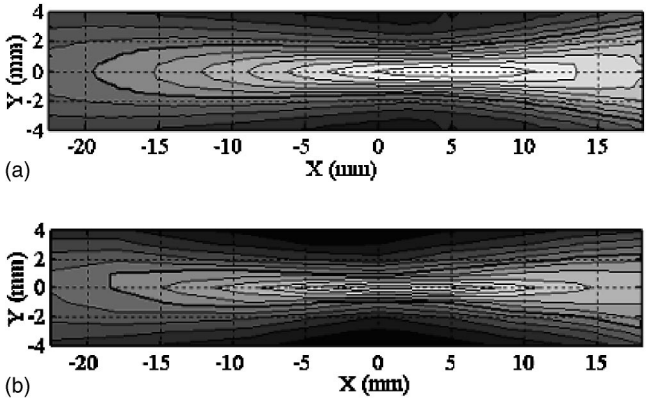


FIG. 8. Numerically simulated (a) and experimental (b) focal zone maps for piezo-detectors with 30° aperture angle.

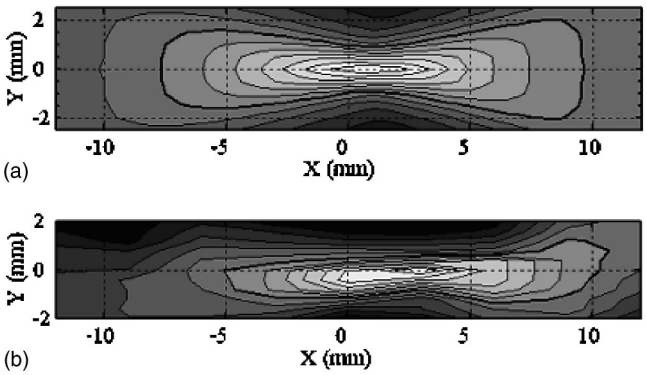


FIG. 9. Numerically simulated (a) and experimental (b) focal zone maps for piezo-detectors with 60° aperture angle.

along the X axis and its size in this direction will be referred to as the ''monitoring depth'' L. This parameter characterizes the depth at which the amplitudes of detected signals are still significant (at 50% level).

Experimentally obtained focal zone maps for detectors with 30° and 60° aperture angles are represented in Figs. 8(b) and 9(b) ($\{x=0,y=0\}$ corresponds to the focus of the detector).

Experimental values of δX and δZ were determined from the array point spread function (PSF). PSF is an image of a point source provided by an array transducer. The size of the image voxel should be smaller than the best resolution of the imaging system (δX and δZ).

For investigation of the array PSF the OA signals from all 32 detectors were collected and then the image of the source was reconstructed using backprojection algorithm. PSFs corresponding to different source locations are represented in Fig. 10.

V. DISCUSSION

A. Focal zone maps

One of the main features of the focused design of the array transducer is enhanced resolution δY in a direction that is perpendicular to the imaging plane. δY is related to the focal zone map of the detector, therefore, it is important to investigate the map in detail. The numerically modeled maps for detectors with 30° and 60° aperture angles are presented in Figs. 8(a) and 9(a). The wider the aperture angle, the better the resolution: $\delta Y_{60°}=1.5$ mm and $\delta Y_{30°}=3.4$ mm correspondingly. On the other hand, the increase of the aperture angle leads to a decrease in the monitoring depth L. For the transducer with aperture angle of 60° the monitoring depth $L_{60}\approx 17$ mm; for 30° transducer L was significantly larger; it could not be measured because the focal zone itself extended up to the surface of the transducer and the region of 50% level of the maximum amplitude region could not be bordered.

The monitoring depth determines the size of maximum sensitivity area and therefore the size of the imaging frame in the X direction. The smaller the frame, the more time it takes to scan through the object. Therefore, a trade-off in terms of the aperture angle must be obtained between the monitoring depth and the resolution δY.

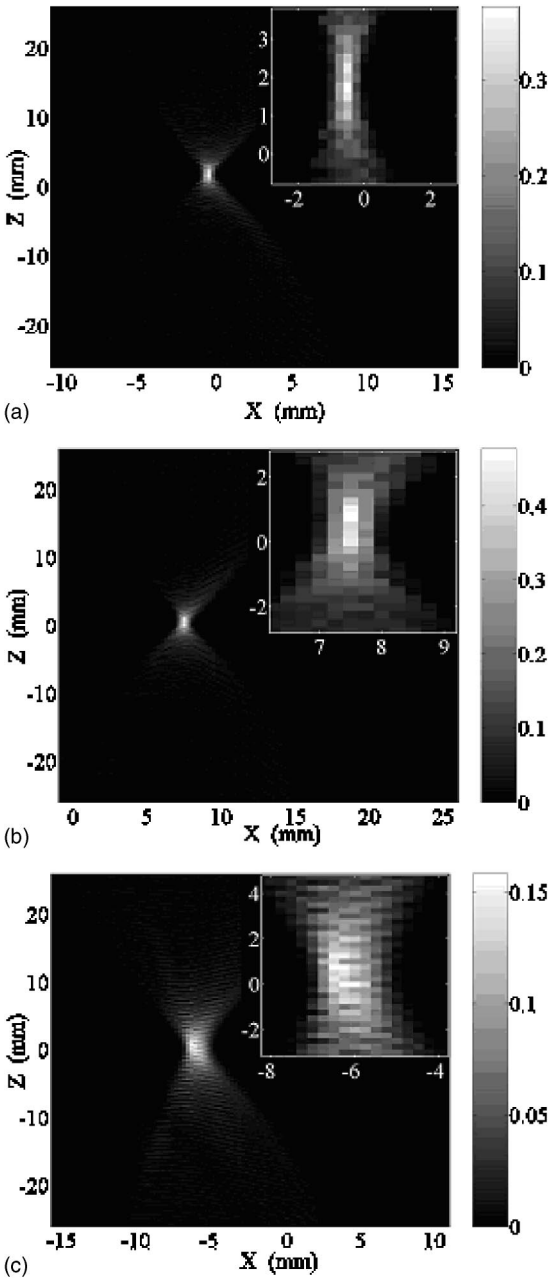


FIG. 10. Experimental point spread function for the array transducer with 30° aperture angle for different positions of the point source: in (a) it is placed in the middle of the focal line of the transducer, in (b) it is 7 mm closer to the array, and in (c) it is 7 mm farther.

Experimental results for focal zone maps of the detectors shown in Figs. 8(b) and 9(b) confirm the results of numerical modeling. However, the experimental value of δY_{30} is 2 mm, which is 1.5 times better than the value obtained from numerical calculations. This disagreement might be related to the fact that the radiation pattern of the OA sources differed from the isotropic one, as we assumed in numerical modeling. According to Eq. (11), the amplitude decreases about three times within the detector aperture angle.

Another characteristic detail of the represented maps is that the area of maximum sensitivity is shifted toward the detector for smaller aperture angles. This shift does not exceed 2 mm for a detector with aperture angle of 60°, however for a 30° detector it is already 6 mm. The same phe-

nomenon is well known for harmonic beam focusing:[23] due to diffraction the actual focus is closer to the transducer than its geometric focus.

B. Point spread function of the array transducer

The resolutions δX and δZ can be determined from the point spread function of the array transducer. In present investigations δX was assigned 0.5 mm, therefore, the pixel size was chosen 0.25 mm$\times$0.25 mm.

In experimental PSF studies the same experimental setup was used (see Fig. 5). The source was located in the imaging plane (XZ) between the 16th and 17th detectors, close to the focal zone where the sensitivity of the array reaches its maximum. The reconstructed image of the source placed exactly at the focus line of the array is presented in Fig. 10(a). The size of the reconstructed source was defined as the size of the area characterized by maximum image brightness, bounded at half maximum level. As seen in the zoomed picture the size of the reconstructed object along the X axis is 0.5 mm, which corresponds to the real size of the source and to the spatial resolution δX. Along the Z axis it is 2 mm, which corresponds to the array angular step of 2°.[24]

The 7 mm shift of the source along the X axis toward the transducer leads to a change in the reconstructed image dimensions. According to Fig. 10(b), δX becomes 0.75 mm, $\delta Z-3$ mm. If the OA source is shifted 7 mm in the opposite direction, δX becomes 1.5 mm and δZ becomes 3 mm [Fig. 10(c)]. At the same time, the maximum amplitude of the signal decreases almost three times. The changes of the image size and brightness are related to the features of acoustic transients temporal profiles described in Sec. IV. If the source is located at the transducer focus, the signal arrives at all its points simultaneously and the duration of the detected signal will be the shortest. With a shift of the source from the focus the OA transient stretches in time, because of asynchronous arrival of a pressure wave at different points of the detector. One more detail of the presented images should be mentioned: the locations of the reconstructed images are shifted $\approx$0.5 mm along the X axis relative to their real position. A possible reason for this disagreement is slight deviations (about 1%) of the real array geometric parameters from those used in the calculations.

The main factors that limit the quality of the reconstructed image are finite width and insufficient number of piezo-detectors. However, narrower detectors have smaller capacitance, therefore their signal-to-noise ratio is also smaller.[6] The increase of detectors number will, on the one hand, decrease the artifact level. On the other hand, a large number of detectors will lead to serious technical difficulties.

As for the time of image reconstruction, the backprojection algorithm alone took about 0.5 s to perform for 32 channels and Pentium IV (2.6 GHz). As the signals from channels were read consecutively, data acquisition time was quite long. However, according to our estimations with the use of multichannel ADC this time can be reduced to 0.5 s.

VI. CONCLUSIONS

A new design of an array transducer for optoacoustic imaging was proposed. The characteristics of the array were studied both theoretically and experimentally.

Transient response of a single focused piezo-detector was studied both theoretically and experimentally. A new simplified approach was applied for the transient response calculation. The results of the calculations were in a good agreement with experimental data that validates the proposed approach.

For estimation of a single detector bandwidth its spectral sensitivity was calculated and yielded the bandwidth equal to 3.5 MHz. It was sufficient to achieve the in-depth resolution $\delta X\approx 0.5$ mm. The measured low-frequency sensitivity of a single detector was about 8 μV/Pa.

The main feature of the proposed array design is enhanced resolution δY in a direction perpendicular to the imaging plane. For determination of δY the focal zone maps for detectors with 30° and 60° aperture angles were plotted in terms of experimentally obtained integrated OA signals. According to these studies, a minimal value of resolution in this direction equal to $\delta Y_{30}=3.4$ mm $\delta Y_{60}=1.5$ mm for detectors with 30° and 60° aperture angles correspondingly. These values are in a good agreement with the results of numerical modeling.

The experimental values of the axial and in-depth resolution in the imaging plane δZ and δX were determined from the point spread function of the array. These values depended on the source location relative to the focal line of the array. Minimal values of δX and δZ were 0.5 and 2 mm.

The useful imaging area of the array was determined. In the X direction it was limited by the monitoring depth L and for a 60° array yielded 17 mm. In the Z direction the size of the useful imaging area is about 50 mm, which is determined by the number of detectors and the first critical angle for a longitudinal wave at water–PMMA boundary.

ACKNOWLEDGMENT

The work was supported by the Russian Foundation for Fundamental Research (RFBR).

[1] V. E. Gusev and A. A. Karabutov, *Laser Optoacoustics* (AIP, New York, 1993).
[2] G. J. Diebold, A. C. Beveridge, and T. J. Hamilton, ''The photoacoustic effect generated by an incompressible sphere,'' J. Acoust. Soc. Am. **112**, 1780–1786 (2002).
[3] G. Paltauf, H. Schmidt-Kloiber, and M. Frenz, ''Photoacoustic waves excited in liquids by fiber-transmitted laser pulses,'' J. Acoust. Soc. Am. **104**, 890–897 (1998).
[4] J. A. Viator and S. A. Prahl, ''Photoacoustic imaging of gelatin phantoms using matched field processing,'' Proc. SPIE **3601**, 276–283 (1999).
[5] V. G. Andreev, A. A. Karabutov, S. V. Solomatin, E. V. Savateeva, V. L. Aleynikov, Y. V. Zhulina, R. D. Fleming, and A. A. Oraevsky, ''Optoacoustic tomography of breast cancer with arc-array transducer,'' in *Biomedical Optoacoustics*, edited by A. A. Oraevsky, [Proc. SPIE **3916**, 36–47 (2000)].
[6] A. Oraevsky and A. Karabutov, ''Ultimate sensitivity of time-resolved opto-acoustic detection,'' *Conference on Biomedical Optoacoustics* [Proc. SPIE **3916**, 228–239 (2000)].
[7] P. C. Beard and T. N. Mills, ''2D line-scan photoacoustic imaging of absorbers in a scattering tissue phantom,'' Proc. SPIE BIOS **4256**, 34–42 (2001).

[8] K. P. Köstli, M. Frenz, H. Bebie, and H. P. Weber, ''Temporal backward projection of optoacoustic pressure transients using Fourier transform methods,'' Phys. Med. Biol. **46**, 1863–1872 (2001).
[9] Y. Xu and L. V. Wang, ''Signal processing in scanning thermoacoustic tomography in biological tissues,'' Med. Phys. **28**, 1519–1524 (2001).
[10] D. W. Blodgett, ''Applications of laser-based ultrasonics to the characterization of the internal structure of teeth,'' J. Acoust. Soc. Am. **114**, 542–549 (2003).
[11] J. A. Viator, G. Au, G. Paltauf, S. L. Jaques, S. A. Prahl, H. Ren, Z. Chen, and J. S. Nelson, ''Clinical testing of a photoacoustic probe for port wine stain depth determination,'' Lasers Surg. Med. **30**, 141–148 (2002).
[12] F. A. Duck, *Physical Properties of Tissue. A Comprehensive Reference Book* (Academic, London, 1990).
[13] R. A. Kruger, D. R. Reinecke, and G. A. Kruger, ''Thermoacoustic computed tomography-technical considerations,'' Med. Phys. **26**, 1832–1837 (1999).
[14] C. G. A. Hoelen, F. F. M. de Mul, R. Pongers, and A. Dekker, ''Three-dimensional photoacoustic imaging of blood vessels in tissue,'' Opt. Lett. **23**, 648–650 (1998).
[15] K. P. Kostli and P. C. Beard, ''Two-dimensional photoacoustic imaging by use of Fourier transform image reconstruction and a detector with an anisotropic response,'' Appl. Opt. **42**, 1899–1908 (2003).
[16] A. Biggs, *Acoustic Microscopy* (Claredon, Oxford, 1992).
[17] V. G. Andreev, A. A. Karabutov, A. E. Ponomaryov, and A. A. Oraevsky, ''Detection of optoacoustic transients with a rectangular transducer of finite dimension,'' in *Biomedical Optoacoustics III*, edited by A. A. Oraevsky [Proc. SPIE **4618**, 153–162 (2002)].
[18] P. R. Stepanishen, ''Acoustic transients from planar axisymmetric vibrators using the impulse function,'' J. Acoust. Soc. Am. **70**, 1176–1181 (1981).
[19] J. A. Jensen, ''A new calculation procedure for spatial impulse responses in ultrasound,'' J. Acoust. Soc. Am. **105**, 3266–3274 (1999).
[20] A. A. Karabutov, V. V. Kozhushko, I. M. Pelivanov, and G. S. Mityurich, ''Interference of opposing longitudinal acoustic waves in an isotropic absorbing plate and a periodic structure with defects,'' Acoust. Phys. **47**, 721–726 (2001).
[21] G. Kino, *Acoustic Waves: Devices, Imaging, and Analog Signal Processing*, Prentice-Hall Signal Processing Series (Prentice–Hall, Englewood Cliffs, NJ, 1987).
[22] V. G. Andreev, A. A. Karabutov, and O. V. Rudenko, ''Method for calibration of the wide band hydrophones in ultrasonic beams of finite amplitude,'' Moscow University Physics Bulletin, Vol. 39, pp. 88–91.
[23] M. B. Vinogradova, O. P. Rudenko, and A. P. Sukhorukov, *Theory of Waves*, 2nd ed. (Nauka, Moscow, 1990).
[24] V. G. Andreev, D. A. Popov, D. V. Sushko, A. A. Karabutov, and A. A. Oraevsky, ''Inverse radon transformation for optoacoustic imaging,'' in *Biomedical Optoacoustics II*, edited by A. A. Oraevsky [Proc. SPIE **4618**, 137–145 (2002)].

Open cycle traveling wave thermoacoustics: Energy fluxes and thermodynamics

Nathan T. Weiland[a)]
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Ben T. Zinn
School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 2 November 2003; revised 25 May 2004; accepted 1 June 2004)

In an open cycle traveling wave thermoacoustic engine, hot gas steadily flows into the hot side of the regenerator, replacing the hot heat exchanger as the primary energy source for the engine. In such an engine, interactions between the acoustic, convective, conductive and thermoacoustic energy fluxes facilitate conversion of the input thermal energy into acoustic energy. This study describes the energy flux interactions throughout the engine, thus clarifying the important role of the mean temperature difference that exists between the mean flow of hot gas and the hot-side regenerator interface in the open cycle engine. Furthermore, this study derives an optimal regenerator interface temperature that maximizes the acoustic power output of the engine for a given thermal energy input. The acoustic power output and thermal efficiency of the open cycle engine are compared to those in a closed cycle engine in which a crossflow heat exchanger is used to supply the required heat input. By accounting for the effectiveness of the heat exchanger, it is shown that the open cycle has the potential to achieve higher efficiencies than the closed cycle in converting the thermal energy in a stream of gas into acoustic energy. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1775277]

PACS numbers: 43.35.Ud, 44.27.+g [RR] Pages: 1507–1517

I. INTRODUCTION

Most thermoacoustics research during the past decade has primarily focused on closed cycle refrigeration and engine systems. While open cycle thermoacoustic systems have received limited attention, there are indications that many typical thermoacoustic applications could benefit from the use of open cycle systems. For example, the superposition of a slow mean flow on the acoustic field in an open cycle device allows for the elimination of one of the heat exchangers used to set up the temperature difference across the stack or regenerator, thereby eliminating its cost, complexity and inefficiency.[1,2] Furthermore, allowing for the passage of a fluid through a thermoacoustic engine opens up the possibility of creating an engine that employs a combustion process, either inside or outside the engine, to generate a stream of hot combustion products that passes through the stack or regenerator to provide the engine's thermal energy input.

In designing an open cycle thermoacoustic device, the utilization of standing wave thermoacoustics is appealing because the inlet and exhaust ports can be located at the pressure nodes in order to minimize the loss of acoustic energy through these ports.[2] However, standing wave thermoacoustic systems rely on irreversible processes for their operation, thus they are inherently less efficient than their traveling wave counterparts.[3] Although the design of an open cycle traveling wave thermoacoustic engine is more challenging, such an engine is, in principle, the most efficient means of converting the thermal energy in a stream of gas into acoustic energy. In support of this claim, the following study compares the energy fluxes and ideal performance indices of both open and closed cycle traveling wave thermoacoustic engines.

Previous work in open cycle thermoacoustics is primarily linked to the standing wave thermoacoustic refrigerator of Reid and Swift,[1,2] though much of the theory developed for this application is not applicable to traveling wave thermoacoustics. Many studies on the effects of acoustic streaming in thermoacoustic devices are related to open cycle thermoacoustics, however these studies do not focus as much on mean flow energy fluxes as they do on the various sources of acoustic streaming, which can generate a mean mass flux in a closed cycle thermoacoustic device.

Gedeon first showed[4] that the mean mass flux flowing through the regenerator in a traveling wave thermoacoustic device is equal to the sum of a second order mean pressure term and an acoustic streaming term, which is generally an unavoidable byproduct of the acoustic energy flux through the regenerator. In a thermoacoustic device with a feedback path, the second order mean pressure gradient is generally small or nonexistent, and the acoustic streaming causes a mean mass flux to travel through the feedback path in the direction of the acoustic energy flux.[4] This case has been recently studied, both theoretically[5] and experimentally,[6,7] although these studies do not consider effects that might counter the mean mass flux due to acoustic streaming. In many thermoacoustic devices, the mean mass flux is forced to be zero, either by utilizing a geometry that does not include a feedback path,[8,9] or by incorporating other means to suppress this mass flux.[3,10,11] In these cases, a second order time average pressure difference, $\Delta p_{2,0}$, develops across the regenerator, which is accompanied by a mean second order

a)Electronic mail: gte852f@mail.gatech.edu

velocity $u_{2,0}$, that counteracts the effect of the acoustic streaming term in generating a mean mass flux through the regenerator.

Open cycle traveling wave thermoacoustic devices differ from either of these two cases, as they include a mean mass flux that travels in a direction opposite to the acoustic energy flux direction. However, open cycle devices can still be analyzed using Waxler's[12] general relationships between the mean mass flux and the pressure gradient. More importantly, the open cycle study presented below can also be applied to the analysis of closed cycle traveling wave thermoacoustic devices in which acoustic streaming effects cause a nonzero mean mass flux to flow through the regenerator.

A simple schematic of the open cycle traveling wave thermoacoustic engine that will be studied in this work is shown in Fig. 1. A slow mean flow of hot gas is superimposed on the acoustic field in the device, and approaches the regenerator from an open duct on the right. The origin of this hot gas is not important for the purposes of this study, but could be supplied by a combustion process, for example, originating from either inside or outside the thermoacoustic device. A cold heat exchanger at the other end of the regenerator rejects heat to the surroundings at the ambient temperature, T_0, and creates a temperature gradient across the regenerator. As the mean flow passes into and through the regenerator, some of its thermal energy is converted into acoustic energy, thus amplifying a traveling acoustic wave as it moves from the cold side to the hot side of the regenerator.[13,14]

In a previous study,[15] it has been shown that a substantial difference exists between the mean temperature of the incoming hot gas, T_h, and the temperature of the solid material at the hot end of the regenerator, T_{re}. The magnitude of this temperature difference significantly impacts the various energy fluxes in the engine, thus critically affecting the efficiency and acoustic power output of the engine. To illustrate these effects, this paper first describes the operation of the open cycle thermoacoustic engine by examining the interactions between the engine's various energy fluxes. This examination reveals that the acoustic power output of the engine can be maximized for particular values of the regenerator's hot-side temperature, T_{re}, thus providing expressions for the engine's ideal power output and thermal efficiency. These ideal performance indices are then compared to those produced by a closed cycle traveling wave thermoacoustic engine, in which a heat exchanger is used to transfer the thermal energy from a stream of hot gas to the engine.

II. OPEN CYCLE ENGINE ENERGY FLUXES

In thermoacoustics, knowledge of the relevant energy fluxes in the device is essential to understanding its operation, and can be used here to help understand the importance of the temperature difference at the regenerator's interface. Of particular importance is the fact that convection energy fluxes are present in addition to acoustic energy fluxes, conduction losses, and thermal energy transport by second order entropy fluxes in the open cycle thermoacoustic engine. The importance of these various energy fluxes has been clearly

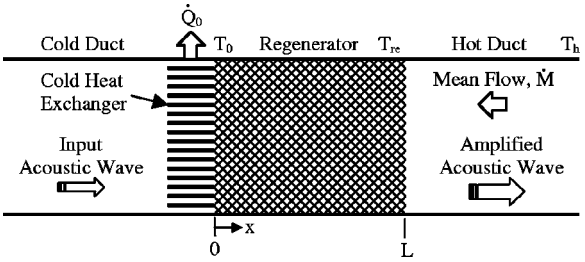


FIG. 1. A schematic of a basic open cycle traveling wave thermoacoustic engine. A steady flow of gas with a mean mass flux $\dot{M}$ and a mean temperature T_h flows into the regenerator, whose hot end is at the temperature T_{re}. A cold heat exchanger rejects heat $\dot{Q}_0$ to the ambient temperature T_0 at the other end of the regenerator. The resulting temperature gradient in the regenerator is used to amplify an acoustic traveling wave that enters the regenerator's cold end and exits its hot end.

illustrated in Swift's book,[16] and the nomenclature and derivations in the following section primarily follow this work.

A. Basic equations

The analysis begins with an expression for the total energy flux in an open cycle thermoacoustic device, Swift's Equation 7.105,[16] which is the sum of the relevant energy fluxes in the device. In deriving this equation, Swift neglects higher order kinetic energy and viscous effects, and assumes that the magnitude of the mean mass flux is a second order quantity. This equation expresses the total second-order energy flux, $\dot{H}$, as

$$\dot{H}=\dot{M}c_p(T_m-T_0)-(Ak+A_sk_s)\frac{dT_m}{dx}+\frac{1}{2}\rho_m c_p\int \mathrm{Re}[T_1\tilde{u}_1]dA. \quad (1)$$

The first term on the right side of Eq. (1) is the convective energy flux, where $\dot{M}$ represents the total mean mass flux in the device, including acoustic streaming, pressure gradient and applied flow effects,[16] c_p is the constant pressure specific heat of the gas, T_m is its mean temperature and T_0 is the ambient reference temperature. The second term on the right side of Eq. (1) is the conductive energy flux, generally a loss term, where A and A_s are the cross-sectional areas of the gas and solid, respectively, and k and k_s are the thermal conductivities of the gas and solid, respectively. The solid components include the regenerator material, the wall surrounding the regenerator, and any other solid materials that are contained within the regenerator. The last term on the right side of Eq. (1) essentially represents the thermoacoustic energy fluxes in the engine, expressed as the time average of the product of T_1, the first order oscillating temperature, and u_1, the first order oscillating axial gas velocity. To ensure that these thermoacoustic effects are the dominant energy fluxes in the engine, this term must be at least as large as the convection and conduction terms in Eq. (1).[16] This generally implies that $\dot{M}\ll\dot{m}_1$ in an open cycle device, i.e., the mean mass flux is much smaller than the oscillating mass flux.

With the ideal gas relationship: $c_pT_1=T_ms_1+p_1/\rho_m$, and the definitions of the acoustic energy flux $\dot{E}$

$= \frac{1}{2}\mathrm{Re}[p_1\tilde{U}_1]$ and the thermoacoustic entropy flux $\dot{S} \equiv \frac{1}{2}\rho_m \int \mathrm{Re}[s_1\tilde{u}_1]dA$, where p_1 is the acoustic pressure, U_1 is the first order oscillating volumetric velocity, ρ_m is the mean gas density, and s_1 is the first order oscillating entropy, the thermoacoustic term in Eq. (1) can be written as

$$\frac{1}{2}\rho_m c_p \int \mathrm{Re}[T_1\tilde{u}_1]dA = \dot{E} + T_m\dot{S}. \tag{2}$$

The second order thermoacoustic entropy flux, $\dot{S}$, is sometimes referred to as a ''thermoacoustic heat flux,'' and is frequently described as a ''bucket brigade'' of gas parcels shuttling heat up or down a temperature gradient.[16]

In open channels and isentropic environments, this second order entropy flux is generally negligible, and Eq. (2) reduces to the acoustic energy flux at that location. Inside a regenerator, however, the second order entropy flux can be fairly large. In an ideal regenerator, perfect solid–gas thermal contact results in an isothermal environment, where $T_1 = 0$. Applying this condition to Eq. (2) yields

$$\dot{S}_{\mathrm{re,ideal}} = -\frac{\dot{E}}{T_m}. \tag{3}$$

This equation expresses the relationship between the ideal entropy and work fluxes in the regenerator, where at any location, a unit of acoustic power traveling to the right is countered by a unit of thermal energy transported to the left by the ideal second order entropy flux. This entropy flux is caused by a combination of the acoustic gas movement and the heat transfer between the solid and gas that accompanies acoustic pressure changes in the regenerator, where the temperature of the gas is ideally constrained to equal the regenerator's local solid temperature.

For imperfect thermal contact in the regenerator, $T_1 \neq 0$, and the last term in Eq. (1) contributes an additional thermoacoustic entropy flux to the total energy flux in the regenerator. Swift has shown that in a parallel plate regenerator, in the limit of very small plate spacing, the last term in Eq. (1) can be approximated to lowest order by[16]

$$\frac{1}{2}\rho_m c_p \int \mathrm{Re}[T_1\tilde{u}_1]dA \approx -\frac{17}{35}\frac{\rho_m c_p}{A\omega}\frac{r_h^2}{\delta_\kappa^2}|U_1|^2\frac{dT_m}{dx} = -\psi\frac{dT_m}{dx}, \tag{4}$$

where ω is the angular frequency of the acoustic oscillations, r_h is the hydraulic radius (half of the regenerator plate spacing), δ_κ is the thermal penetration depth, $\delta_\kappa = \sqrt{2k/\omega\rho_m c_p}$, and

$$\psi \equiv \frac{17}{35}\frac{\rho_m c_p}{A\omega}\frac{r_h^2}{\delta_\kappa^2}|U_1|^2. \tag{5}$$

A nonzero value of ψ expresses the effect of imperfect solid–gas thermal contact in the regenerator. In contrast, perfect solid–gas thermal contact in the regenerator results when the plate spacing or pore size is very small relative to the thermal penetration depth, i.e., $r_h^2/\delta_\kappa^2 \ll 1$, yielding $\psi \approx 0$ in Eq. (5) and $T_1 \approx 0$ in Eq. (4). Note that the particular form of ψ in Eq. (5) only applies to parallel plate regenerators, though similar expressions can be derived for other regenerator geometries. This term, divided by the cross-sectional area, has been referred to as an ''effective'' thermal conductivity in other studies on energy fluxes in closed cycle thermoacoustic devices,[17,18] although this terminology tends to de-emphasize the acoustic nature of this energy flux.

Substituting Eq. (4) into Eq. (2) yields the following expression for the thermoacoustic entropy flux inside the regenerator:

$$\dot{S}_{\mathrm{re}} \approx \frac{-1}{T_m}\left[\dot{E} + \psi\frac{dT_m}{dx}\right] = \dot{S}_{\mathrm{re,ideal}} + \dot{S}_{\mathrm{re,loss}}, \tag{6}$$

where

$$\dot{S}_{\mathrm{re,loss}} \equiv -\frac{\psi}{T_m}\frac{dT_m}{dx} \tag{7}$$

is the entropy flux loss. Equation (6) indicates that the ''bucket brigade'' heat transfer, as described by the second order entropy flux, is composed of both perfect and imperfect regenerator thermal contact components. As will be shown below, $\dot{S}_{\mathrm{re,ideal}}$ carries thermal energy that is used to amplify the acoustic wave in the regenerator, while $\dot{S}_{\mathrm{re,loss}}$ is an unwanted entropy flux that carries thermal energy out of the system without converting it to work in the form of acoustic energy.

B. Energy fluxes at the regenerator/open duct interface

Hot gas at temperature T_h flows towards the regenerator through the adjoining open duct, as shown in Fig. 1. Since the oscillations in the open duct can be assumed to be approximately isentropic and the mean gas temperature there is nearly uniform, the conduction and entropy flux terms can be neglected in Eqs. (1) and (2), respectively, to yield the total energy flux in the open duct:

$$\dot{H}_h = \dot{M}c_p(T_h - T_0) + \dot{E}_h, \tag{8}$$

where the subscript ''h'' denotes conditions at the hot-side open duct.

Denoting the temperature at the hot side of the regenerator as T_{re}, Eqs. (1) and (4) can be used to express the total energy flux at the right side of the regenerator:

$$\dot{H}_{\mathrm{re},h} = \dot{M}c_p(T_{\mathrm{re}} - T_0) - (\psi + Ak + A_s k_s)\frac{dT_m}{dx}. \tag{9}$$

Assuming that the periphery of the engine is well-insulated, the total energy flux remains constant everywhere in the engine except at the cold heat exchanger, where heat is transferred out of the engine at the ambient temperature, T_0. Therefore, equating the total energy fluxes of Eqs. (8) and (9) yields the following expression for the acoustic energy flux in the open duct:

$$\dot{E}_h = \dot{M}c_p(T_{\mathrm{re}} - T_h) - (\psi + Ak + A_s k_s)\frac{dT_m}{dx}. \tag{10}$$

This equation states that the acoustic energy output of the regenerator is a function of the difference in the convected

thermal energy across the regenerator interface, minus the conductive and entropy flux losses down the regenerator. The same relationship was derived by very different means in a previous study,[15] though without the entropy flux loss term, since $\psi=0$ in the perfect thermal contact assumption of that study.

Rearranging Eq. (10) with the help of Eq. (6) and the state relation $dh=c_p\,dT$, where h is the enthalpy of the gas, an alternate description of the energy balance at the regenerator interface is obtained:

$$\dot{M}h_h=\dot{M}h_{\mathrm{re}}+T_{\mathrm{re}}\dot{S}_{\mathrm{re,ideal}}+T_{\mathrm{re}}\dot{S}_{\mathrm{re,loss}}-(Ak+A_sk_s)\frac{dT_m}{dx}. \tag{11}$$

This equation states that the input thermal energy convected to the regenerator interface, $\dot{M}h_h$, is split into four thermal energy fluxes upon entering the regenerator. Three of these thermal energy fluxes are loss mechanisms that transport heat towards the regenerator's cold end without converting it into acoustic energy: the conductive heat flux, the thermoacoustic entropy flux due to imperfect thermal contact in the regenerator, $\dot{S}_{\mathrm{re,loss}}$, and the convective heat flux at lower enthalpy, $\dot{M}h_{\mathrm{re}}$. The remaining thermal energy flux entering the regenerator is due to the ideal entropy flux, $\dot{S}_{\mathrm{re,ideal}}$.

C. The acoustic energy gain in the regenerator

The interface energy balance of Eq. (11) shows that acoustic energy is neither created nor destroyed at the regenerator interface. Rather, it indicates that the ideal entropy flux carries thermal energy that is converted into acoustic energy as it travels down the regenerator's temperature gradient. To demonstrate this point, note first that $\dot{S}_{\mathrm{re,ideal}}$ is a constant if viscous effects are neglected in the regenerator. Although this proof is somewhat tedious and will not be shown here, a constant ideal entropy flux can be used with Eq. (3) to show that the acoustic gain across the regenerator is

$$\dot{E}_h-\dot{E}_c=-(T_{\mathrm{re}}\dot{S}_{\mathrm{re,ideal}}-T_0\dot{S}_{\mathrm{re,ideal}}). \tag{12}$$

This equation, which applies to both ideal and nonideal regenerators, states that the acoustic energy gain is equal to the difference in the thermal energy carried by the ideal entropy fluxes at the hot and cold ends of the regenerator. This is basically a statement of the first law of thermodynamics for an ideal regenerator, where the work produced in the regenerator is equal to its heat input minus its heat output.

Since $\dot{S}_{\mathrm{re,ideal}}$ is a constant in the regenerator, differentiating Eq. (3) with respect to x yields

$$\frac{d\dot{E}}{dx}\approx\frac{\dot{E}}{T_m}\frac{dT_m}{dx}. \tag{13}$$

This expression has been previously derived by other means,[16] and is frequently used to approximate the acoustic energy gain in a regenerator with good solid–gas thermal contact in a traveling wave thermoacoustic device, although it does not include losses from viscous effects. Note that the conductive and entropy flux loss terms have been retained in this analysis because they are functions of the local temperature gradient in the engine, and are needed for estimating the regenerator temperature profile and the acoustic energy output below. On the other hand, the viscous regenerator loss is not a function of the local temperature gradient, and is therefore neglected in this study.

D. The regenerator temperature profile

Since the total energy flux in the regenerator is constant, it follows that $d\dot{H}_{\mathrm{re}}/dx=0$. Substituting Eqs. (1) and (4) into this expression yields, upon rearrangement,

$$\frac{d^2T_m}{dx^2}-\Xi\frac{dT_m}{dx}=0, \tag{14}$$

where the parameter Ξ has been defined as

$$\Xi\equiv\frac{\dot{M}c_p}{\psi+Ak+A_sk_s}. \tag{15}$$

By specifying the temperatures at either end of the regenerator (i.e., $T_m=T_0$ at $x=0$ and $T_m=T_{\mathrm{re}}$ at $x=L$, as depicted in Fig. 1), Eq. (14) can be solved for the mean temperature profile in the regenerator:

$$T_{m,\mathrm{re}}(x)=T_0+(T_{\mathrm{re}}-T_0)\frac{1-e^{\Xi x}}{1-e^{\Xi L}}. \tag{16}$$

In deriving this equation, it has been assumed that Ξ is a constant, which essentially assumes that c_p, k, and k_s are independent of temperature in the regenerator. This is not entirely accurate, but since these variables are not very strong functions of temperature, Eq. (16) still provides a reasonable approximation for the temperature profile in the regenerator. Also, note that Reid and Swift[2] have derived an equation similar to Eq. (16), though their definition of Ξ and the assumptions applied in their derivation are only valid for standing wave thermoacoustic systems.

Once the regenerator's cold and hot end temperatures are known, Eq. (16) allows one to calculate Ξ with a single temperature measurement at any point within the regenerator. This can then be used to find $\dot{M}$ from Eq. (15), and thus could be a valuable means of evaluating the streaming mass flux in a closed cycle traveling wave thermoacoustic system. Such a procedure was first suggested by Gedeon,[4] and was later used by Job *et al.*[6] to determine the streaming mass flux passing through a quasi-adiabatic stack (i.e., where $r_h\approx\delta_\kappa$) in a looped tube. Although the theory Job *et al.* developed to describe the temperature profile includes radial heat loss effects not modeled here, the present work includes thermoacoustic heat pumping effects not accounted for in their analysis.[6]

Finally, note that Eq. (16) can be differentiated with respect to x to evaluate the temperature gradient at the regenerator's hot-side interface. Substituting this result and Eq. (15) into Eq. (10), and assuming that the quantity $(\psi+Ak+A_sk_s)$ is independent of temperature, the following simplified expression for the hot-side acoustic energy flux is attained:

$$\dot{E}_h=\dot{M}c_p\left(T_{\mathrm{re}}-T_h-\frac{(T_0-T_{\mathrm{re}})e^{\Xi L}}{1-e^{\Xi L}}\right). \tag{17}$$

E. Cold side energy fluxes

An energy balance for the cold heat exchanger can be written as

$$\dot{H}_c = \dot{H}_{\mathrm{re},c} + \dot{Q}_0, \tag{18}$$

where $\dot{H}_c$ is the total energy flux in the open duct to the left of the cold heat exchanger and $\dot{Q}_0$ is the heat transferred out of the engine at the cold heat exchanger, as shown in Fig. 1. The total energy flux at the cold side of the regenerator, $\dot{H}_{\mathrm{re},c}$, can be found by setting $T_m = T_0$ in Eqs. (1) and (4), i.e.,

$$\dot{H}_{\mathrm{re},c} = -(\psi + Ak + A_s k_s)\left.\frac{dT_m}{dx}\right|_0. \tag{19}$$

Since the cold-side acoustic energy flux $\dot{H}_c = \dot{E}_c$, Eqs. (3), (18), and (19) can be used to show that

$$\dot{Q}_0 = -T_0 \dot{S}_{\mathrm{re,ideal}} + (\psi + Ak + A_s k_s)\left.\frac{dT_m}{dx}\right|_0. \tag{20}$$

This equation states that in addition to the heat rejection that would occur in an ideal thermoacoustic engine, $-T_0 \dot{S}_{\mathrm{re,ideal}}$, the heat transfer out of the cold heat exchanger also includes the thermal energy losses at the cold end of the regenerator, as described in Eq. (19).

F. Energy flux results

To facilitate the analysis of the energy fluxes in the open cycle thermoacoustic engine, the above equations are cast in a nondimensional form by dividing energy flux terms by $-\dot{M} c_p T_0$, temperatures by the ambient temperature T_0, and lengths by the regenerator length, L. Since this nondimensionalization is fairly straightforward, it will not be performed here, although it is worth noting that all of the dimensionless energy fluxes in the engine are dependent on just three dimensionless parameters: T_h/T_0, T_{re}/T_0, and ΞL. In the following section, it will be shown that an optimal value for T_{re}/T_0 can be computed from T_h/T_0 and ΞL, indicating that only two dimensionless parameters are really needed to compute the dimensionless energy fluxes that maximize the acoustic power output of the engine.

Solving the dimensionless forms of the above equations yields the temperature profile and various energy fluxes shown in Fig. 2(b) and Fig. 2(c), respectively, for representative conditions in a basic open cycle traveling wave thermoacoustic engine, similar to that shown in Fig. 2(a). The temperature plot in Fig. 2(b) shows a significant temperature drop at the regenerator's interface with the hot duct, followed by an exponential temperature profile within the regenerator. The convective energy flux in Fig. 2(c) mirrors this profile, as it is linearly dependent on the local gas temperature.

Figure 2(c) shows that in the hot duct to the right of the regenerator, the total energy flux consists of the sum of the acoustic and convective energy fluxes, as described by Eq. (8), while just inside the regenerator, the total energy flux is equal to the sum of the convective energy flux and the regenerator's thermal energy losses, as expressed in Eq. (9).

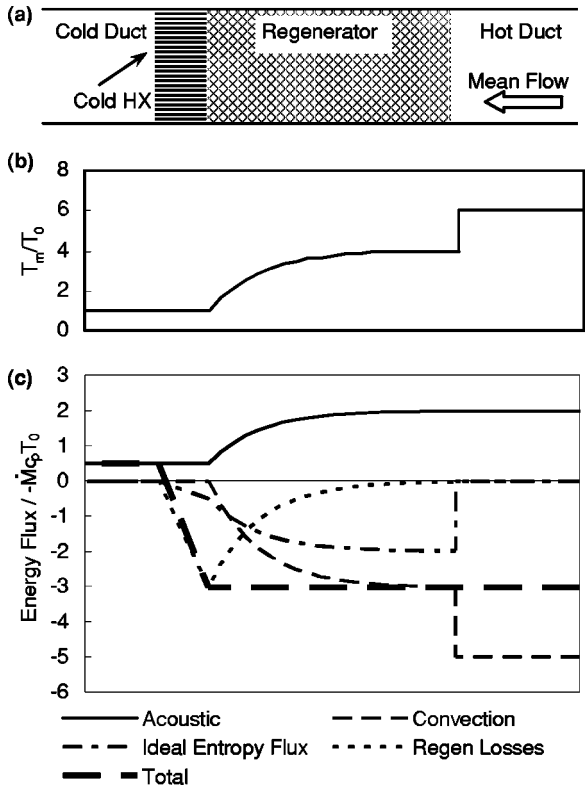


FIG. 2. Plots of (b) the spatial dependence of the dimensionless temperature and (c) the various dimensionless energy fluxes in an open cycle thermoacoustic engine (a), for the representative conditions: $T_h/T_0 = 6$, $T_{\mathrm{re}}/T_0 = 4$, and $\Xi L = -5$. Note that the horizontal scales in (b) and (c) correspond to the horizontal scale of the engine schematic (a).

Note that in addition to these two energy fluxes in Fig. 2(c), there is an acoustic energy flux within the regenerator that is "cancelled" by the thermal energy transport associated with the ideal second order entropy flux, as described in Eqs. (2) and (6).

In Fig. 2(c), the second order entropy flux resulting from imperfect thermal contact in the regenerator, $\dot{S}_{\mathrm{re,loss}}$, has been lumped together with the conductive energy flux to form the "Regen Losses" energy flux, which is proportional to the local temperature gradient in the regenerator. Since the convective energy flux decreases with the temperature in the regenerator, the thermal energy losses must increase as the temperature decreases to maintain a constant total energy flux in the regenerator. As a result, the temperature gradient is low at the regenerator's hot end, and high at its cold end, explaining the exponential shape of the regenerator's temperature profile as expressed in Eq. (16). Unfortunately, this also means that all of the convective thermal energy flux at the hot end of the regenerator, $\dot{M} c_p (T_{\mathrm{re}} - T_0)$, is gradually converted into heat conduction and additional thermoacoustic entropy flux losses as it convects down the regenerator's temperature gradient. This energy, equal to the total energy flux of Eq. (19), is essentially wasted, and leaves the system through the cold heat exchanger, as described in Eq. (20). Since the details of the heat transfer out of the cold heat exchanger are not of interest in this study, the heat fluxes are simply depicted in Fig. 2(c) as linearly decreasing to zero over the length of the cold heat exchanger.

Recovery of the convective thermal energy loss could substantially improve the thermal efficiency of the thermoacoustic engine, without having significant effects on its operation. One method that has been suggested for recovery of this thermal energy[19] involves preheating the reactants of a combustion process that supplies the hot stream of gas to the engine, possibly by routing the pipes carrying the combustion process reactants around or through the regenerator so that some of this waste heat is transferred to the combustion process reactants. Such preheating would increase the combustion process temperature, thereby increasing the input temperature, T_h, and the efficiency of the engine. Equivalently, the preheating could instead be used to reduce the amount of fuel energy required to reach a desired combustion temperature, T_h, thus increasing the fuel energy-to-acoustic energy conversion efficiency. In either case, the transfer of waste heat from the regenerator into the combustion process reactants would result in a more linear temperature profile, thereby reducing the conductive and thermoacoustic entropy fluxes that leave the system through the cold heat exchanger. As the total energy flux through the regenerator is no longer constant in this situation, the analysis becomes more complex, and will be left as a topic for future study.

G. Mean mass flux and pressure gradient

As noted in the Introduction, the flow direction of the mean mass flux in an open cycle thermoacoustic device opposes that of the acoustic energy flux. In contrast, acoustic streaming, in the absence of other effects, produces a mean mass flux that flows in the same direction as the acoustic energy flux,[4] thus an external force is required to cause mass to flow in the desired direction in the open cycle configuration. This external force is provided by imposing a mean pressure gradient across the regenerator, and the magnitude of the mean mass flux is controlled by adjusting the mean inlet and outlet gas pressures on either side of the regenerator. An equation relating the mean mass flux to the pressure difference across a parallel plate regenerator, Δp_2, is given by Waxler.[12] Bailliet *et al.*[8] and Backhaus and Swift[20] note that Waxler's analysis does not include the temperature dependence of viscosity and thermal conductivity, though their analyses also point out that these effects are not important for regenerators in the small plate spacing approximation, $r_h \ll \delta_\kappa$, which applies to the regenerator in a traveling wave thermoacoustic device.

A generalized version of Waxler's regenerator pressure difference equation is[12]

$$\Delta p_2 = -\frac{\alpha \mu_m L}{\rho_m}\dot{M} + \Delta p_2|_{\dot{M}=0}, \tag{21}$$

where μ_m is the gas viscosity evaluated at the regenerator's mean temperature, and the factor α accounts for the geometry of the regenerator (e.g., $\alpha = 3/Ar_h^2$ for a parallel plate geometry). The term $\Delta p_2|_{\dot{M}=0}$ is the pressure difference that results when $\dot{M}=0$, and is primarily generated by acoustic streaming effects in the regenerator. While Waxler derives an expression for $\Delta p_2|_{\dot{M}=0}$ in a parallel plate regenerator,[12] for the order of magnitude calculation presented here it is sufficient to use the more general result of Gedeon,[4] whose studies imply that an acoustic streaming mass flux $\dot{M}_{\mathrm{st}} \approx \rho_0 \dot{E}_c / p_0$ develops when $\Delta p_2 \approx 0$ across the regenerator, where the density and acoustic energy flux are evaluated at the ambient temperature location.

In an idealized open cycle thermoacoustic engine, the acoustic energy gain across the regenerator can be related to its hot and cold temperatures by integrating Eq. (13) across the regenerator to yield[16]

$$\frac{\dot{E}_h}{\dot{E}_c} \approx \frac{T_{\mathrm{re}}}{T_0}. \tag{22}$$

By setting Eq. (21) equal to zero, using the ideal gas law, Gedeon's estimate for $\dot{M}_{\mathrm{st}}$, and Eq. (22) yields a rough estimate for $\Delta p_2|_{\dot{M}=0}$:

$$\Delta p_2|_{\dot{M}=0} = \frac{\alpha \mu_m L}{\rho_m}\dot{M}_{\mathrm{st}} \approx \frac{\alpha \mu_m L}{\rho_m}\frac{\dot{E}_h}{R T_{\mathrm{re}}}. \tag{23}$$

Substituting Eq. (23) back into Eq. (21), and neglecting the regenerator loss terms in Eq. (10) to estimate $\dot{M}$, the required pressure difference across the regenerator is approximated by

$$\Delta p_2 \approx \frac{\alpha \mu_m L}{\rho_m}\frac{\dot{E}_h}{R T_{\mathrm{re}}}\left[1 + \frac{\gamma - 1}{\gamma (T_h / T_{\mathrm{re}} - 1)}\right]. \tag{24}$$

Since the acoustic energy flux is a second order quantity and all the other terms are of order zero, Eq. (24) verifies that the mean pressure difference across the regenerator is a second order quantity. Note that the first term in brackets on the right hand side of Eq. (24) is the pressure gradient required to counteract the effects of acoustic streaming in the regenerator, while the second term is the additional pressure required to force mass to flow at the desired rate from hot to cold through the regenerator. As T_{re} approaches T_h, the denominator in this term approaches zero, though the acoustic energy flux also approaches zero according to Eq. (10). Even for $\gamma = 1.4$ and a moderate interface temperature ratio, $T_h / T_{\mathrm{re}} = 1.3$, the additional pressure required to force mean flow in the desired direction is about the same as that required to negate the effects of acoustic streaming in the regenerator. Note also that the power required to force the mean flow through the regenerator is a fourth order quantity proportional to $\dot{M}\Delta p_2$, therefore neglecting this effect in considering the total second-order energy flux is well justified.

III. POWER OUTPUT AND THERMAL EFFICIENCY

In the implementation of an open cycle traveling wave thermoacoustic engine, T_{re} would not typically be known. Instead, $\dot{M}$, c_p, and T_h of the incoming hot gas would be the known quantities, in addition to the length of the regenerator, L, and the temperature of the ambient reservoir, T_0. Further, the possible geometries for the open cycle engine fall into two general categories: a linear device as first envisioned by Ceperley,[13] and a looped device similar to the TASHE.[3] In a linear engine, the input acoustic energy, $\dot{E}_c$, is typically a

known quantity, and $\dot{E}_h$ and T_{re} are determined using Eqs. (10) and (22), for example. In the looped configuration, a portion of $\dot{E}_h$ is routed through a feedback loop to provide the acoustic energy input $\dot{E}_c$, while the remainder of the acoustic energy is applied to an acoustic load, $\dot{E}_L$, such as a thermoacoustic refrigeration process. In this case, $\dot{E}_L \approx \dot{E}_h - \dot{E}_c$ is the known quantity, and an optimal T_{re} exists that yields the maximum possible $\dot{E}_L$ for the given thermal energy input to the engine. This particular case is studied below, although the same results can be used to determine the maximum acoustic gain $\dot{E}_g = \dot{E}_h - \dot{E}_c$ across the regenerator of a linear thermoacoustic engine.

A. Optimal regenerator interface temperature

With the aid of Eqs. (17) and (22), the amount of acoustic energy available for the system load, $\dot{E}_L$, is ideally equal to

$$\dot{E}_L \approx \dot{E}_h - \dot{E}_c = \dot{M} c_p \left(T_{\mathrm{re}} - T_h - \frac{(T_0 - T_{\mathrm{re}}) e^{\Xi L}}{1 - e^{\Xi L}} \right) \left(1 - \frac{T_0}{T_{\mathrm{re}}} \right), \tag{25}$$

where the acoustic load is shown to be a nonlinear function of T_{re}. In the absence of an acoustic load (i.e., $\dot{E}_L = 0$), Eq. (25) indicates that a temperature difference exists across the regenerator interface due solely to the thermal energy losses at this location:

$$T_{\mathrm{re,no\ load}} = T_h - (T_h - T_0) e^{\Xi L}, \tag{26}$$

where the parameter Ξ in the second term on the right is a function of the thermal losses in the regenerator.

As $\dot{E}_L$ increases, T_{re} decreases from the value given in Eq. (26), and the acoustic energy exiting the regenerator, $\dot{E}_h$, increases as shown in Eq. (10). However, as T_{re} decreases, so does the acoustic gain across the regenerator, according to Eq. (22). Given these two competing effects that increase and decrease $\dot{E}_h$, an optimum regenerator interface temperature must exist that maximizes the ideal acoustic power output of the engine, $\dot{E}_L$. To find this optimum temperature, $T_{\mathrm{re,opt}}$, Eq. (25) is differentiated with respect to T_{re}, and the resulting equation is set equal to zero, yielding

$$T_{\mathrm{re,opt}} = T_0 \sqrt{\frac{T_h}{T_0} - \left(\frac{T_h}{T_0} - 1 \right) e^{\Xi L}}. \tag{27}$$

Substituting Eq. (27) into Eq. (25) then gives, after some manipulation, the maximum acoustic power output of the engine for a given thermal energy input and thermal loss in the regenerator:

$$\dot{E}_{L,\max} = -\dot{M} c_p \left(\frac{T_0 e^{\Xi L}}{1 - e^{\Xi L}} + T_h \right) \times \left[1 - \sqrt{\frac{T_0}{T_h}} \sqrt{\frac{1}{1 - (1 - T_0/T_h) e^{\Xi L}}} \right]^2. \tag{28}$$

B. Simplification for large ΞL

The quantity ΞL essentially represents the ratio of the regenerator's convective energy flux to the thermal loss energy flux. Ideally, this ratio should be large to reduce the losses in the engine. Note also that according to Eq. (15) above, Ξ is negative since $\dot{M} < 0$ in this study.

Consequently, Eqs. (27) and (28) can be simplified to a reasonable degree of accuracy for large values of ΞL. Since in these cases, $e^{\Xi L} \ll 1$, the optimal regenerator interface temperature can be approximated by

$$T_{\mathrm{re,opt}} \approx \sqrt{T_0 T_h}, \tag{29}$$

and the maximum acoustic power output is approximately

$$\dot{E}_{L,\max} \approx -\dot{M} c_p T_h \left(1 - \sqrt{\frac{T_0}{T_h}} \right)^2. \tag{30}$$

For reasonable hot and cold temperatures (e.g., $T_h/T_0 > 1.5$), Eqs. (29) and (30) approximate Eqs. (27) and (28), respectively, to within 1.5% error for $\Xi L \leqslant -4$.

C. Engine efficiency

Having determined the maximum acoustic power output of the engine for a given thermal energy input, various engine efficiencies can now be computed. The thermal efficiency of the engine is equal to the acoustic power output, $\dot{E}_L$, divided by the thermal energy input, $-\dot{M} c_p (T_h - T_0)$. The thermal efficiency for the open cycle's maximum power output is then given by

$$\eta_{\mathrm{th,oc}} = \frac{\dot{E}_{L,\max}}{-\dot{M} c_p (T_h - T_0)} \approx \frac{(1 - \sqrt{T_0/T_h})^2}{1 - T_0/T_h}, \tag{31}$$

where Eq. (30) has been used for $\dot{E}_{L,\max}$, since it represents an upper limit of $\dot{E}_L$ for negligible thermal energy losses in the regenerator.

To estimate the effectiveness of the cycle, an availability analysis gives the maximum amount of work that could be extracted from the convected thermal energy input:[21]

$$\dot{E}_{L,\mathrm{avail}} = -\dot{M} c_p [T_h - T_0 - T_0 \ln(T_h/T_0)]. \tag{32}$$

Therefore, the second law efficiency, which expresses the ratio of the actual power output to the power output available in the flow, is given by

$$\eta_{\mathrm{II,oc}} = \frac{\dot{E}_{L,\max}}{\dot{E}_{L,\mathrm{avail}}} = \frac{(1 - \sqrt{T_0/T_h})^2}{1 - T_0/T_h (1 - \ln(T_0/T_h))}, \tag{33}$$

where, again, Eq. (30) has been used for $\dot{E}_{L,\max}$.

IV. CLOSED CYCLE COMPARISON

In order to evaluate the feasibility of the open cycle traveling wave thermoacoustic engine, its thermal efficiencies and acoustic power output should be compared to corresponding performance parameters in closed cycle thermoacoustic engines. Since the goal of the open cycle engine is to convert the thermal energy in a stream of gas into acoustic energy, the closed cycle engines will be evaluated by deter-

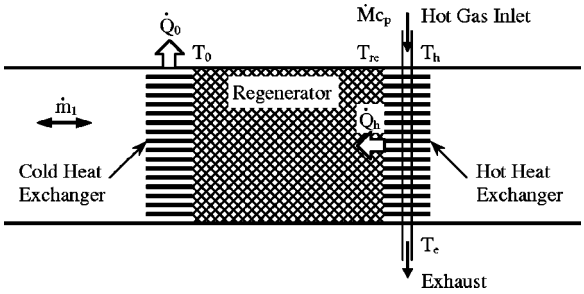


FIG. 3. A schematic of a closed cycle thermoacoustic engine that employs a crossflow heat exchanger to transfer heat $\dot{Q}_h$ from a stream of hot gas with a mean mass flux $\dot{M}$, specific heat c_p, and a mean temperature T_h, to the regenerator of the thermoacoustic engine. The stream of hot gas leaves the heat exchanger at an exhaust temperature, T_e, and the hot side of the regenerator is assumed to have a uniform temperature, T_{re}. On the cold side of the regenerator, heat $\dot{Q}_0$ is rejected to the ambient temperature T_0 through another crossflow heat exchanger, and the resulting temperature gradient across the regenerator is used to amplify a traveling acoustic wave.

mining the acoustic power output that they can derive from a similar stream of hot gas, carrying a thermal energy given by $\dot{M}c_p(T_h-T_0)$.

Two basic methods can be used to transfer the thermal energy in the stream of gas to the closed cycle thermoacoustic engine. In the first method, the stream of gas is used to generate electrical energy, which is then converted back into thermal energy in the closed cycle engine with an electric resistance heater. While this method is certainly convenient from the end-user's standpoint, the conversion to electrical energy, transmission and conversion back to thermal energy is much less efficient than direct use of the thermal energy. Furthermore, such a system is not very practical in remote settings, as electrical generation hardware only adds to the cost and maintenance of the overall system. Therefore, this study will focus on the second method of heat transfer, which uses a heat exchanger to directly transfer the thermal energy in the hot gas stream into the thermoacoustic engine. The performance obtained with this method will be analyzed and compared to that provided by the open cycle engine in the following sections.

A. The crossflow hot heat exchanger

A crossflow heat exchanger, as depicted in Fig. 3, provides a practical means for transferring the heat from the hot stream of gas to the thermoacoustic engine because all of the heat transfer occurs over an acoustic displacement length, which is normally much smaller than the diameter of the engine. However, the design of a crossflow heat exchanger will not allow all of the heat in the hot gas stream to be transferred into the engine. As will be shown below, this configuration also produces a maximum acoustic power output for an optimum heat transfer and hot regenerator temperature.

The heat transfer from the hot gas stream to the thermoacoustic engine can be expressed as

$$\dot{Q}_h=\dot{M}c_p(T_h-T_e)=C_h(T_h-T_e), \tag{34}$$

where T_e is the exhaust temperature of the hot gas stream as it exits the heat exchanger. The heat capacity rate is defined as $C=\dot{M}c_p$, and C_h in Eq. (34) is the heat capacity rate for the hot side of the heat exchanger.

The complex physical processes occurring on the thermoacoustic side of the heat exchanger make calculation of the heat transfer and relevant gas temperatures difficult. For instance, the temperature is nonuniform in the transverse direction because hot gas enters one side of the heat exchanger and colder exhaust gas exits the other side. To simplify matters, it is assumed that the thermoacoustic gas rapidly mixes and/or that transverse thermal conduction in the regenerator creates a uniform temperature, T_{re}, at the hot side of the regenerator.

The determination of the temperature of the thermoacoustic gas as it returns to the regenerator from the crossflow heat exchanger is also problematic. In this case, gas enters one side of the heat exchanger at T_{re}, traverses to the other side and comes back again, with heat being transferred to the gas from the surface of the heat exchanger in the process. In addition to the temperature change that may occur during this heat transfer process, the gas temperature periodically varies due to its expansion and compression by the acoustic waves.

B. Heat exchanger effectiveness

Most or all of the complexities associated with using a crossflow heat exchanger in an acoustic environment can be avoided by considering its effectiveness,[22] ε, which is defined as the heat transfer rate through the heat exchanger, $\dot{Q}_h$, divided by the maximum heat transfer rate possible, $\dot{Q}_{\max}$, for the temperatures and heat capacity rates involved, i.e.,

$$\varepsilon\equiv\frac{\dot{Q}_h}{\dot{Q}_{\max}}. \tag{35}$$

$\dot{Q}_{\max}$ is obtained for the minimum heat capacity rate between the two fluids and between the maximum and minimum temperatures present in the heat exchanger, i.e.,

$$\dot{Q}_{\max}=C_{\min}(T_{\max}-T_{\min}). \tag{36}$$

Ideally, this maximum heat transfer rate could only be achieved in an infinitely long counterflow heat exchanger, where one of the two fluids experiences the maximum temperature change. Note that $\dot{Q}_{\max}$ is obtained with the minimum heat capacity rate, $C_{\min}$, not the maximum, $C_{\max}$, since, in the example of the counterflow heat exchanger, the fluid experiencing the largest temperature difference must also have the lower heat capacity rate.[22]

In this analysis, the maximum temperature is chosen as $T_{\max}=T_h$, while the minimum temperature is given by $T_{\min}=T_{re}$, because the thermoacoustic gas leaves the regenerator at the temperature T_{re}, is heated within the heat exchanger and returns to the regenerator hotter than when it left. Furthermore, $T_{re}<T_e$, would be transferred out of some portion of the thermoacoustic engine back into the heat exchanger fluid.

The choice for $C_{\min}$ in Eq. (36) deserves careful consideration. To help make this choice, the following expressions

for the heat capacity rate on the thermoacoustic side (subscript ''ta'') and the hot gas side (subscript ''hg'') of the heat exchanger are derived using the ideal gas law and the relationship $c_p/R=\gamma/(\gamma\text{-}1)$:

$$C_{\mathrm{ta}}\equiv(\overline{|\dot{m}_1|}c_p)_{\mathrm{ta}}\approx(\overline{|U_1|}\rho_0 c_p)_{\mathrm{ta}}=\left(\frac{\gamma}{\gamma-1}\frac{\overline{|U_1|}p_0}{T_0}\right)_{\mathrm{ta}}, \quad (37)$$

$$C_{\mathrm{hg}}\equiv(\dot{M}c_p)_{\mathrm{hg}}=(U_0\rho_0 c_p)_{\mathrm{hg}}=\left(\frac{\gamma}{\gamma-1}\frac{U_0 p_0}{T_0}\right)_{\mathrm{hg}}, \quad (38)$$

where $\overline{|U_1|}$ is the time average of the magnitude of the acoustic volumetric velocity on the thermoacoustic side.

In the open cycle engine, $U_{0,\mathrm{hg}}\ll U_{1,\mathrm{hg}}$ in order to keep convection effects from dominating the acoustic effects in the device.[16] Since the acoustic power outputs of the open and closed cycle cases are expected to be comparable, $U_{1,\mathrm{hg}}\approx U_{1,\mathrm{ta}}$, so it is then reasonable to expect that $U_{0,\mathrm{hg}}<U_{1,\mathrm{ta}}$. Furthermore, since closed cycle thermoacoustic engines typically operate at elevated pressures to increase the acoustic power density in the engine, it is reasonable to assume that $p_{0,\mathrm{hg}}\leqslant p_{0,\mathrm{ta}}$, while heat is only transferred into the engine for $T_{0,\mathrm{ta}}<T_{0,\mathrm{hg}}$. Considering these factors in comparing Eqs. (37) and (38), and recognizing that γ has little influence on the heat capacity rates, it follows that $C_{\min}=C_{\mathrm{hg}}$. Substituting this result and Eq. (36) into Eq. (35) yields a simple expression for the heat transfer into the thermoacoustic engine:

$$\dot{Q}_h=\varepsilon C_{\mathrm{hg}}(T_h-T_{\mathrm{re}}), \quad (39)$$

where the details of the heat exchanger geometry and heat exchange process are contained in the effectiveness term, ε.

The effectiveness parameter is a function of two dimensionless parameters,[22] $C_{\min}/C_{\max}$ and the ''Number of Transfer Units'' (NTU), defined as

$$\mathrm{NTU}=\frac{UA}{C_{\min}}, \quad (40)$$

where U and A are the heat transfer coefficient and surface area of the heat exchanger, respectively. While an effectiveness-NTU relationship has not been derived for a crossflow heat exchanger in an acoustic environment, these relationships exist for many types of heat exchangers.[22] Note, however, that all heat exchangers attain their maximum effectiveness with respect to the heat capacity ratio in the limiting case where $C_{\min}\ll C_{\max}$, for which[22]

$$\varepsilon\approx 1-\exp(-\mathrm{NTU}). \quad (41)$$

Since the determination of $C_{\max}$ is somewhat problematic in this case, Eq. (41) can be used as an estimate for the maximum effectiveness of the crossflow heat exchanger. Note that as the NTU increases, $\varepsilon\to 1$, though this also indicates an increasing UA for the heat exchanger according to Eq. (40). Since the effective length of the heat exchanger in a thermoacoustic device is limited to the acoustic displacement length, increasing the heat exchanger surface area, A, would require the addition of more fins to the heat exchanger. This would increase the viscous dissipation in the heat exchanger, thus placing a practical limit on the magnitude of the heat transfer coefficient UA, and the magnitudes of the NTU and ε.

C. Maximum power output

The expression for the heat input, $\dot{Q}_h$ in Eq. (39), can be linked to the acoustic energy output of the closed cycle engine by investigating the energy balance around the heat exchanger. Noting that the total energy fluxes in the closed cycle engine are the same as those for the open cycle engine above, with the convective terms excluded, this energy balance can be expressed as

$$\dot{Q}_h=\dot{H}_{h,\mathrm{cc}}-\dot{H}_{\mathrm{re},h,\mathrm{cc}}=\dot{E}_{h,\mathrm{cc}}+\left[(\psi+Ak+A_s k_s)\frac{dT_m}{dx}\right]_{\mathrm{cc}}, \quad (42)$$

where Eqs. (8) and (9) have been applied without the convection terms, viscous dissipation in the heat exchanger has been neglected, and the subscript ''cc'' denotes quantities evaluated in the closed cycle engine. Since the thermal losses in the regenerator should be small compared to the acoustic energy flux in an efficient closed cycle engine, Eq. (42) indicates that $\dot{Q}_h\approx\dot{E}_{h,\mathrm{cc}}$. Using this approximation and Eqs. (22) and (39), the following expression for the acoustic power output of the closed cycle engine with a crossflow hot heat exchanger is obtained:

$$\dot{E}_{L,\mathrm{cc}}=\dot{E}_{h,\mathrm{cc}}-\dot{E}_{c,\mathrm{cc}}\approx\varepsilon\dot{M}c_p(T_h-T_{\mathrm{re}})\left(1-\frac{T_0}{T_{\mathrm{re}}}\right). \quad (43)$$

As was done for the open cycle case, Eq. (43) can be optimized with respect to T_{re} to yield a maximum acoustic power output. The optimum regenerator temperature, $T_{\mathrm{re,opt}}$, turns out to be the same as that of the ideal open cycle case, given in Eq. (27) above. The resulting maximum power output for the closed cycle engine with a crossflow heat exchanger is

$$\dot{E}_{L,\mathrm{cc,max}}=\varepsilon\dot{M}c_p T_h\left(1-\sqrt{\frac{T_0}{T_h}}\right)^2, \quad (44)$$

which is identical to the power output of the ideal open cycle case in Eq. (30), multiplied by the heat exchanger effectiveness, ε. [Note that the negative sign in Eq. (30) is due to the sign convention adopted for the direction of the mean flow in the open cycle case.] Not surprisingly, since $\varepsilon<1$ for a crossflow heat exchanger, it follows that the open cycle engine delivers more acoustic power than a closed cycle engine with a crossflow heat exchanger, given the same input flow of hot gas.

D. Thermal efficiency

Using Eq. (27) for $T_{\mathrm{re,opt}}$, the optimum heat flux into the closed cycle engine is given by

$$\dot{Q}_{h,\mathrm{opt}}=\varepsilon\dot{M}c_p T_h\left(1-\sqrt{\frac{T_0}{T_h}}\right), \quad (45)$$

and the thermal efficiency is given by

$$\eta_{\mathrm{th,cc,1}}=1-\sqrt{\frac{T_0}{T_h}}. \quad (46)$$

This efficiency does not represent the maximum thermal efficiency, but the thermal efficiency at the point of maximum power output, which is larger than that given for the open cycle engine in Eq. (31). This occurs because the hot exhaust gas temperature, T_e, is larger than the ambient temperature T_0, so the heat in this exhaust gas could still be applied to some useful purpose. In this regard, the closed cycle engine with a crossflow hot heat exchanger holds an advantage over the open cycle engine, which consumes all of the input thermal energy to produce acoustic power. However, if the thermal energy in the exhaust stream is not used, then a more appropriate thermal efficiency would be based on the amount of heat available from the hot stream of gas, i.e.,

$$\eta_{\mathrm{th,cc,2}}=\frac{\varepsilon(1-\sqrt{T_0/T_h})^2}{1-T_0/T_h}. \tag{47}$$

Again, this differs from the thermal efficiency of the open cycle case, Eq. (31), by the heat exchanger effectiveness, ε. Finally, a second law efficiency can be computed based on the available work in the stream of hot gas, as expressed in Eq. (32):

$$\eta_{\mathrm{II,cc}}=\frac{\dot{E}_{L,\mathrm{cc,max}}}{\dot{E}_{L,\mathrm{avail}}}=\frac{\varepsilon(1-\sqrt{T_0/T_h})^2}{1-T_0/T_h(1-\ln(T_0/T_h))}. \tag{48}$$

This second law efficiency also differs from that of the open cycle case by the effectiveness parameter, ε, and is, therefore, smaller than the open cycle second law efficiency since $\varepsilon<1$.

E. Comparison of open and closed cycles

Figure 4(a) shows the normalized maximum acoustic power outputs of the open cycle and closed cycle engines, as expressed in Eqs. (30) and (43), in addition to the maximum available power output, as described by Eq. (32). The difference between the open and closed cycle cases is accounted for by the crossflow heat exchanger effectiveness, ε, which was assumed to equal 0.8 in Fig. 4. This factor can be made to approach unity by increasing the overall heat transfer coefficient and heat transfer surface area of the heat exchanger, or by decreasing the heat capacity ratio, $C_{\min}/C_{\max}$. However, cost, construction complexity, and other practical constraints such as viscous loss minimization will generally limit the effectiveness to $\varepsilon<0.9$, thus the acoustic power output of the open cycle thermoacoustic engine should always be larger than that for the closed cycle engine.

The open cycle thermal efficiency of Eq. (31) and second law efficiency of Eq. (33) are plotted as functions of T_h/T_0 in Fig. 4(b). Also plotted for the closed cycle case is $\eta_{\mathrm{th,cc,1}}$ of Eq. (46), based on the actual heat transferred into the thermoacoustic engine, $\eta_{\mathrm{th,cc,2}}$ of Eq. (47), where the thermal energy in the exhaust gas remains unused, and $\eta_{\mathrm{II,cc}}$ from Eq. (48). As discussed above, Fig. 4(b) shows that $\eta_{\mathrm{th,cc,1}}>\eta_{\mathrm{th,oc}}$, since the exhaust gas still carries thermal energy that can be used in another application. A similar situation, which was alluded to previously, also exists in the open cycle thermoacoustic engine, where some of the wasted convective thermal energy in the regenerator is recovered by

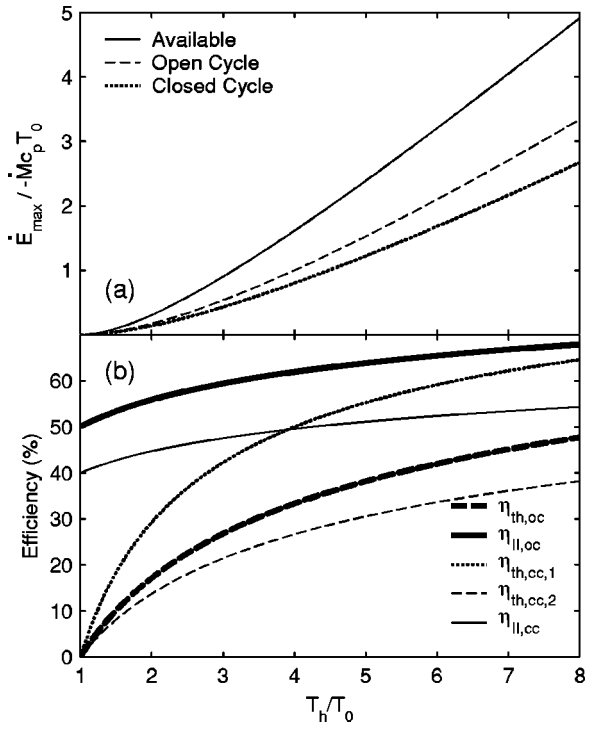


FIG. 4. The maximum dimensionless acoustic energy outputs (a), and the corresponding thermal and second law efficiencies (b), of open and closed cycle thermoacoustic engines, where an effectiveness of $\varepsilon=0.8$ has been assumed for the crossflow hot heat exchanger in the closed cycle case. Also plotted in (a) is the maximum power output available from the thermal energy in the hot stream of gas, based on a second law analysis of this gas stream. In (b), the subscript ''th'' refers to the thermal efficiency, ''II'' refers to the second law efficiency, ''oc'' refers to the open cycle, and ''cc'' to the closed cycle. Also, the first thermal efficiency for the closed cycle case assumes that the thermal energy in the exhaust gas is available for use, while the second thermal efficiency assumes that it is exhausted to the environment.

preheating the combustion process reactants.[19] Although this situation is not analyzed in this study, it is expected to yield a thermal efficiency comparable to $\eta_{\mathrm{th,cc,1}}$ by analogy to the closed cycle crossflow heat exchanger case. However, if the thermal energy in the exhaust gas of the heat exchanger is not used, the thermal efficiency of the closed cycle engine is shown to be smaller than that for the open cycle engine by the effectiveness factor, ε, as was the case for the acoustic power outputs in Fig. 4(a).

Finally, it should be noted that the acoustic power outputs and efficiencies in Fig. 4 represent the maximum attainable values for the configurations considered in these devices, and only account for the primary loss mechanisms associated with the crossflow heat exchanger effectiveness and the ''lost'' thermal energy in the convective heat flux passing through the regenerator in the open cycle engine. Although there are similarities between other loss mechanisms in the open and closed cycle engines, some differences do exist. For instance, the thermal losses in the regenerator are dependent on the temperature gradient and physical characteristics of the regenerator, which may vary between the open and closed cycle engines. Furthermore, the viscous dissipation losses of the crossflow hot heat exchanger are absent in the open cycle configuration, but additional viscous losses are produced in the regenerator by the exponential tempera-

ture profile and the resulting increase in the average acoustic velocity and viscosity coefficient in the regenerator. Thus, an improved evaluation of the difference in power output and thermal efficiency between the open and closed cycle engines would require a further investigation of the various open and closed cycle loss mechanisms.

V. CONCLUSIONS

This paper shows that the difference between the mean temperature of the hot incoming gas and the temperature at the hot side of the regenerator in an open cycle thermoacoustic engine plays a crucial role in the conversion of the thermal energy in the input stream of hot gas into acoustic energy, and significantly impacts the various energy fluxes in the open cycle engine. For a constant convective thermal energy input, increasing the acoustic load essentially increases the magnitude of this temperature difference, thereby increasing the amount of input thermal energy that is converted into ideal thermoacoustic entropy flux at the regenerator interface. The remainder of the input thermal energy entering the regenerator is essentially wasted and leaves the engine through the cold heat exchanger, having been transported down the regenerator temperature gradient by a combination of convective energy flux, conductive energy flux, and second order entropy flux due to imperfect thermal contact in the regenerator. However, since the hot-side regenerator temperature determines the acoustic gain across the regenerator, it has been shown that an optimal hot-side regenerator temperature exists that maximizes the acoustic power output of the engine for a given thermal energy input.

Furthermore, an effectiveness-NTU analysis of a crossflow heat exchanger, which can be used to bring thermal energy into a closed cycle traveling wave thermoacoustic engine, has shown that the open cycle configuration has the potential to be the more efficient means of converting thermal energy in a stream of gas into acoustic energy. Although this is a somewhat intuitive result, the effectiveness formulation employed could serve as a valuable tool in evaluating the efficiency of crossflow heat exchangers in closed cycle thermoacoustic devices.

Since open cycle thermoacoustic systems have not received much attention to date, however, more research is required to assess the feasibility of these devices. For instance, a more thorough investigation of the losses in open cycle thermoacoustic engines needs to be performed, and the results should be compared to those obtained for closed cycle systems. In addition, the effects of Rayleigh streaming, turbulence, and radial heat losses on the operation of the open cycle thermoacoustic engine need to be investigated. Also, as was suggested above, the open cycle engine's power output and efficiency could be further improved by recovering some of the convected thermal energy that is lost through the cold heat exchanger to, e.g., preheat combustion reactants that generate the hot stream of input gas. Such a system holds considerable promise for attaining high fuel energy-to-acoustic energy conversion efficiencies, and deserves further consideration. Finally, laboratory experiments and open cycle engine prototypes need to be designed, built, and tested, in order to check the validity of the theoretical predictions of this and other studies of open cycle traveling wave thermoacoustic systems.

[1] R. S. Reid, W. C. Ward, and G. W. Swift, ''Cyclic thermodynamics with open flow,'' Phys. Rev. Lett. **80**, 4617–4620 (1998).
[2] R. S. Reid and G. W. Swift, ''Experiments with a flow-through thermoacoustic refrigerator,'' J. Acoust. Soc. Am. **108**, 2835–2842 (2000).
[3] S. Backhaus and G. W. Swift, ''A thermoacoustic-Stirling heat engine: Detailed study,'' J. Acoust. Soc. Am. **107**, 3148–3166 (2000).
[4] D. Gedeon, ''DC gas flows in Stirling and pulse-tube cryocoolers,'' *Cryocoolers 9*, edited by R. G. Ross, Ed. (Plenum, New York, 1997), pp. 385–392.
[5] V. Gusev, S. Job, H. Bailliet, P. Lotton, and M. Bruneau, ''Acoustic streaming in annular thermoacoustic prime-movers,'' J. Acoust. Soc. Am. **108**, 934–945 (2000).
[6] S. Job, V. Gusev, P. Lotton, and M. Bruneau, ''Acoustic streaming measurements in annular thermoacoustic engines,'' J. Acoust. Soc. Am. **113**, 1892–1899 (2003).
[7] Y. Ueda, T. Biwa, U. Mizutani, and T. Yazaki, ''Experimental studies of a thermoacoustic Stirling prime mover and its application to a cooler,'' J. Acoust. Soc. Am. **115**, 1134–1141 (2004).
[8] H. Bailliet, V. Gusev, R. Raspet, and R. A. Hiller, ''Acoustic streaming in closed thermoacoustic devices,'' J. Acoust. Soc. Am. **110**, 1808–1821 (2001).
[9] D. L. Gardner and G. W. Swift, ''A cascade thermoacoustic engine,'' J. Acoust. Soc. Am. **114**, 1905–1919 (2003).
[10] G. W. Swift, D. L. Gardner, and S. Backhaus, ''Acoustic recovery of lost power in pulse tube refrigerators,'' J. Acoust. Soc. Am. **105**, 711–724 (1999).
[11] I. Charles, L. Duband, and A. Ravex, ''Permanent flow in low and high frequency pulse tube coolers—experimental results,'' Cryogenics **39**, 777–782 (1999).
[12] R. Waxler, ''Stationary velocity and pressure gradients in a thermoacoustic stack,'' J. Acoust. Soc. Am. **109**, 2739–2750 (2001).
[13] P. H. Ceperley, ''A pistonless Stirling engine—The traveling wave heat engine,'' J. Acoust. Soc. Am. **66**, 1508–1513 (1979).
[14] P. H. Ceperley, ''Gain and efficiency of a short traveling wave heat engine,'' J. Acoust. Soc. Am. **77**, 1239–1244 (1985).
[15] N. T. Weiland and B. T. Zinn, ''Open cycle traveling wave thermoacoustics: Mean temperature difference at the regenerator interface,'' J. Acoust. Soc. Am. **114**, 2791–2798 (2003).
[16] G. W. Swift, *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerators* (Acoustical Society of America, Melville, NY, 2002).
[17] N. Rott, ''Thermally driven acoustic oscillations, Part III: Second order heat flux,'' Z. Angew. Math. Phys. **26**, 43–49 (1975).
[18] J. H. Xiao, ''Thermoacoustic theory for cyclic flow regenerators. Part I: fundamentals,'' Cryogenics **32**, 895–901 (1992).
[19] N. T. Weiland, B. T. Zinn, and G. W. Swift, ''Traveling-wave thermoacoustic engines with internal combustion,'' U.S. Patent No. 6,732,515, 11 May 2004.
[20] S. Backhaus and G. W. Swift, ''An acoustic streaming instability in thermoacoustic devices utilizing jet pumps,'' J. Acoust. Soc. Am. **113**, 1317–1324 (2003).
[21] A. Bejan, *Advanced Engineering Thermodynamics*, 2nd ed. (Wiley, New York, 1997).
[22] F. P. Incropera and D. P. DeWitt, *Fundamentals of Heat and Mass Transfer*, 4th ed. (Wiley, New York, 1996).

Material discrimination using bispectral signatures

Paul A. Nyffenegger,[a] Melvin J. Hinich, Dillon Ritter, and Stephen Hansen
Applied Research Laboratories, University of Texas at Austin, Austin, Texas 78713-8029

(Received 30 March 2003; revised 1 June 2004; accepted 14 June 2004)

One goal of investigative signal processing techniques is to discriminate between types of materials composing an object. This paper explores a method for material discrimination and characterization using bispectral signatures acquired from an object actively probed with acoustic pulses. The mathematical foundations of the bispectrum are presented, and the proposed technique is tested with an ultrasonic apparatus. Results indicate that at ultrasonic frequencies this technique provides signatures with the potential of discriminating between classes of materials such as plastic, aluminum, and rock. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1780169]

PACS numbers: 43.35.Yb, 43.60.Cg, 43.58.+z [EJS] Pages: 1518–1523

I. INTRODUCTION

One goal of investigative signal processing techniques is to discriminate between types of materials composing an object or target. Such diverse areas as remote sensing, medical imaging, nondestructive evaluation, and automated detection and classification commonly share a variety of well-known approaches including spectral analysis and correlation techniques designed to characterize the transfer function of the target.

Within the last decade many researchers have developed bispectral techniques to recover information contained within time series that would otherwise remain hidden. These methods were originally developed for analyzing the statistical properties of stationary stochastic processes that satisfy the standard mixing conditions.[1] Such situations typify the problem of source detection in a passively monitored system consisting of an unknown source embedded in a media with an unknown structure (e.g., the standard passive ranging framework[2]). Bispectral techniques have found utility in a diverse array of fields such as economics,[3] condition-based maintenance,[4] biomedical research,[5] speech processing,[6] and image processing and reconstruction.[7] There is ongoing research applying higher order moments, cumulants, and cumulant spectra for system identification and reconstruction[8] and signal detection.[9] Higher-order moments, cumulants, and cumulant spectra techniques have also been applied to active ranging problems in both radar[10] and sonar[11] where considerable deterministic components exist.

Much of the body of research applying bispectra and cumulant spectra seeks to detect a specific condition, gain insight into the root causes of the condition, and/or mitigate the effects of the condition. Whereas the physical mechanisms resulting in a significant bispectrum from classes of transient signals are generally understood,[12] the mechanisms responsible for the bispectral signatures revealed using active ranging have not been well studied. In particular, little is known about the relative contributions to the bispectrum originating in the physical properties of the target material itself rather than from target structural acoustics and the propagation media.

In this paper we present a method for material discrimination and characterization using bispectral signatures acquired from a target actively probed with acoustic pulses. Here the analysis is not applied to the typical stationary stochastic time series these techniques were developed for. Instead, significant deterministic components exist in the data due to repeated trials with a known transmitted waveform, a constant velocity media, and a fixed target. In this case, we are not making statistical statements about any single time series but rather comparative statements between the echoes from the three targets. First we present the mathematical foundations of the bispectrum and derive the bispectral signature, then test the proposed technique with an ultrasonic apparatus. We present evidence that at ultrasonic frequencies this technique provides signatures with the potential of discriminating classes of materials.

II. BACKGROUND

The received echo from a sonar pulse incident on a target is often modeled as a linear system. Let $\{e(t)\}$ denote a sonar pulse waveform input to a discrete-time *causal linear system* with a measured output $x(t)$ of the form $x(t)=\Sigma_{k=0}^{\infty}h(k)e(t-k)$. The system's complex transfer function is $H(f)=\Sigma_{k=0}^{\infty}h(k)\exp(-i2\pi kf)=|H(f)|\exp[i\phi(f)]$. Excluding the source and receiver, the transfer function of the system is the product of the target's transfer function and the transfer functions of the propagation environment for the outgoing pulse and the reflected pulse. The goal is to characterize the target's transfer function to the extent that discrimination or classification between types of targets is possible.

The third order cumulant of the measured output $x(t)$, also known as the bicovariance, is the expected value of the signal at three time points t, $t+\tau_1$, and $t+\tau_2$ and expressed as $C_x^3(\tau_1,\tau_2)=E[x(t)x(t+\tau_1)x(t+\tau_2)]$, assuming zero mean and stationarity. The stationary bispectrum of $x(t)$ defined as the two-dimensional Fourier transform of the third order cumulant C_x^3 is then[13]

[a]Author to whom correspondence should be addressed; Electronic mail: hinich@mail.la.utexas.edu

 © 2004 Acoustical Society of America

$$B_x(f_1,f_2)=\sum_{\tau_1=-\infty}^{\infty}\sum_{\tau_2=-\infty}^{\infty}C_x^3(\tau_1,\tau_2)\times\exp[-i2\pi(f_1\tau_1+f_2\tau_2)]. \quad (1)$$

Cast in terms of the frequency domain transfer function, the power spectrum of the output of the linear system is the familiar $S_x(f)=S_e(f)H(f)H(-f)=S_e(f)|H(f)|^2$ where $S_e(f)$ is the spectrum of $e(t)$ the input of the system. Similarly, the bispectrum of the output of the linear system can be expressed as

$$\begin{aligned}B_x(f_1,f_2)&=H(f_1)H(f_2)H^*(f_1+f_2)B_e(f_1,f_2)\\&=|H(f_1)H(f_2)H(f_1+f_2)|\\&\quad\times\exp[i(\phi(f_1)+\phi(f_2)-\phi(f_1+f_2))]\\&\quad\times B_e(f_1,f_2),\end{aligned} \quad (2)$$

where $B_e(f_1,f_2)$ is the bispectrum of the input of the system. Note that the phase of the bispectrum contains information about the phase of the complex linear transfer function that is absent from the power spectrum.[14] Also note that nonlinearity is not required to obtain a bispectrum but that a broad frequency band is. As demonstrated in what follows, phase information is crucial in the use of the bispectrum to characterize the transfer function of a target.

III. COMPUTING THE BISPECTRUM

The spectrum and bispectrum, and thus the normalized bispectrum, is computed using conventional nonparametric methods. When computing the bispectrum we take advantage of the following two properties: (1) bispectrum values are approximately normally distributed[15] and (2) bispectral estimators are approximately independent across frequencies.[16] We use the frame-averaging spectrum estimation method to shed light on the statistical issues of bispectrum normalization.

Suppose that we have a sample $\{x(1),\ldots,x(N)\}$ that we partition into $P=[N/L]$ nonoverlapping frames of length L where the last frame is dropped if it has less than L observations. The pth frame is $\{x_p(1),\ldots,x_p(L)\}=\{x((p-1)L+1),\ldots,x(pL)\}$. The discrete Fourier transform of the pth frame is

$$X_p(k)=\sum_{t=1}^{L}x_p(t)\exp\left(-i2\pi\frac{kt}{L}\right)$$

and the periodogram of the mth frame is

$$\frac{1}{L}|X_p(k)|^2=\frac{1}{L}X_p(k)X_p(-k).$$

Because $N\simeq LP$ the frame-averaged estimate of the spectrum at frequency $f_k=k/L$ is

$$\hat{S}_x(f_k)=\frac{1}{N}\sum_{p=1}^{P}|X_p(k)|^2. \quad (3)$$

Then the expectation $E[\hat{S}_x(f_k)]=S_x(f_k)+O(1/L)$ where the error term of order $1/L$ is due to the frame windowing of the spectrum, and the variance of the estimate for large values of L and P is $(1/P)S_x^2(f_k)$.

Similarly, the frame-averaged estimate of the bispectrum at frequencies (f_{k_1},f_{k_2}) is

$$\hat{B}_x(f_{k_1},f_{k_2})=\frac{1}{N}\sum_{p=1}^{P}X_p(k_1)X_p(k_2)X_p(-k_1-k_2) \quad (4)$$

with $E[\hat{B}(f_{k_1},f_{k_2})]=B(f_{k_1},f_{k_2})+O(1/L)$ and variance for large L and P expressed as $(L/P)S(f_{k_1})S(f_{k_2})S(f_{k_1}+f_{k_2})$. The spectrum-normalized estimate of the bispectrum is then

$$\hat{b}_x(f_{k_1},f_{k_2})=\frac{1/L\sqrt{N}\hat{B}_x(f_{k_1},f_{k_2})}{\sqrt{S(f_{k_1})S(f_{k_2})S(f_{k_1}+f_{k_2})}}. \quad (5)$$

Hinich[14] shows that the estimators $\hat{b}_x(f_{k_1},f_{k_2})$ can readily be manipulated into the form

$$\hat{\beta}_x=2|\hat{b}_x(f_{k_1},f_{k_2})|^2=\frac{2N/L^2|\hat{B}_x(f_{k_1},f_{k_2})|^2}{S(f_{k_1})S(f_{k_2})S(f_{k_1}+f_{k_2})} \quad (6)$$

which, for discrete frequencies (f_{k_1},f_{k_2}), is asymptotically distributed as independent, two degree-of-freedom, noncentral chi-squared variates $\chi^2(2,\lambda)$. The real and complex parts of the estimate constitute the two degrees of freedom, and the noncentrality parameter λ is approximated by the skewness. In the case of Gaussianity with skewness zero, the expected value of the bispectrum is zero. We determine the cumulative density function of $\hat{\beta}_x$ and then transform the result into chi-squared probability values that, when displayed as images, we call the bispectral signature.

IV. DATA ACQUISITION

The testing environment consists of a water-filled Lucite tank measuring 102.5 cm long by 57.5 cm wide by 45.5 cm deep with a motorized four-axis positioning system situated above the tank. A jig mount housing both the source and receiver transducers is suspended within the tank and connected to the bottom of a mast controlled with the Z and θ axis motors of the positioning system. The tank is placed on foam blocks to decouple it from ambient random vibrations.

Signal connectivity is illustrated in Fig. 1. The waveforms pass through a 50 dB RF power amplifier between the signal generator and the transmitter: a 1.5 in. diam 1 MHZ unfocused Panametrics transducer (#A392S-SU). A Valpey–Fisher pinpoint transducer (#VP-1093) acts as the receiver with analog signals passing through a Panametrics 5662 ultrasonic preamplifier set at 34 dB gain before entering the data acquisition PC. For the acquisition of signal echoes, the receive transducer projects about 10 cm in front of the transmit transducer and also acquires the outgoing waveform for quality control. The transmitter and receiver mount at the bottom of the positioning mast move as a unit. The receiver mount partially but not completely isolates it from signals traveling through the mount.

A PC running LabView functions as the apparatus primary acquisition component, controlling the triggering of the signal generator and digitizing hardware and software within

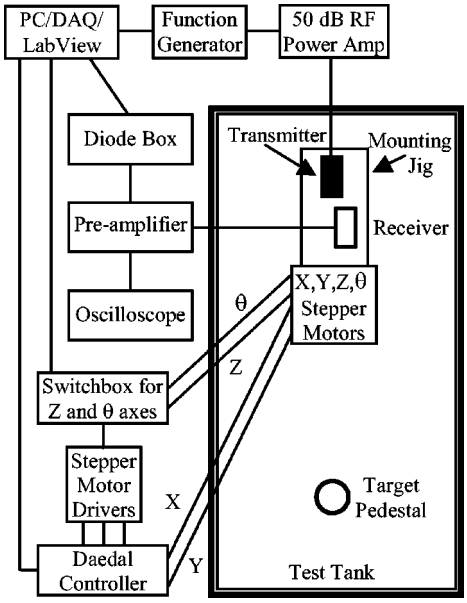


FIG. 1. Signal connectivity scheme for ultrasonic bispectral signature measurement.

the PC. The acquired signal passes through an analog anti-aliasing filter and then is digitized at 20 MHz. Multiple cycle acquisition, initial data display, and data file management is automated through LabView.

For this set of experiments, three solid cylindrical targets composed of aluminum, polyethylene, and an oolitic-fossiliferous limestone were chosen (Table I). The targets were placed on a Lucite pedestal with their long axis vertical such that they would sit in the center of the water column and in the center of the transmit beam. The limestone target contains natural irregularly shaped pore spaces; the partially marbleized nature of the sample indicates that some of the pore spaces will not become saturated with water even after the sample comes to equilibrium.

Two waveforms were designed specifically for this experiment and downloaded into the function generator. We generated a 150 microsecond sinusoidal pulse at 1 MHz (a tonal pulse), convolved with a Gaussian taper to limit ringing in the transmit transducer. The sinusoidal waveform acts as a control data set throughout the experiment because the limited bandwidth should produce no significant bispectral signature. The second waveform is a 150 microsecond linear-frequency-modulated (LFM) pulse swept from 100 kHz to 2500 kHz, also convolved with a Gaussian function to prevent ringing.

TABLE I. Target physical characteristics.

Target composition	Diameter (cm)	Height (cm)	Density (g/cm^3)	Acoustic velocity ($mm/\mu s$)
Polyethylene	2.54	15.24	0.979	2.43[a]
Limestone	3.81	10.16	2.44	~3.8[b]
Aluminum	2.54	15.24	2.68	~6.3[c]

[a]*Handbook of Chemistry and Physics*, 45th ed. (Chemical Rubber Co., Cleveland, OH, 1964), p. E-28.
[b]*Handbook of Physical Constants*, edited by S. P. Clark, Jr. (The Geological Society of America Memoir 97, 1966).
[c]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982), p. 461.

We acquire the one-way transmission and reflected pulse time series using the following algorithm.

For one-way data:

Set waveform

(i) Set initial distance (20 cm)
(ii) Acquire noise record
(iii) Acquire data for 200 transmit-receive acquisition cycles
(iv) Increment distance by 5 cm and repeat acquisition up to 60 cm

Change waveform and repeat procedure.

For target reflection data:

Set waveform

Set target

(i) Set initial distance (20 cm)
(ii) Acquire noise record
(iii) Acquire data for 200 transmit-receive acquisition cycles
(iv) Increment distance by 5 cm and repeat acquisition up to 60 cm

Change target and repeat acquisition

Change waveform and repeat procedure.

Acquisition geometry offsets are reported for the distance between the transmitter and receiver for the one-way geometry and between the receiver and the target for target reflection acquisition.

V. DATA PROCESSING AND RESULTS

Before estimating the spectrum and bispectrum, we perform several operations to condition the time series. First, we apply a standard low pass filter-decimation stage in order to remove signal aliases. The filter corner frequency is 5 MHz and the final time series sample rate is 10 MHz. A sample spectrogram of a time series after initial processing is shown in Fig. 2.

Because the target reflection is the focus of our analysis, we gate out the outgoing pulse, the tank back-wall reflections, and reflections of the target echo off of the mounting jig from the time series. Thus the spectral and bispectral estimates are performed over a 2000 sample window centered on the reflected pulse, and are averaged over the 200 data acquisition cycles.

Spectral estimates demonstrate that, as expected, the bulk of the power for the transmitted signal and the target echo falls between 500 kHz and 2500 kHz (Fig. 3). In addition, significant 2ν, 3ν, and 4ν harmonics created by the transmitter are readily apparent in both Fig. 2 and in the spectral estimates of the one-way transmissions shown in Fig. 3. The relationship between the fundamental frequency and the harmonics is confirmed in the ''control'' group of tonal 1000 kHz signals (not shown). The presence of the harmonics is beneficial in this study because they act to fill out the bispectral principal domain. A problem that became apparent after initial processing is the interference of two

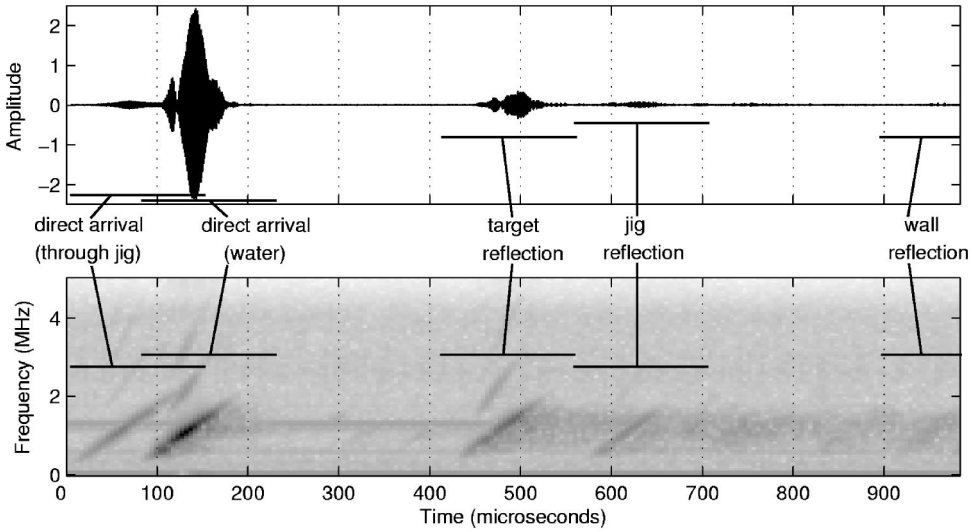


FIG. 2. Time record and spectrogram of the acquired signal averaged over 40 acquisition cycles. The test waveform is a linear frequency modulated sweep. The target for this set of experiments is an aluminum cylinder at a range of 250 cm with its long axis oriented perpendicular to the sonar beam axis. The spectral lines at about 1300 KHz and 590 KHz are the signal from nearby radio station transmission towers. Note the harmonics apparent in both the outgoing signal and the target echo.

radio stations with our electronic equipment, seen as high-power spectral lines at 590 kHz and 1300 kHz in Figs. 2 and 3. The differences between the spectra of the three target echoes (Fig. 4) are small; thus developing a discrimination algorithm based upon spectral differences or narrow-band comparisons would be difficult at best.

The bispectral signature of the one-way transmission for two offsets is shown in Fig. 5. Note that because of the wide frequency band and the harmonics present in the LFM pulse we would expect to see a nonzero bispectrum for the one-way transmission: nonlinearity is not required to explain this observation. There is very little difference between the bispectral signature for the two offsets even though the offset has doubled. This demonstrates that for our experimental design, frequency attenuation does not greatly affect the bispectral signature for ranges between these two offsets

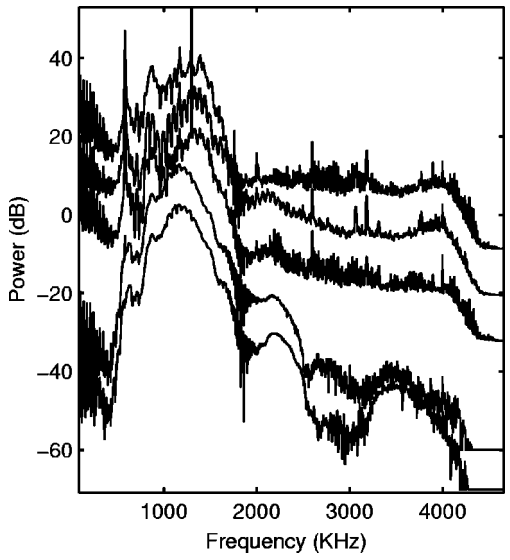


FIG. 3. Spectral estimates of source signal and target echoes offset vertically for display purposes. From the bottom most plotted line to the top most we show the spectral estimates of the one-way signal at a source-receiver offset of 25 cm (less 10 dB), the one-way signal at 50 cm, the aluminum reflection (+10 dB), the polyethylene reflection (+20 dB), and the limestone reflection (+30 dB). The reflections are each acquired at a distance of 35 cm. Each estimate is taken as the average over 200 frames, where the frame length is 2000 points. The spectral lines of two interfering AM radio stations are clearly seen in the target echoes at 590 and 1300 kHz.

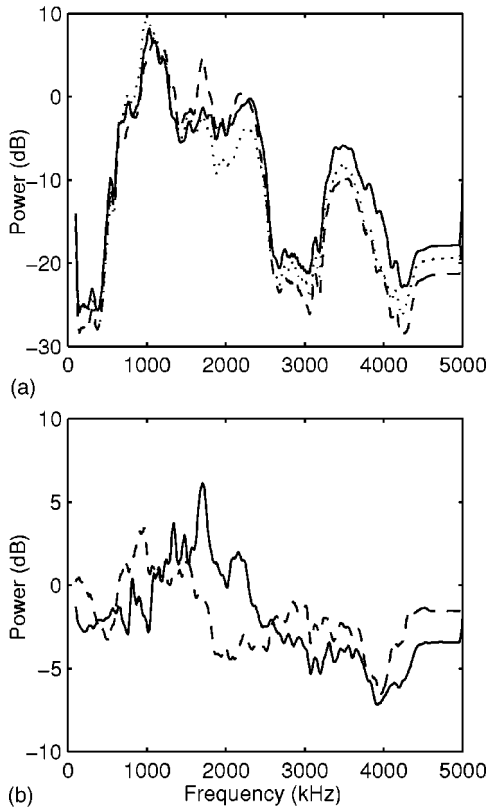


FIG. 4. Spectral comparisons: (A) difference between one-way signal at 25 cm and reflections of aluminum (solid), limestone (dashed), and polyethylene (dotted) at 35 cm; and (B) differences between reflection from aluminum and those of polyethylene (solid line) and limestone (dashed line). The spectral differences are smoothed using a 19-point triangular moving average.

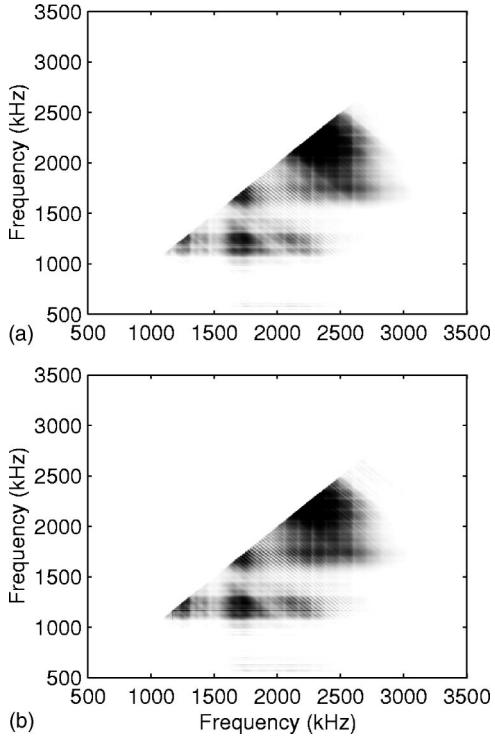


FIG. 5. Bispectral signature of the one-way transmission for (A) 25 cm and (B) 50 cm offsets. Displayed is the bispectral signature calculated over the principal bispectral domain for the case where no target is present, averaged over 200 acquisition cycles. The image intensity scale varies as probability of significance from 0 (white) to 1 (black).

over frequencies from approximately 1 MHz to about 2.8 MHz.

In comparison, the bispectral signatures for the three cylinders (Fig. 6) demonstrate remarkable variation. None of the three resemble the one-way bispectral signatures shown in Fig. 5, and the dissimilarity between the three is large even though differences in the power spectra were small. For example, the region above 2000 KHz is well filled in for the aluminum cylinder (a), dominated by a pattern reminiscent of coarse stippling for the polyethylene cylinder (b), and nearly absent from the limestone cylinder (c). Furthermore, the dominant response in both aluminum and limestone at about 1500 kHz is nearly absent from the bispectral signature of the polyethylene cylinder. The bispectral signatures for the targets, and thus the bispectral comparative differences, are nearly identical for all the offsets tested. In addition, note that the two interfering AM radio stations that contribute strong spectral lines in Fig. 3 contribute little or no value to the bispectral signature in Fig. 6.

VI. SUMMARY

The techniques applied in this study were originally developed for analyzing the statistical properties of stationary stochastic processes typical for the problem of source detection in a passively monitored environment. However, in this experiment significant deterministic components exist due to the repeated trials with a known transmitted waveform, a constant media, and a fixed target. The observed bispectral signatures then become useful as comparative statements between the echoes from the three targets rather than a statistical characterization of any single time series. Some small differences in the power spectra of the echoes from the three targets exist; differences at any given frequency are less than 10 dB. In contrast, even an untrained eye can see differences in the bispectral signatures, and the structure of the bispectral signature may lead to methods for discriminating between materials.

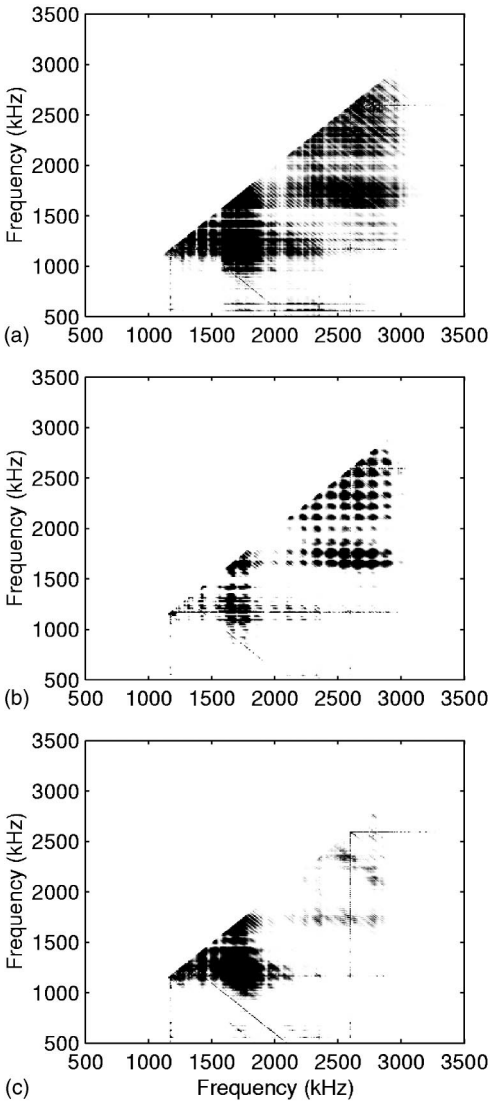


FIG. 6. Bispectral signatures of (A) aluminum, (B) polyethylene, and (C) limestone cylinders for the LFM waveform at an offset of 35 cm, averaged over 200 acquisition cycles. Image intensity scale is as described in Fig. 5.

At this point our observations suggest that the major variations in the three bispectral signatures originate in the material differences of the three cylinders because other factors such as transmitted waveform, dimensions of the cylinder, and target offset were held constant as the targets were changed. In this investigation, we intentionally limited our scope to an empirical demonstration of bispectral signature variation with changes in composition, and avoided developing an explanation of how the elastic properties of the materials give rise to variations in their bispectra. More work remains to be done in order to understand a material's bispectral response as a function of composition and wave-

form so that we may begin to explore physical models explaining the response.

For these observations to prove useful in a broader context, several aspects of the bispectral response of materials must be explored. Primarily, we must demonstrate scale invariance because simply scaling the cylindrical targets should not greatly affect phase if the main interaction is from the near surface of the target. Second, we must determine sensitivity to shape because variation in the target shape, especially topography of the target surface, should affect the relative phase of arriving echoes and in turn the bispectral response, especially if the main interaction is from the near surface of the target. Last, additional targets of various composition including heterogeneously composed targets ought to be investigated in order to determine the range of possible bispectral signatures, using a variety of broadband signals such as hyperbolic frequency modulated pulses and pseudorandom noise signals.

ACKNOWLEDGMENT

Funding for Dillon Ritter and Stephen Hansen was provided through Applied Research Laboratories' Independent Research and Development office.

[1] D. R. Brillinger, *Time Series Data Analysis and Theory* (Holt, Rinehart and Winston, New York, 1975).

[2] M. J. Hinich and G. R. Wilson, ''Time delay estimation using the cross bispectrum,'' IEEE Trans. Signal Process. **40**, 106–113 (1992); M. K. Broadhead, L. A. Pflug, and R. L. Field, ''Use of higher order statistics in source signature estimation,'' J. Acoust. Soc. Am. **107**, 2576–2585 (2000); L. A. Pflug, G. E. Ioup, and J. W. Ioup, ''Known source detection predictions for higher order correlators,'' *ibid.* **103**, 2469–2477 (1998).

[3] For example, A. Abhyankar, L. S. Copeland, and W. Wong, ''Nonlinear dynamics in real-time equity market indices: evidence from the United Kingdom,'' Econom. J. **105**, 864–880 (1995); P. L. Brockett, M. J. Hinich, and D. Patterson, ''Bispectral-based tests for the detection of Gaussianity and nonlinearity in time series,'' J. Am. Stat. Assoc. **83**, 657–664 (1988); R. B. Ashley, D. M. Patterson, and M. J. Hinich, ''A diagnostic test for nonlinear serial dependence in time series fitting errors,'' J. Time Ser. Anal. **7**, 165–178 (1986).

[4] For example, G. D. Meegan, H. R. Nelson, M. L. Barlett, and G. R. Wilson, ''Fault detection in engines through higher order spectral analysis of acoustic signatures,'' J. Acoust. Soc. Am. **106**, L1–L6 (1999); A. C. McCormick and A. K. Nandi, ''Bispectral and trispectral features for machine condition diagnosis,'' IEE Proc. Vision Image Signal Process. **146**, 229–234 (1999); R. W. Barker, G.-A. Klutke, M. J. Hinich, C. N. Ramirez, and R. J. Thornhill, ''Development and application of a statistically based feature extraction algorithm for monitoring tool wear in circuit board assembly,'' Circuits Syst. Signal Process. **13**, 411–433 (1994).

[5] For example, M. Shen, F. H. Y. Chan, L. Sun, and P. J. Beadle, ''Parametric bispectral estimation of EEG signals in different functional states of the brain,'' IEE Proc.: Sci., Meas. Technol. **147**, 374–377 (2000); J. Muthuswamy, D. L. Sherman, and N. V. Thakor, ''Higher order spectral analysis of burst patterns in EEG,'' IEEE Trans. Biomed. Eng. **46**, 92–99 (1999); P. Husar and G. Henning, ''Bispectrum analysis of visually evoked potentials,'' IEEE Eng. Med. Biol. Mag. **16**, 57–63 (1997).

[6] For example, J. Navarro-Mesa and A. Moreno-Bilbao, ''An improved speech endpoint detection system in noisy environments by means of third-order spectra,'' IEEE Signal Process. Lett. **6**, 224–226 (1999); B. Boyanov, S. Hadjitodorov, and T. Ivanov, ''Analysis of voiced speech by means of bispectrum,'' Electron. Lett. **27**, 2267–2268 (1991).

[7] For example, S. Bhattacharya, N. C. Ray, and S. Sinha, ''2-D signal modelling and reconstruction using third-order cumulants,'' Signal Process. **62**, 61–72 (1997); V. Chandran and S. Elgar, ''A general procedure for the derivation of principal domains of higher-order spectra,'' IEEE Trans. Image Process. **6**, 703–712 (1997); M. Michael Chang, A. Murat Tekalp, and A. Tanju Erdem, ''Blur identification using the bispectrum,'' IEEE Trans. Signal Process. **39**, 2323–2325 (1991).

[8] B. Chen, L. De Lathauwer, and A. P. Petropulu, ''Blind identification of complex convolutive mimo systems with 3 sources and 2 sensors,'' 2002 IEEE International Conference on Acoustics and Speech Signal Processes, Volume 2, 12–17 May, 2002, pp. 1669–1672; G. Jacovitti, G. Panci, and G. Scarano, ''Bussgang-zero crossing equalization: and integrated HOS-SOS approach,'' IEEE Trans. Signal Process. **49**, 2798–2812 (2001); J. K. Tugnait, ''Identification of multivariable stochastic linear systems via polyspectral analysis given noisy input-output time-domain data,'' IEEE Trans. Autom. Control **43**, 1084–1100 (1998); D. Yellin and E. Weinstein, ''Criteria for multichannel signal separation,'' IEEE Trans. Signal Process. **42**, 2158–2168 (1994).

[9] M. J. Hinich, IEEE Trans. Acoust., Speech, Signal Process. **38**, 1277–1283 (1990); M. J. Hinich and G. R. Wilson, ''Detection of non-Gaussian signals in non-Gaussian noise using the bispectrum,'' *ibid.* **38**, 1126–1131 (1990); K. M. Hock, ''Narrowband weak signal detection by higher order spectrum,'' IEEE Trans. Signal Process. **44**, 874–879 (1996); J. K. Tugnait, ''Detection of random signals by integrated polyspectral analysis,'' *ibid.* **44**, 2102–2108 (1996).

[10] I. Jouny, ''Description of radar targets using the bispectrum,'' IEE Proc. Radar Sonar Nav. **141**, 159–163 (1994); I. Jouny, F. D. Garber, and R. L. Moses, ''Target classification with the bispectrum: A comparative study,'' IEEE Trans. Aerosp. Electron. Syst. **31**, 69–77 (1995); F. Berizzi, G. Corsini, and F. Gini, ''HOS based algorithm for autofocusing of spotlight SAR images,'' Electron. Lett. **33**, 628–630 (1997).

[11] R. F. Dwyer, ''Range and doppler information from fourth-order spectra,'' IEEE J. Ocean. Eng. **16**, 233–243 (1991); J. A. Tague, C. M. Pike, and E. J. Sullivan, ''Active sonar detection in multipath: a new bispectral analysis approach,'' Circuits Syst. Signal Process. **13**, 455–466 (1994); A. Trucco, ''Detection of objects buried in the seafloor by a pattern recognition approach,'' IEEE J. Ocean. Eng. **26**, 769–782 (2001).

[12] P. L. Brockett, M. Hinich, and G. R. Wilson, ''Nonlinear and non-Gaussian ocean noise,'' J. Acoust. Soc. Am. **82**, 1386–1394 (1987); M. L. Barlett, K. W. Baugh, and G. R. Wilson, ''Transient detection using the nonstationary bispectrum,'' J. Acoust. Soc. Am. **99**, 3018–3028 (1996).

[13] M. J. Hinich and C. S. Clay, ''The application of the discrete fourier transform in the estimation of power spectra, coherence, and bispectra of geophysical data,'' Rev. Geophys. **6**, 347–363 (1968).

[14] M. Rosenblatt, *Stationary Sequences and Random Fields* (Birkhauser, Boston, 1985).

[15] M. J. Hinich, ''Testing for Gaussianity and linearity of a stationary time series,'' J. Time Ser. Anal. **3**, 169 (1982).

[16] M. J. Hinich and H. Messer, ''On the principal domain of the discrete bispectrum of a stationary signal,'' IEEE Trans. Signal Process. **43**, 2130 (1995).

Spectral methods for modelling guided waves in elastic media

A. T. I. Adamou and R. V. Craster
Imperial College London, Department of Mathematics, London SW7 2BZ, United Kingdom

(Received 21 November 2003; revised 7 June 2004; accepted 7 June 2004)

An efficient, accurate, and flexible numerical scheme based on spectral methods is developed here to determine dispersion curves and displacement/stress profiles for modes in elastic guiding structures, possibly curved, layered, damped, inhomogeneous, or anisotropic. Numerical examples and comparisons with existing techniques are given to demonstrate the efficiency and flexibility of the scheme. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1777871]

PACS numbers: 43.35.Zc, 43.58.Ta, 43.20.Mv, 43.35 [JBS] Pages: 1524–1535

I. INTRODUCTION

Guided elastic waves are used extensively in the nondestructive evaluation (NDE) of various structures.[1] An important step in the development of inspection techniques is the accurate and efficient calculation of dispersion curves and the associated displacement and stress profiles (or *mode shapes*). Existing techniques achieve this in three steps. The exact dispersion relation, expressed as the vanishing of the so-called *characteristic function*,

$$F(\text{frequency, wave number}) = 0,$$

is calculated; the roots of this characteristic function, which give the dispersion curves, are found using a numerical root-finding algorithm; and the mode shapes are obtained by substituting these roots back into the governing equations and solving. Considerable attention has been directed over the last 50 years to the optimization of the first of these steps, the calculation of the characteristic function. This is not particularly strenuous for the Lamb modes of a flat, homogeneous, isotropic, elastic plate. However, the introduction of curvature, layering, damping, inhomogeneity, or anisotropy can lead to expensive and tedious calculations. The review by Lowe,[2] which focuses on layered media in particular, describes the progress made in this area up to 1995. The techniques developed fall into two main categories: those based on the transfer matrix formulation of Thomson[3] and Haskell;[4] and those based on the global matrix formulation of Knopoff.[5] Both involve constructing a matrix description of the system in terms of quantities at its external boundaries, either the displacements and stresses or the incoming and outgoing wave amplitudes. Application of the desired boundary conditions requires that the determinant of part or all of the matrix be zero, and this determinant is identified as the characteristic function. The transfer matrix formulation has the advantage that the matrices involved are always small (e.g., 4×4 for in-plane displacements), so the characteristic function is easily evaluated. However, Dunkin[6] discovered that this method was susceptible to numerical instability at high frequencies or layer thicknesses (the so-called large-fd problem), and suggested a reformulation which avoided this problem. Various other reformulations have been suggested since. The global matrix method involves expressing all the boundary and interfacial conditions as a single matrix system and does not suffer from numerical instability. However, the size of the matrix increases with the number of layers, so the characteristic function can be prohibitively slow to compute for all but the simplest structures. Excellent progress has been made since 1995 to find even faster and more stable reformulations of the characteristic function, most notably the stiffness matrix method developed by Wang and Rokhlin[7] and the use of surface impedances by Hosten and Castaings.[8] Nevertheless the concept behind all these methods remains unchanged: calculate the characteristic function, then find its roots numerically. Henceforth we shall refer to this concept as *root finding*.

An alternative to root finding, which we explore in this paper, is to solve the underlying differential equations directly, by numerical interpolation. This approach is very fast but at the cost that it is inexact (in the sense that, unlike root finding, it is not based on the exact dispersion relations). However, we aim to show that by performing the interpolation with *spectral methods* instead of more familiar finite difference or finite element schemes, the results obtained are so accurate (*spectrally accurate*, to use the terminology in the literature) that the apparent problem of inexactness is immaterial. Spectral methods have become a standard computational tool in fluid dynamics since they were introduced in the 1970s by Kreiss and Oliger,[9] Orszag,[10] and Fornberg,[11] amongst others. More recently they have been used to solve elastic wave problems in seismology, further details of which can be found in the recent review by Carcione *et al.*[12] However, they seem to have been largely ignored as a means of solving similar problems in acoustic engineering, although it should be mentioned that a related method, the use of Fourier analysis and the fast Fourier transform algorithm, is advocated for wave propagation problems in the book by Doyle.[13] For introductory material on spectral methods the reader is referred to the books by Fornberg,[14] Trefethen,[15] and Boyd[16] and the software suite by Weideman and Reddy.[17] The basic idea is to use spectral differentiation matrices to cast the underlying equations in the form of a matrix eigenvalue problem, with the eigenvalues corresponding to, say, the frequency at a given wave number. This is then solved using a standard algorithm (we use MATLAB's `eig` routine which employs the QZ algorithm) and dispersion curves are generated by varying the wave number and repeating the procedure. Spectral methods are particularly well suited to solving eigenvalue problems because the number of interpolation

 © 2004 Acoustical Society of America

points, N, required for a given accuracy is considerably smaller than for finite difference or element schemes. This proves important as the QZ algorithm requires $O(N^3)$ operations to find the eigenvalues and eigenvectors of an $N \times N$ matrix, so N is at a premium.

For many applications root finding remains an entirely satisfactory approach. For example, the characteristic functions for flat guiding structures contain only simple trigonometric functions which are rapidly evaluated. A practitioner already using root finding for such an application would gain no significant benefit from replacing his existing software with routines based on spectral methods. However, we shall in this paper investigate certain scenarios where root finding does not work so well, such as curved, damped, inhomogeneous, and anisotropic media, where we believe spectral methods represent a marked improvement. Furthermore, spectral methods have a number of advantages in addition to their speed and accuracy that we believe make them a useful addition to the practitioner's toolbox:

(i) The coding effort required to implement them is minimal (see the example code in Appendix A).
(ii) The eigenvectors, which represent accurate approximations to the mode shapes, are automatically generated by the QZ algorithm, negating the need for separate calculations (the third step of the root-finding procedure).
(iii) The lowest eigenvalues are guaranteed to be found *without any being missed*. One of the pitfalls of root finding is that roots can be missed, particularly when the characteristic function varies rapidly.[18,19]
(iv) The approach is flexible, i.e., it can easily be extended to handle a wide variety of scenarios, such as those mentioned above.

The paper is structured as follows. Section II describes the method and provides a detailed comparison with the root finding method. Section III shows how the method can be extended to more complicated scenarios. We conclude with Sec. IV where we discuss our main findings and the scope for further work.

II. METHOD

It might seem sensible to describe our method by first implementing it for a flat wave guide. However, as mentioned in the Introduction, the characteristic functions for such a guide contain only trigonometric functions that are rapidly evaluated. Tests performed for flat guides show that spectral methods can compete with root finding in terms of efficiency, but there is very little between the two approaches and we would be hard pressed to convince a practitioner to throw out his old software and implement ours. We therefore choose to demonstrate our technique on a curved wave guide, specifically a circular annulus. This is not without applications: annular structures have attracted much recent attention in NDE due to their widespread use in the aerospace, oil, and nuclear industries.[20] The characteristic functions now contain Bessel functions which are less rapidly evaluated than trigonometric functions and, as Gridin *et al.*[21] note, can be difficult to evaluate at all for large orders or arguments. The spectral method bypasses this problem altogether since it does not require the manipulation of Bessel (or other special) functions whatsoever; one could view this as another advantage of this approach.

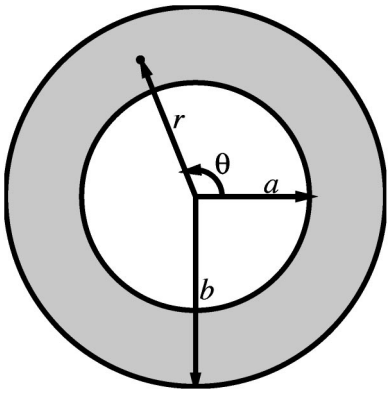


FIG. 1. The geometry of the annulus.

A. Out-of-plane displacements

We consider a circular annulus of inner radius a and outer radius b containing a homogeneous, isotropic, linearly elastic medium with longitudinal wave speed α and transverse wave speed β. We use polar coordinates as shown in Fig. 1. First we consider the case where the displacement, u, is perpendicular to the (r,θ) plane. We study time-harmonic waves propagating circumferentially, i.e., in the θ direction, whence u takes the form

$$u(r,\theta,t) = U(r)e^{i(\nu\theta-\omega t)}.$$

For a complete annulus periodicity requires integer ν. However, it is worth noting that noninteger values of ν do have physical meaning inasmuch as they can be applied to a circumferential wave propagating in an isolated section of an annulus. The equation of motion in polar coordinates,

$$\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\right)u=\frac{1}{\beta^2}\frac{\partial^2 u}{\partial t^2},$$

becomes an ordinary differential equation for $U(r)$,

$$\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{\nu^2}{r^2}\right)U=-\frac{\omega^2}{\beta^2}U. \tag{1}$$

We assume stress free boundary conditions on $r=a,\ b$ which for the out-of-plane case are simply

$$\frac{dU}{dr}=0, \quad r=a,b. \tag{2}$$

Analytically one would continue by writing Eq. (1) in the form of Bessel's equation with general solutions composed of linear superpositions of Bessel functions. The boundary conditions, Eq. (2), then give the characteristic function as

$$F(\omega,\nu)=J'_\nu(\omega a/\beta)Y'_\nu(\omega b/\beta)-Y'_\nu(\omega a/\beta)J'_\nu(\omega b/\beta), \tag{3}$$

where J_ν and Y_ν are the νth order Bessel functions of the first and second kinds, respectively, and the prime indicates the derivative with respect to the argument. We proceed differently by noting that for given ν, Eq. (1) is an eigenvalue problem where ω (strictly ω^2) plays the role of the eigenvalue and U the eigenfunction. (One could alternatively fix ω and treat ν as the eigenvalue.)

We wish to represent the differential operators in Eqs. (1) and (2) by differentiation matrices (hereafter referred to as DMs), which we introduce as follows. Consider a column vector $\mathbf{f}$ whose elements are the values of the function $f(x)$ evaluated at the N interpolation points x_i, $i=1,...,N$. The mth derivative is approximated by multiplying $\mathbf{f}$ by the $N \times N$ matrix $D^{(m)}$, i.e.,

$$\mathbf{f}^{(m)} \approx D^{(m)}\mathbf{f}.$$

Here the differentiation occurs on a bounded interval, $a \leqslant r \leqslant b$, so the natural choice is to use Chebyshev DMs. The MATLAB program `chebdif`[17] generates these for the interval $-1 \leqslant x \leqslant 1$. The interpolation points on this interval are

$$x_i = \cos\left(\frac{(i-1)\pi}{N-1}\right), \quad i=1,...,N. \tag{4}$$

We note that these points are not evenly spaced but are clustered at the ends of the interval. The appropriate interpolation points for r are given by

$$r_i = \frac{hx_i + a + b}{2}, \tag{5}$$

where $h=b-a$ is the thickness of the annulus. The DMs generated by `chebdif` approximate d^m/dx^m, but we require approximations to d^m/dr^m. The coordinate transformation, Eq. (5), gives us

$$D_r^{(m)} = \left(\frac{2}{h}\right)^m D_x^{(m)},$$

where the subscript denotes the differentiation variable. Henceforth, we will be dealing exclusively with r derivatives, so we omit the subscript.

The differential operator in Eq. (1),

$$\mathcal{L} = \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{\nu^2}{r^2},$$

is represented by the matrix

$$L = D^{(2)} + \mathrm{diag}\left(\frac{1}{r}\right)D^{(1)} - \mathrm{diag}\left(\frac{\nu^2}{r^2}\right), \tag{6}$$

where $\mathrm{diag}(f(r))$ is the matrix whose elements are $f(r_i)$ along the leading diagonal and zero elsewhere. This allows Eq. (1) to be written as a matrix eigenvalue equation,

$$L\mathbf{u} = -\frac{\omega^2}{\beta^2}\mathbf{u}, \tag{7}$$

where $\mathbf{u}$ is the column vector whose elements are $U(r_i)$. We note that this is only a partial description of the problem as it does not include the boundary conditions. These are incorporated by constructing the matrix $\tilde{L}$ by replacing the first and last rows of L with the first and last rows of $D^{(1)}$, i.e.,

$$\tilde{L} = \begin{pmatrix} \boxed{D^{(1)}} \\ \boxed{L} \\ \boxed{D^{(1)}} \end{pmatrix}.$$

Thus the first and last elements of $\tilde{L}\mathbf{u}$ give dU/dr at $r_1=b$ and $r_N=a$, respectively. The corresponding elements on the right are set to zero by introducing the following $N \times N$ matrix:

$$M(x) = -\frac{1}{x^2}\begin{pmatrix} 0 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0 \end{pmatrix}. \tag{8}$$

The factor $-1/x^2$ allows us to write the complete problem in the following neat form and will prove to be a useful notation later

$$\tilde{L}\mathbf{u} = \omega^2 M(\beta)\mathbf{u}.$$

This is known as a *generalized* eigenvalue problem and is solved by the MATLAB command `eig(L,M)`. We note that $\det M = 0$, so it is not possible to invert M and solve the more familiar *regular* eigenvalue problem

$$M(\beta)^{-1}\tilde{L}\mathbf{u} = \omega^2\mathbf{u}.$$

Finally, we note that numerical accuracy is improved by rescaling the first and last rows of $\tilde{L}$ by a factor $1/a$ so that they are of the same magnitude, $O(1/r^2)$, as the other rows. This is permissible only if the boundary conditions are homogeneous.

We now present typical results using the following physical parameters (which represent a steel annulus): $\alpha = 5960$ m/s, $\beta = 3260$ m/s, $a = 20$ mm, and $b = 23$ mm. The programs were run on a personal computer with a 1.7 GHz processor and 512 Mb of RAM. Appendix A contains the MATLAB code for solving the out-of-plane problem. For specified N and ν the code calculates the N lowest eigenvalues, ω, and the corresponding eigenvectors, $\mathbf{u}$. This example code is included to demonstrate the minimal coding effort required to solve this and subsequent problems.

Before examining the output from this code we must highlight pitfalls which the practitioner should be alert to. The first is that the matrices may, in addition to the eigenvalues of the physical problem, have some eigenvalues which do not correspond to real physical phenomena in the annulus, but are mere artifacts of the discretization process. Fortunately these nonsense eigenvalues are easily spotted since they are generally of a different magnitude to the physical eigenvalues or are complex (indicating modes which grow or decay with time, which is not possible in this scenario). Most tellingly, they change with N and this provides a useful check. Not all complex eigenvalues should be rejected, however. The error in the eigenvalue is not constrained to be real and may be, in part or in full, imaginary.

TABLE I. A comparison of the eigenvalues generated by the spectral method, ω_{sp}, with the exact values, ω_{ex}, for out-of-plane displacements in the steel annulus with $\nu=10$. The spectral method used $N=100$ and the loss of accuracy after the first $N/2$ modes is clearly seen.

m	$\omega_{sp}\,(10^6\ s^{-1})$	$\omega_{ex}\,(10^6\ s^{-1})$	Relative error
1	1.516 551 355 209 277	1.516 551 355 238 739	1.9×10^{-11}
10	30.762 634 741 805 80	30.762 634 741 802 87	9.5×10^{-14}
20	64.881 357 000 087 09	64.881 357 000 163 52	1.2×10^{-12}
30	99.013 811 558 054 86	99.013 811 558 055 61	7.5×10^{-15}
40	133.149 437 814 6480	133.149 437 815 5262	6.6×10^{-12}
50	167.286 291 165 0491	167.286 294 149 2643	1.8×10^{-8}
60	201.431 628 220 3012	201.423 755 145 4420	3.9×10^{-5}
70	242.304 294 949 5498	235.561 557 918 1455	0.029
80	344.330 256 713 5115	269.699 572 685 0146	0.28
90	646.594 891 745 9849	303.837 727 989 1275	1.1
100	∞	337.975 981 244 1500	n/a

Therefore, eigenvalues with very small imaginary parts must be examined to see whether they are physical. In practice these minor issues are easily resolved.

Let us also consider the theoretical accuracy of the results. For simplicity we consider a function $f(x)$ on our original Chebyshev interval, $-1\leqslant x\leqslant 1$. In order to accurately resolve f we require at least two interpolation points per wavelength. The interpoint spacing is greatest in the center of the interval and from Eq. (4) is asymptotically π/N for large N. Suppose f has a typical wavelength of λ, and this wavelength is roughly constant over the domain. Then our condition for accurate resolution is

$$\frac{\pi}{N}\leqslant\frac{\lambda}{2}\quad\text{or}\quad N\geqslant\frac{2\pi}{\lambda}.$$

Admittedly this is a rather heuristic argument, but the interested reader can find a formal demonstration (based on the convergence rates of Chebyshev series) in Gottlieb and Orszag.[22] So what does this mean for our spectral method? It tells us that for a given N we should expect accurate eigenvalues for the lower modes where the above condition holds, and then reduced accuracy for the higher modes as λ decreases. If we assume an inverse relationship between mode number, m, and λ (this being exact for the flat plate, where each unit increment in m corresponds to an extra half-oscillation across the guide) then our accuracy condition becomes $m\leqslant kN$ where k is some constant of proportionality. Table I shows the eigenvalues for the previously defined steel annulus with $\nu=10$ generated by the spectral method with $N=100$, together with the exact eigenvalues obtained using a root-finding method, which are correct to 16 digits. We see that the accuracy is extremely high for the first 50 or so modes and then starts to fall off, giving a value for k of around 1/2.

Having established the accuracy of the method and the conditions required to obtain it, we now consider the computational cost. Figure 2 shows the time taken to compute the first m eigenvalues to at least 10-digit accuracy by the spectral method (solid line) and by root finding (dashed line). For the spectral method we set $N=2m+10$ to get the required accuracy. We note that the graph is closely cubic in m because the QZ algorithm requires $O(N^3)$ operations to solve an $N\times N$ matrix. The root finding method works as follows: the approximate locations of the roots of the characteristic function are found by evaluating it at regular intervals in ω to find where it changes sign, and then using the MATLAB routine `fzero` to home in on the roots to the chosen accuracy. Performing the initial sweep reduces the probability of missing roots, and improves the overall speed by giving `fzero` a good initial guess to work with. Under this method it takes approximately the same amount of time to find each root, so the resulting graph is linear in m. We see that the spectral method is faster than the root-finding method for calculating fewer than around 400 eigenvalues, and for fewer than 100 the difference is large (up to two orders of magnitude, e.g., 0.020 s compared with 2.2 s to find the first 30 modes). This represents a significant time saving when plotting a set of dispersion curves, such as Fig. 3 which took 7 s to plot and is accurate to 10 digits. Our basic point here is that the method is very fast and easy to implement.

B. In-plane displacements

We now look at the case where the displacements are in the (r,θ) plane. Let $(u,v,0)$ be the displacement field in cylindrical polar coordinates. We follow the standard procedure (see, *inter alia*, Achenbach[23]) of writing it in terms of two scalar potentials, Φ and Ψ, as follows:

$$(u,v,0)=\nabla\Phi+\nabla\wedge(0,0,\Psi),$$

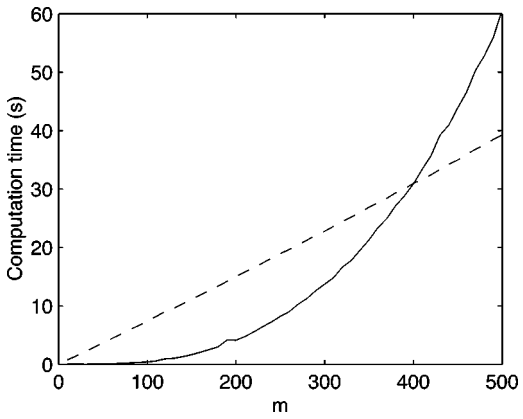


FIG. 2. Time taken to compute the first m eigenvalues to 10-digit accuracy for the out-of-plane problem with $\nu=10$ using the spectral method (solid line) and the root finding method (dashed line).

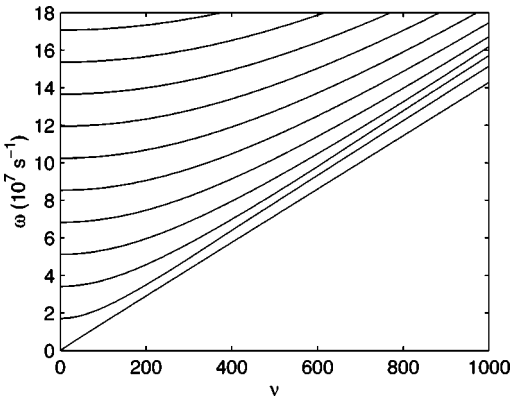


FIG. 3. The dispersion curves for the out-of-plane problem. Every fifth mode is shown starting from the lowest mode. A step size of 10 for ν, and $N=110$ were used to give a minimum of 10-digit accuracy everywhere.

which, in cylindrical polars, gives

$$u=\frac{\partial\Phi}{\partial r}+\frac{1}{r}\frac{\partial\Psi}{\partial\theta},\quad v=\frac{1}{r}\frac{\partial\Phi}{\partial\theta}-\frac{\partial\Psi}{\partial r}. \tag{9}$$

We pose similar forms for the potentials as for the displacement in the out-of-plane case,

$$[\Phi(r,\theta,t),\Psi(r,\theta,t)]=[\phi(r),i\psi(r)]e^{i(\nu\theta-\omega t)}. \tag{10}$$

Under this formulation the equations of motion reduce to two uncoupled ordinary differential equations for ϕ and ψ:

$$\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{\nu^2}{r^2}\right)\phi=-\frac{\omega^2}{\alpha^2}\phi, \tag{11}$$

$$\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{\nu^2}{r^2}\right)\psi=-\frac{\omega^2}{\beta^2}\psi, \tag{12}$$

Coupling occurs via the boundary conditions, which we take to be stress free on $r=a$, b. After some manipulation, the radial and tangential stresses are given by

$$\frac{\sigma_{11}}{\mu}=\left[\kappa^2\frac{d^2}{dr^2}+(\kappa^2-2)\left(\frac{1}{r}\frac{d}{dr}-\frac{\nu^2}{r^2}\right)\right]\phi+2\nu\left(\frac{1}{r^2}-\frac{1}{r}\frac{d}{dr}\right)\psi, \tag{13}$$

$$\frac{i\sigma_{12}}{\mu}=2\nu\left(\frac{1}{r^2}-\frac{1}{r}\frac{d}{dr}\right)\phi+\left(\frac{d^2}{dr^2}-\frac{1}{r}\frac{d}{dr}+\frac{\nu^2}{r^2}\right)\psi, \tag{14}$$

where μ is the shear modulus and $\kappa=\alpha/\beta$. Written in this way, the terms on the right-hand side are of the same order as those in Eqs. (11) and (12), so we can combine the governing equations and boundary conditions into matrix form without rescaling rows.

We wish to write the two governing equations as a single matrix equation, so we require a single column vector to represent both potentials ϕ and ψ. This is done by simply making $\mathbf{u}$ a vector of length $2N$ whose first N elements are $\phi(r_i)$ and whose last N elements are $\psi(r_i)$. Each of the potentials are acted on by the same differential operator, $\mathcal{L}$, as in the out-of-plane case, so the vector $\mathbf{u}$ is acted on by the $2N\times 2N$ matrix

$$P=\begin{pmatrix}L & 0\\ 0 & L\end{pmatrix}.$$

We now turn to the boundary conditions. Let $\mathbf{s}$ be the vector whose elements are $\sigma_{11}(r_i)/\mu$ followed by $i\sigma_{12}(r_i)/\mu$, and let this be related to $\mathbf{u}$ via the matrix S, i.e.,

$$\mathbf{s}=S\mathbf{u}.$$

By considering the differential operators acting on ϕ and ψ in Eqs. (13) and (14) we see that S has the form

$$S=\begin{pmatrix}A & B\\ B & C\end{pmatrix},$$

where $A-C$ are the following $N\times N$ matrices:

$$A=\kappa^2D^{(2)}+(\kappa^2-2)\left[\operatorname{diag}\left(\frac{1}{r}\right)D^{(1)}-\operatorname{diag}\left(\frac{\nu^2}{r^2}\right)\right],$$

$$B=2\nu\left[\operatorname{diag}\left(\frac{1}{r^2}\right)-\operatorname{diag}\left(\frac{1}{r}\right)D^{(1)}\right],$$

$$C=D^{(2)}-\operatorname{diag}\left(\frac{1}{r}\right)D^{(1)}+\nu^2\operatorname{diag}\left(\frac{1}{r^2}\right).$$

The boundary conditions are imposed by requiring elements 1, N, $N+1$, and $2N$ of $\mathbf{s}$ to be zero. We therefore replace these rows of P with the corresponding rows of S and denote the resulting matrix by $\tilde{P}$. For clarity we provide a pictorial representation of this

$$\tilde{P}=\begin{pmatrix}S\\ L\ \ 0\\ S\\ S\\ 0\ \ L\\ S\end{pmatrix}.$$

Elements 1, N, $N+1$, and $2N$ of $\tilde{P}\mathbf{u}$ are now the elements of $\mathbf{s}$. We set these to zero by constructing the matrix Q,

$$Q=\begin{pmatrix}M(\alpha) & 0\\ 0 & M(\beta)\end{pmatrix},$$

and solving

$$\tilde{P}\mathbf{u}=\omega^2Q\mathbf{u}.$$

We present the results in a similar fashion to the out-of-plane case, noting the same pitfalls and accuracy considerations. Table II shows that we do slightly better here than getting the first $N/2$ modes accurately: the accuracy only starts to drop at around the 70th mode for $N=100$. Figure 4 shows that the spectral method is faster than the root finder as far as the first 180 modes or so, and significantly faster for fewer than 50 modes. Figure 5 shows the dispersion curves for the in-plane problem. We note that, unlike the out-of-plane problem, the dispersion curves appear to cross each other (see Liu and Qu[18] for a discussion of this). Upon closer inspection, however, some of these crossovers are in fact two

TABLE II. A comparison of the eigenvalues generated by the spectral method, ω_{sp}, with the exact values, ω_{ex}, for in-plane displacements in the steel annulus with $\nu=10$. The spectral method used $N=100$.

Mode	ω_{sp} (10^6 s^{-1})	ω_{ex} (10^6 s^{-1})	Relative error
1	0.792 037 618 085 718	0.792 037 122 235 856	6.3×10^{-7}
10	18.872 592 380 927 40	18.872 592 381 105 08	9.4×10^{-12}
20	40.987 898 344 786 29	40.987 898 344 773 13	3.2×10^{-13}
30	62.472 863 302 674 63	62.472 863 302 560 36	1.8×10^{-12}
40	85.357 999 780 751 26	85.357 999 781 031 65	3.3×10^{-12}
50	106.138 599 827 3447	106.138 599 827 3571	1.2×10^{-13}
60	129.734 235 748 5949	129.734 235 748 6150	1.5×10^{-13}
70	150.217 969 333 2550	150.217 969 600 3870	1.8×10^{-9}
80	174.108 685 465 9129	174.111 799 440 0534	1.8×10^{-5}
90	195.406 299 672 6103	194.596 330 539 5677	4.2×10^{-3}
100	249.667 441 002 7489	218.444 295 862 5206	0.14

curves approaching closely without crossing, as can be seen in the inset figure. Many data points are required to differentiate between a crossover and a close approach so plotting the dispersion curves can be a very time-consuming business. Thus the spectral method, with its greater speed, particularly for the first 50 modes, comes into its own. Figure 5 took 39 s to plot using the spectral method; for the same accuracy and number of data points it would take around half an hour using the root-finding method. It must, however, be mentioned that this situation can be improved by employing the asymptotic simplifications of Gridin *et al.*[21]

III. FURTHER APPLICATIONS

The spectral method approach has been described by implementing it on a relatively simple guided wave problem, and its accuracy and efficiency advantages over root finding have been considered in some detail. We now discuss further applications. First we consider how to adapt the method to model waves in layered media. Although there are many reformulations of the transfer matrix method[7,8] which make root finding an excellent option for this purpose, we feel it is a sufficiently important topic to warrant attention. We then consider three applications for which root finding is inefficient or impossible: damped, inhomogeneous, and anisotropic media. Brevity dictates that these be presented as brief sketches, so results and comparisons with root finding are kept short or omitted. In any case, the aim of this section is not to solve an exhaustive list of propagation problems in minute detail. Instead we hope that by suggesting some ways in which the spectral method approach can be extended we can highlight its flexibility, which we consider one of its major strengths.

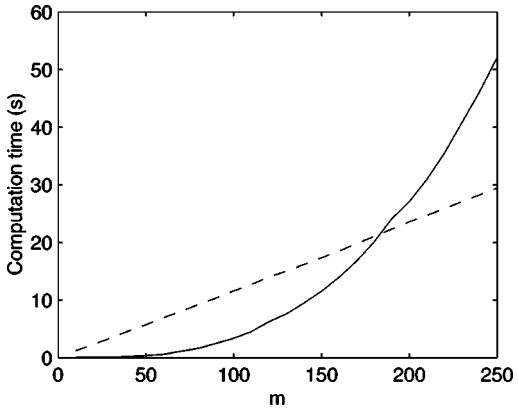


FIG. 4. Time taken to compute the first m eigenvalues to 10-digit accuracy for the in-plane problem with $\nu=10$ using the spectral method (solid line) and the root finding method (dashed line).

A. Layering

1. Two layers

We demonstrate the extension of our approach to layered media by modelling waves in a bilayered annulus with out-of-plane displacements. The case of in-plane displacements follows from the ideas of the preceding section. Consider a circular annulus composed of two layers of different homogeneous, isotropic, linearly elastic media with interface at $r=a$. The outer layer has thickness h_1, longitudinal wave speed α_1, transverse wave speed β_1, and density ρ_1. We similarly denote the properties of the inner layer by replacing 1 with 2, and require $h_2<a$ for a physical annulus. The structure is shown in Fig. 6. Let the displacements u_1 in the outer layer and u_2 in the inner layer be perpendicular to the (r,θ) plane. Each displacement satisfies the appropriate version of Eq. (1) in its own layer, with zero stress on the outer and inner surfaces and continuity of stress and displacement at the interface, i.e.,

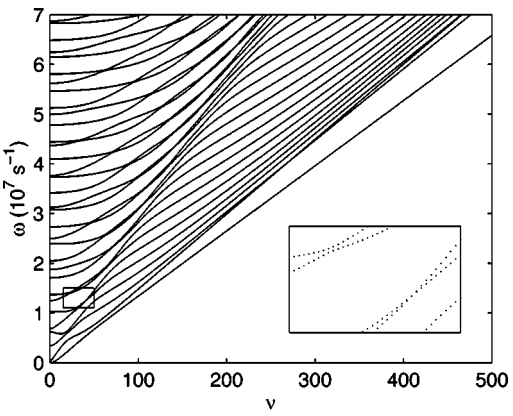


FIG. 5. The dispersion curves for the in-plane problem. All modes are shown. A step size of 1 for ν, and $N=57$ were used to give a minimum of 10-digit accuracy. The inset is a magnification of the small rectangular region shown and contains a crossover (bottom right) and a close approach (top left).

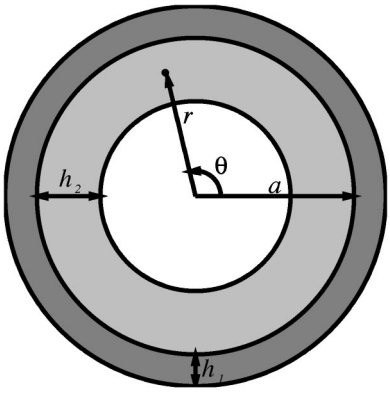


FIG. 6. The geometry of the bilayer.

$$\frac{dU_1}{dr}=0, \quad r=a+h_1,$$

$$\left[\mu\frac{dU}{dr}\right]_2^1=[U]_2^1=0, \quad r=a,$$

$$\frac{dU_2}{dr}=0, \quad r=a-h_2.$$

We construct the matrices $D_1^{(m)}$, L_1, and M_1 for the outer layer and the corresponding matrices with a subscript 2 for the inner layer. It is possible at this point to choose different numbers of interpolation points for each layer, which might be appropriate in certain cases. For example, if the function $U(r)$ varied over a considerably shorter length scale in one layer than in the other, it might be sensible to use more interpolation points in that layer to better resolve it. For the purposes of illustration we choose the same N in each layer. In this case M_1 and M_2 are identical and are referred to as M. Since there are two displacements, U_1 and U_2, to be found, we use the same technique as in the single-layer in-plane case of constructing a vector $\mathbf{u}$ whose elements are $U_1(r_{1i})$ followed by $U_2(r_{2i})$, where r_{1i} are the Chebyshev interpolation points in the outer layer and so on. Accordingly the governing equations are represented by the matrix

$$\begin{pmatrix} L_1 & 0 \\ 0 & L_2 \end{pmatrix}.$$

The boundary conditions are applied on rows 1 and $2N$ and the interface conditions on rows N and $N+1$ by forming the matrix

$$\tilde{L}=\begin{pmatrix} D_1^{(1)} & 0 \\ L_1 & 0 \\ \mu_1 D_1^{(1)},\ \text{row } N & -\mu_2 D_2^{(1)},\ \text{row } 1 \\ I,\ \text{row } N & -I,\ \text{row } 1 \\ 0 & L_2 \\ 0 & D_2^{(1)} \end{pmatrix},$$

where I is the $N\times N$ identity matrix. The matrix

$$\tilde{M}=\begin{pmatrix} M(\beta_1) & 0 \\ 0 & M(\beta_2) \end{pmatrix},$$

sets the corresponding elements of $\tilde{L}\mathbf{u}$ to zero. Our eigenvalue equation is

$$\tilde{L}\mathbf{u}=\omega^2\tilde{M}\mathbf{u}.$$

As this is a brief sketch we present neither results nor discussion of the speed and accuracy of the method, except to point out that since the matrices $\tilde{L}$ and $\tilde{M}$ have dimensions $2N\times 2N$ the QZ algorithm takes $2^3=8$ times as long as for the equivalent single layer.

2. Three or more layers

The case of an n-layered annulus where $n>2$ can be solved by building the matrices $\tilde{L}$, $\tilde{M}$ for the out-of-plane problem (or appropriate matrices for the in-plane problem) from n copies of the single-layer matrices L, M along the diagonal, and applying the boundary and interface conditions as described above. The final matrices are n times the size of their single-layer equivalents, which means that the spectral method suffers from the same drawback as the global matrix method, in that the size of the matrix grows with the number of layers. However, the greater speed of the spectral method, particularly for low mode numbers, means that it can still compete with transfer matrix methods for $n\leqslant 5$. After this point, the $O(N^3)$ number of operations required by the QZ algorithm renders the spectral method uncompetitive for plotting dispersion curves, although it may still be of use for generating displacement and stress profiles which are automatically calculated, or as an accuracy check.

B. Damping

Let us now turn our attention to wave attenuation in damped media. There are numerous constitutive models for describing such media. We shall concentrate on the widely used Kelvin–Voigt viscoelastic model in which a velocity-dependent damping force is introduced to the equation of motion (see, *inter alia*, Malvern[24] for details). This is done by replacing the Lamé constants with the operators

$$\lambda\to\lambda_r-\frac{\lambda_i}{\omega}\frac{\partial}{\partial t}, \quad \mu\to\mu_r-\frac{\mu_i}{\omega}\frac{\partial}{\partial t},$$

which, for time-harmonic disturbances, become simply $\lambda_r+i\lambda_i$ and $\mu_r+i\mu_i$. Henceforth the subscripts r and i shall denote real and imaginary parts. It follows that the wave speeds are now complex quantities,

$$\alpha\to\alpha_r+i\alpha_i\equiv\left(\frac{\lambda_r+2\mu_r+i(\lambda_i+2\mu_i)}{\rho}\right)^{1/2},$$

$$\beta\to\beta_r+i\beta_i\equiv\left(\frac{\mu_r+i\mu_i}{\rho}\right)^{1/2}.$$

This formulation results in a complex wave number, $\nu_r+i\nu_i$, the real part of which describes the propagation and the imaginary part, the attenuation of the wave. Thus for a given frequency, ω, we are now required to find two quantities instead of one.

TABLE III. The real and imaginary parts of the wave number for the first five modes for the single-layer out-of-plane problem with $\beta=1+i$, $a=1$, $b=3$, $\omega=1$ (arbitrary units) calculated using a 32-point spectral method. The modulus of the characteristic function is also shown to verify accuracy.

Mode	ν_r	ν_i	$\lvert F\rvert$
1	0.916 232 558 671 261	0.990 358 986 078 937	2.8×10^{-12}
2	0.353 687 438 706 428	2.862 805 026 186 14	1.2×10^{-13}
3	0.163 812 318 494 137	5.720 067 965 131 59	9.2×10^{-13}
4	0.107 517 269 887 583	8.579 135 171 191 1	1.2×10^{-10}
5	0.080 179 218 783 537 7	11.438 557 368 885	8.3×10^{-9}

Root finding methods encounter problems here. The task of finding complex roots of the characteristic function is substantially harder than for real roots. This difficulty is easily explained: from a given seed value there are only two directions in which a real root could lie (forwards or backwards along the real line), but a complex root could lie in any direction in the complex plane. Recent work by Castaings and Hosten[25] employs a bisection method to locate initial roots for a given frequency, and then a Newton–Raphson method to follow these roots as the frequency-changes. Experimental data is presented to validate the real part of the solution (which gives the phase velocity) but not the imaginary part (the attenuation). Unfortunately no indication of computation time is given but, as Lowe[2] points out, complex variants of root-finding procedures are typically slower and less robust than their real counterparts. In the case of Newton–Raphson, which requires both the characteristic function and its derivative to be evaluated at each step, the speed reduction is due to the additional computational expense of calculating complex derivatives.[26]

The spectral method, however, handles damped media with no loss of speed or accuracy because the QZ algorithm calculates complex eigenvalues just as easily as real ones. We simply replace the wave speeds and Lamé constants with the appropriate complex quantities and proceed as before. The only difference is that ω is now specified and ν plays the role of the eigenvalue. This requires a minor modification of our previous formulation, which we leave to the interested reader. We provide a brief demonstration for the single-layer annulus with out-of-plane displacements. Table III shows the real and imaginary parts of the first five complex wave numbers, calculated using a 32-point spectral method, for the specified problem parameters. The time taken to generate these is the same as for an undamped problem with the same number of grid points. Their accuracy is verified by evaluating the characteristic function, $F(\omega,\nu)$, of Eq. (3) to see how close to zero it is. The absolute values, $|F|$, are shown in Table III, confirming the same high accuracy seen in undamped problems.

C. Inhomogeneity

One area in which root finding cannot be used is that of inhomogeneous media. Let us again consider the case of out-of-plane displacements in a single-layer annulus. There are two relevant material parameters: the Lamé constant, μ, and the density, ρ. Instead of being constant as before, these parameters now vary with r. The transverse wave speed becomes a function of r, given by

$$\beta(r)^2=\frac{\mu(r)}{\rho(r)}.$$

For general $\beta(r)$ the governing equation, Eq. (1), is no longer amenable to analytical solution, so an exact dispersion relation, and hence characteristic function, cannot be obtained. Consequently root-finding cannot be used. The spectral method, however, can be adapted for very little extra work. We simply change the matrix M to

$$M=-\operatorname{diag}\left(\frac{1}{\beta(r)^2}\right),\quad M_{11}=0,\quad M_{NN}=0,$$

or, pictorially,

$$M=-\begin{pmatrix}0 & & & & \\ & \beta(r_2)^{-2} & & & \\ & & \ddots & & \\ & & & \beta(r_{N-1})^{-2} & \\ & & & & 0\end{pmatrix}.$$

There are two caveats here. The first is that the material parameters must be continuous functions of r. The reason for this is as follows. The problem is defined mathematically by the governing equation, Eq. (1), the boundary conditions, Eq. (2), and the physical requirements that the displacement, U, and stress, $\mu U'$, be continuous. If μ and ρ are continuous, then so is β, and Eq. (1) guarantees that U' will be too. Thus the stress continuity condition is automatically satisfied and the problem is completely expressed by Eqs. (1) and (2). This is solved by the above prescription, and we note that the solution depends only on β, i.e., on the ratio of μ to ρ and not their actual values. If, however, μ has a discontinuity, say at r_0, then the stress continuity condition is not automatically satisfied and a complete expression of the problem requires not only Eqs. (1) and (2) but also

$$[U]=[\mu U']=0,\quad r=r_0.$$

This is not solved by the above prescription as it does require the actual value of μ, but is solved by the method outlined in Sec. III A. In other words, genuinely layered media must be treated using genuinely layered methods. Somewhat perversely this means that the ostensibly more difficult case of a continuously varying medium is actually easier to treat under the spectral method approach than a layered medium. The second caveat is more straightforward: sufficiently large N must be chosen to accurately resolve the function $\beta(r)$.

Some typical results are presented in Fig. 7. The graphs on the left-hand side show the transverse wave speed and the

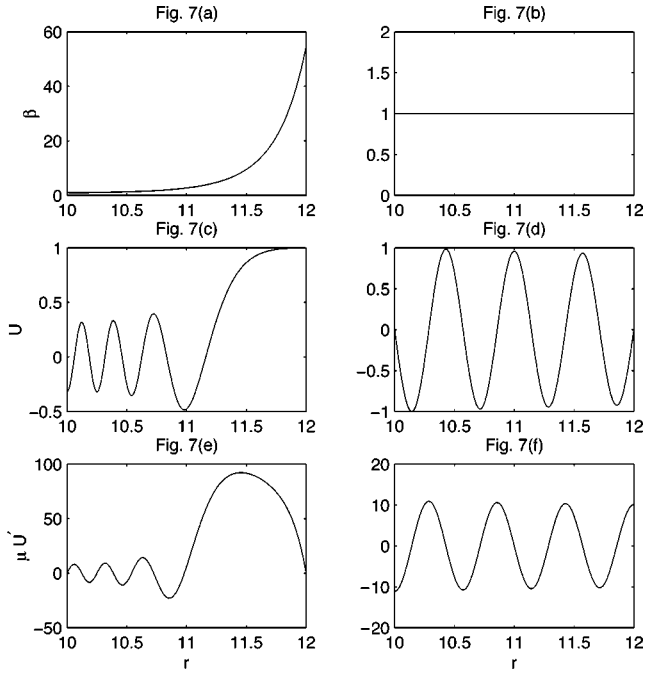


FIG. 7. (a) and (b) show the transverse wave speeds for inhomogeneous and homogeneous annuli, respectively. The inhomogeneous wave speed is given by $\beta(r)=\exp[(r-a)^2]$, the homogeneous wave speed is $\beta=1$. Both wave speeds have the same value at $r=a$. The density is $\rho=1$ in both cases. (c) and (d) show the displacement profiles for the eighth $\nu=1$ mode in each case. (e) and (f) show the corresponding stress profiles. The dimensions of the annulus are $a=10$ and $b=12$. All values are in arbitrary units.

displacement and stress profiles for a mode in an inhomogeneous annulus. Those on the right-hand side are for the same mode in a homogeneous annulus. Details of the parameters used can be found in the figure caption. For convenience we take $\rho=1$ in both cases, whence $\mu=\beta^2$. In the inhomogeneous medium β, and hence μ, remain roughly constant in the inner half of the annulus and increase rapidly in the outer half. Consequently the mode exhibits high-shear behavior in the inner region where the stiffness is low, and low-shear behavior towards the stiff outer surface, particularly for $r>11.5$ where the behavior is close to rigid. In contrast an almost homogeneous mode shape is observed in the homogeneous annulus, as one would expect.

D. Anisotropy

As a final example we consider an annulus composed of an anisotropic material. In this case it is no longer possible to decompose the motion so that the governing equations uncouple so we may not, as before, consider modes with purely out-of-plane or in-plane displacements. Instead we have to solve a system of three coupled differential equations. This was recently studied by Towfighi *et al.*[27] and we shall adopt their formulation with some minor modifications, but without derivation. The displacement field in cylindrical polars is

$$[u_r(r),u_\theta(r),u_z(r)]e^{i(\nu\theta-\omega t)}.$$

The constitutive properties of the material are defined by the stiffness matrix, C, which relates stresses, σ_{ij}, and strains, ϵ_{ij}, via the 21 independent elastic constants. Towfighi *et al.* define C as follows:

$$\begin{pmatrix}\sigma_{\theta\theta}\\ \sigma_{zz}\\ \sigma_{rr}\\ \sigma_{\theta z}\\ \sigma_{r\theta}\\ \sigma_{rz}\end{pmatrix}=\begin{pmatrix}C_{11}&C_{12}&C_{13}&C_{14}&C_{15}&C_{16}\\ C_{12}&C_{22}&C_{23}&C_{24}&C_{25}&C_{26}\\ C_{13}&C_{23}&C_{33}&C_{34}&C_{35}&C_{36}\\ C_{14}&C_{24}&C_{34}&C_{44}&C_{45}&C_{46}\\ C_{15}&C_{25}&C_{35}&C_{45}&C_{55}&C_{56}\\ C_{16}&C_{26}&C_{36}&C_{46}&C_{56}&C_{66}\end{pmatrix}\begin{pmatrix}\epsilon_{\theta\theta}\\ \epsilon_{zz}\\ \epsilon_{rr}\\ 2\epsilon_{\theta z}\\ 2\epsilon_{r\theta}\\ 2\epsilon_{rz}\end{pmatrix}.$$

It should be noted that reordering the stresses and strains results in different definitions of C which may be found elsewhere in the literature. Under this definition the equations of motion can be written as

$$\mathcal{A}u_r+\mathcal{B}u_\theta+\mathcal{D}u_z=-\rho\omega^2u_r,$$

$$\mathcal{E}u_r+\mathcal{F}u_\theta+\mathcal{G}u_z=-\rho\omega^2u_\theta,$$

$$\mathcal{H}u_r+\mathcal{J}u_\theta+\mathcal{K}u_z=-\rho\omega^2u_z,$$

where the differential operators $\mathcal{A}$–$\mathcal{K}$ are given in Appendix B. Let A–K be the $N\times N$ matrices representing these operators. These are built from Chebyshev DMs in a similar way to the original isotropic single-layer matrix, L, in Eq. (6), so we omit the details. The governing equations are represented by the following $3N\times 3N$ matrix system:

$$\left(\begin{array}{c|c|c}A&B&D\\ \hline E&F&G\\ \hline H&J&K\end{array}\right)\begin{pmatrix}u_r(r_1)\\ \vdots\\ u_r(r_N)\\ \hline u_\theta(r_1)\\ \vdots\\ u_\theta(r_N)\\ \hline u_z(r_1)\\ \vdots\\ u_z(r_N)\end{pmatrix}$$

We denote the large matrix by X and the vector by $\mathbf{u}$. The boundary conditions are

$$\left[\frac{C_{33}}{r}\frac{d}{dr}+C_{13}+iC_{35}p\right]u_r+\left[\frac{C_{35}}{r}\frac{d}{dr}-C_{35}+iC_{13}p\right]u_\theta+\left[\frac{C_{36}}{r}\frac{d}{dr}+iC_{34}p\right]u_z=0,$$

$$\left[\frac{C_{35}}{r}\frac{d}{dr}+C_{15}+iC_{55}p\right]u_r+\left[\frac{C_{55}}{r}\frac{d}{dr}-C_{55}+iC_{15}p\right]u_\theta+\left[\frac{C_{56}}{r}\frac{d}{dr}-iC_{45}p\right]u_z=0,$$

$$\left[\frac{C_{36}}{r}\frac{d}{dr}+C_{16}+iC_{56}p\right]u_r+\left[\frac{C_{56}}{r}\frac{d}{dr}-C_{56}+iC_{16}p\right]u_\theta+\left[\frac{C_{66}}{r}\frac{d}{dr}+iC_{46}p\right]u_z=0.$$

These can also be written as a $3N\times 3N$ matrix system: $S\mathbf{u}=\mathbf{0}$. We take the rows 1, N, $N+1$, $2N$, $2N+1$, and $3N$ of S, which correspond to the boundary conditions evaluated at $r=a$, b, and insert them into X, denoting the resulting matrix $\tilde{X}$. Finally we construct the matrix,

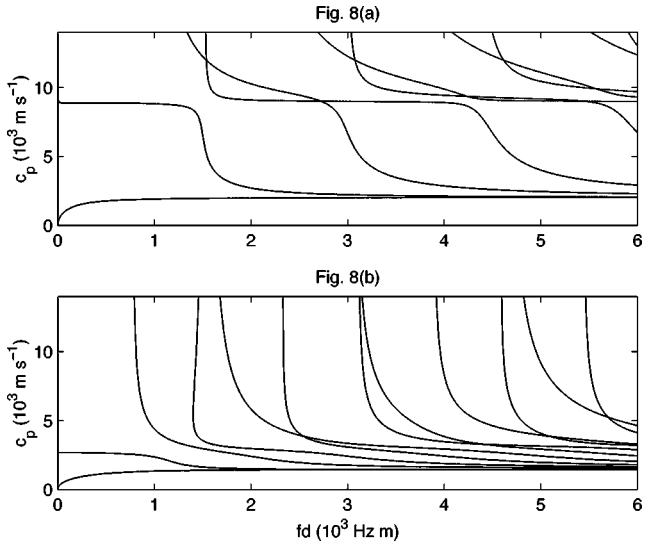


FIG. 8. Dispersion curves (in phase velocity versus frequency-thickness format) for an annulus of inner radius 1 m and thickness 1 mm composed of a unidirectional composite whose fibers are aligned in the θ direction (a) and the z direction (b). Stiffness matrices for each case are given in Appendix C and $\rho=1580$ kg/m^3 in both cases. Each plot was generated in approximately 30 s using a 16-point spectral method with a variable ν step size to give unbroken curves.

$$Y=\rho\begin{pmatrix} M(1) & 0 & 0 \\ 0 & M(1) & 0 \\ 0 & 0 & M(1) \end{pmatrix},$$

which sets the boundary conditions to zero, and solve the eigenvalue equation,

$$\tilde{X}\mathbf{u}=\omega^2 Y\mathbf{u},$$

the form of which is by now familiar.

Towfighi *et al.* solved this by a completely new approach. They performed Fourier series expansions of the displacements, used a method of weighted residuals to turn the governing equations into a set of linear equations, and then found the roots of an appropriate matrix determinant. The dispersion curves obtained by this method showed good agreement with the known curves for anisotropic flat plates (in Rose[28]) in the small h/a limit but, as they noted, contained many broken lines. Unfortunately no indication was given of computation time for this method. Figure 8, generated by a 16-point spectral method in approximately 60 s, shows two sets of dispersion curves for a unidirectional composite where the material fibers are oriented along the θ direction [Fig. 8(a)] and the z direction [Fig. 8(b)] for an annulus with $h/a=10^{-3}$. The stiffness matrices for each of these orientations are given in Appendix C. Both plots show excellent agreement with the corresponding flat plate curves in Rose (not reproduced here). Indeed, fine detail such as the $fd\to 0$ behavior of the lowest two modes, which is absent in the plots by both Rose and Towfighi *et al.*, becomes evident.

IV. CONCLUSIONS

We have shown that solving the underlying ordinary differential equations using a numerical scheme based on spectral methods presents an attractive alternative to root finding for calculating the dispersion curves and displacement/stress profiles of guided waves in elastic media.

In Sec. II the method is applied to a circular, single-layered annulus. Large (up to 100-fold) efficiency savings over existing techniques are demonstrated for finding the first 100 modes or fewer. Further advantages include: low coding effort in MATLAB; automatic generation of mode shapes; no eigenvalues missed; no need to manipulate special functions.

In Sec. III we present extensions of the method to demonstrate its flexibility. One extension is to layered media, where the method proves computationally efficient for fewer than five layers. Other extensions are to areas where existing techniques suffer problems or fail completely, namely damped, inhomogeneous, and anisotropic media.

There is some scope for further work. So far only modal solutions have been found, but it may be possible to use the method to find response solutions as well. Furthermore, the curved structure chosen here is particularly simple as its geometry is separable: an investigation of more general guiding structures, such as that in Gridin *et al.*,[29] shows that the spectral method can be used for nonseparable geometries too. Finally, the high speed and accuracy of the method may make it a convenient tool for solving time-dependent problems, where the modal problem is solved repeatedly, once for each time step; we have recently implemented this.

ACKNOWLEDGMENTS

The authors would like to thank Mike Lowe, Jimmy Fong, and Francesco Simonetti of the Imperial College NDT group and Dmitri Gridin, our colleague in the Mathematics Department, for sharing their expertise during many helpful and enjoyable discussions.

APPENDIX A: MATLAB CODE FOR OUT-OF-PLANE ANNULUS

The MATLAB program `chebdif` is from the software suite by Weideman and Reddy[17] and can be downloaded at ftp.mathworks.com/pub/contrib/v5/diffeq/dmsuite.

```
% set problem parameters
N=32; beta=3260; a=20e-3; b=23e-3; nu=1;
% generate Chebyshev differentiation matrices
[x,D]=chebdif(N,2);
h=b-a;
r=(h*x+b+a)/2; % r-coordinate
D1=(2/h)*D(:,:,1); % first r-derivative
D2=(2/h)^2*D(:,:,2); % second r-derivative
% construct matrices L and M
L=D2+diag(r.^-1)*D1-nu^2*diag(r.^-2);
L(1,:)=D1(1,:)/a; % boundary
L(N,:)=D1(N,:)/a; % conditions
M=-beta^-2*eye(N); M(1,1)=0; M(N,N)=0;
% solve system using QZ algorithm
[U,E]=eig(L,M);
[w,ii]=sort(sqrt(diag(E)));
U=U(:,ii);
% w is a vector containing the frequency
```

% eigenvalues in ascending order, U is a
% matrix whose nth column is the eigenvector
% (i.e., displacement) for the eigenvalue w(n)

APPENDIX B: ANISOTROPIC GOVERNING EQUATIONS

The operators $\mathcal{A}$–$\mathcal{K}$ are as follows:

$$\mathcal{A}=C_{33}\frac{d^2}{dr^2}+(C_{33}+2iC_{35}p)\frac{1}{r}\frac{d}{dr}-(C_{55}p^2+C_{11})\frac{1}{r^2},$$

$$\mathcal{B}=C_{35}\frac{d^2}{dr^2}+[-C_{15}+i(C_{13}+C_{55})p]\frac{1}{r}\frac{d}{dr}$$
$$-[C_{15}p^2+i(C_{11}+C_{55})p-C_{15}]\frac{1}{r^2},$$

$$\mathcal{D}=C_{36}\frac{d^2}{dr^2}+[C_{36}-C_{16}+i(C_{34}+C_{56})p]\frac{1}{r}\frac{d}{dr}$$
$$-(C_{45}p^2+iC_{14}p)\frac{1}{r^2},$$

$$\mathcal{E}=C_{35}\frac{d^2}{dr^2}+[C_{15}+2C_{35}+i(C_{13}+C_{55})p]\frac{1}{r}\frac{d}{dr}$$
$$-[C_{15}p^2-i(C_{11}+C_{55})p-C_{15}]\frac{1}{r^2},$$

$$\mathcal{F}=C_{55}\frac{d^2}{dr^2}+(C_{55}+2iC_{15}p)\frac{1}{r}\frac{d}{dr}-(C_{11}p^2+C_{55})\frac{1}{r^2},$$

$$\mathcal{G}=C_{56}\frac{d^2}{dr^2}+[2C_{56}+i(C_{16}+C_{45})p]\frac{1}{r}\frac{d}{dr}$$
$$-(C_{14}p^2-iC_{45}p)\frac{1}{r^2},$$

$$\mathcal{H}=C_{36}\frac{d^2}{dr^2}+[C_{16}+C_{36}+i(C_{34}+C_{56})p]\frac{1}{r}\frac{d}{dr}$$
$$-(C_{45}p^2-iC_{14}p)\frac{1}{r^2},$$

$$\mathcal{J}=C_{56}\frac{d^2}{dr^2}+i(C_{16}+C_{45})p\frac{1}{r}\frac{d}{dr}-(C_{14}p^2+iC_{45}p)\frac{1}{r^2},$$

$$\mathcal{K}=C_{66}\frac{d^2}{dr^2}+(C_{66}+2iC_{46}p)\frac{1}{r}\frac{d}{dr}-C_{44}p^2\frac{1}{r^2}.$$

APPENDIX C: STIFFNESS MATRICES FOR MATERIALS IN FIG. 8

The stiffness matrices for the materials in Figs. 8(a) and 8(b) are

$$S_a=\begin{pmatrix} 128.2 & 6.9 & 6.9 & 0 & 0 & 0 \\ 6.9 & 14.95 & 7.33 & 0 & 0 & 0 \\ 6.9 & 7.33 & 14.95 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6.73 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_b=\begin{pmatrix} 14.95 & 6.9 & 7.33 & 0 & 0 & 0 \\ 6.9 & 128.2 & 6.9 & 0 & 0 & 0 \\ 7.33 & 6.9 & 14.95 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.81 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

All elastic constants are in GPa.

[1] D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247 (1997).
[2] M. J. S. Lowe, "Matrix techniques for modeling ultrasonic waves in multilayered media," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 525 (1995).
[3] W. T. Thomson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89 (1950).
[4] N. A. Haskell, "Dispersion of surface waves on multilayered media," Bull. Seismol. Soc. Am. **43**, 17 (1953).
[5] L. Knopoff, "A matrix method for elastic wave problems," Bull. Seismol. Soc. Am. **54**, 431 (1964).
[6] J. W. Dunkin, "Computation of modal solutions in layered, elastic media at high frequencies," Bull. Seismol. Soc. Am. **55**, 335 (1965).
[7] L. Wang and S. I. Rokhlin, "Stable reformulation of transfer matrix method for wave propagation in layered anisotropic media," Ultrasonics **39**, 413 (2001).
[8] B. Hosten and M. Castaings, "Surface impedance matrices to model the propagation in multilayered media," Ultrasonics **41**, 501 (2003).
[9] H. O. Kreiss and J. Oliger, "Comparison of accurate methods for the integration of hyperbolic equations," Tellus **24**, 199 (1972).
[10] S. A. Orszag, "Comparison of pseudospectral and spectral approximation," Stud. Appl. Math. **51**, 253 (1972).
[11] B. Fornberg, "On a Fourier method for the integration of hyperbolic equations," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **12**, 509 (1975).
[12] J. M. Carcione, G. C. Herman, and A. P. E. ten Kroode, "Seismic modeling," Geophysics **67**, 1304 (2002).
[13] J. F. Doyle, *Wave Propagation in Structures: an FFT-based Spectral Analysis Methodology* (Springer-Verlag, New York, 1989).
[14] B. Fornberg, *A Practical Guide to Pseudospectral Methods* (Cambridge University Press, Cambridge UK, 1996).
[15] L. N. Trefethen, *Spectral Methods in MATLAB* (SIAM, Philadelphia, 2000).
[16] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, 2nd ed. (Dover, New York, 2001).
[17] J. A. C. Weideman and S. C. Reddy, "A MATLAB differentiation matrix suite," ACM Trans. Math. Softw. **26**, 465 (2000).
[18] G. Liu and J. Qu, "Guided circumferential waves in a circular annulus," J. Appl. Mech. **65**, 424 (1998).
[19] C. Valle, J. Qu, and L. J. Jacobs, "Guided circumferential waves in layered cylinders," Int. J. Eng. Sci. **37**, 1369 (1999).
[20] C. Valle, M. Niethammer, J. Qu, and L. J. Jacobs, "Crack characterization using guided circumferential waves," J. Acoust. Soc. Am. **110**, 1282 (2001).
[21] D. Gridin *et al.*, "The high-frequency asymptotic analysis of guided waves in a circular elastic annulus," Wave Motion **38**, 69 (2003).
[22] D. Gottlieb and S. A. Orszag, *Numerical Analysis of Spectral Methods: Theory and Applications* (SIAM, Philadelphia, 1977).
[23] J. D. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland,

Amsterdam, 1973).
[24] L. E. Malvern, *Introduction to the Mechanics of a Continuous Medium* (Prentice-Hall, Englewood Cliffs, NJ, 1969).
[25] M. Castaings and B. Hosten, ''Guided waves propagating in sandwich structures made of anisotropic, viscoelastic, composite materials,'' J. Acoust. Soc. Am. **113**, 2622 (2003).
[26] A. Neumaier, *Introduction to Numerical Analysis* (Cambridge University Press, Cambridge UK, 2001).
[27] S. Towfighi, T. Kundu, and M. Ehsani, ''Elastic wave propagation in circumferential direction in anisotropic cylindrical curved plates,'' J. Appl. Mech. **69**, 283 (2002).
[28] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge UK, 1999).
[29] D. Gridin, A. T. I. Adamou, and R. V. Craster, ''Electronic eigenstates in quantum rings: Asymptotics and numerics,'' Phys. Rev. B **69**, 155317 (2004).

Thin, lightweight electroacoustic projector for low frequency underwater applications[a)]

James F. Tressler[b)]
Naval Research Laboratory, Washington, DC 20375

Thomas R. Howarth
Naval Sea Systems Command, Division Newport, Newport, Rhode Island 02841

Walter L. Carney
Naval Sea Systems Command, Division Crane, Crane, Indiana 47522

(Received 25 June 2003; revised 30 April 2003; accepted 12 May 2004)

A thin, lightweight oil-filled electroacoustic projector has been designed and fabricated for application on unmanned underwater vehicles (UUVs) and autonomous unmanned vehicles (AUVs) used in littoral environments. The projector has dimensions of 381 mm×280 mm×64 mm and weighs 26 N in water. The acoustic aperture consists of 294 cymbal-type miniature Class V flextensional drivers. These 294 elements are divided into six independent modules of 49 cymbals wired together electrically in parallel. The six modules, in turn, are wired so that they can be easily combined together electrically either in parallel, series, or as a combination of both. Based on the VA power curve and a low duty cycle input voltage, an acoustic source level of at least 180 dB (re: 1 μPa @ 1 m) is attainable across the frequency span ~3–15 kHz. This particular design is best suited for shallow water, but is shown to still be operable at water depths up to 200 meters without significant degradation in performance. [DOI: 10.1121/1.1778742]

PACS numbers: 43.38.Fx, 43.38.Hz, 43.30.Yj [AJZ] Pages: 1536–1543

I. INTRODUCTION

This paper describes the design, fabrication procedure, and underwater acoustic performance of a lightweight, thin profile, low frequency underwater sound projector. The motivation behind this work derives from the interest of the U.S. Navy in gathering broadband data in littoral environments, for which small, unmanned underwater vehicles (UUVs) and autonomous unmanned vehicles (AUVs) are best suited.[1,2] Due to the limited amount of real estate available to house the projector on these vehicles, it is critical to minimize the volumetric projector dimensions.

Of the potential broadband acoustic source candidates that can be utilized for this application, 1–3 type piezocomposites meet the dimensional requirements; unfortunately, their acoustic output at frequencies below 10 kHz is lower than desired.[3] Other technologies used to generate high power sound waves at frequencies under 30 kHz include free-flooded piezoelectric ceramic rings, electromagnetic drivers, tonpilz transducers, and flextensional devices.[4,5] However, because of their large size and weight, these technologies are not easily adaptable for mounting on the smaller advanced underwater vehicle platforms of AUVs and UUVs.

The challenge has thus been to design a thin profile acoustic projector with adequate source level at frequencies below 30 kHz. It is extremely difficult to design a single projector to cover this entire band. The particular projector construction described in this paper measures 381 mm×280 mm×64 mm, weighs 26 N in water, and is capable of delivering an acoustic output of at least 180 dB (re: 1 μPa @ 1 m) over an extended frequency range below 30 kHz. It is based upon the Class V flextensional transducer.[6]

II. PROJECTOR DESIGN

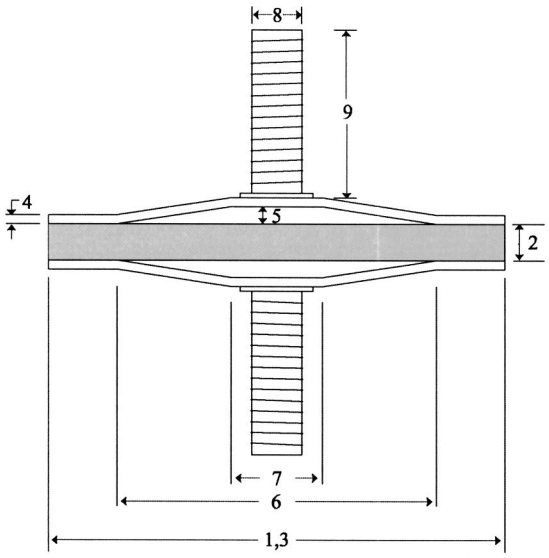
A. Cymbal drivers

Miniaturizing the projector volume makes necessary the use of a vibrating structure that can produce large displacements.[7] This projector utilizes 294 miniature Class V flextensional electro-mechanical drivers that are known as cymbals.[8,9] These devices are very small compared to the wavelength of sound in water (over our frequency range of interest) and amplify the displacement of the ceramic by over an order of magnitude.[9] A cross-sectional sketch, accompanied by a photograph, of the cymbal driver is shown in Fig. 1.

Table I provides the dimensions of the cymbals used in this projector. The active material in each cymbal is a PZT-5H (a.k.a. Navy Type VI)[10] piezoelectric ceramic disk poled in its thickness direction. PZT-5H was selected for this study because of its very large piezoelectric d_{31} coefficient (-250×10^{-12} m/V, as compared to -135×10^{-12} m/V for PZT-4 and -105×10^{-12} m/V for PZT-8).[11] Consequently, since our voltage supply was limited, PZT-5H would generate the largest displacement. Titanium endcaps with the geometry and dimensions shown in Fig. 1 and Table I, respectively, are coupled both mechanically and electrically in parallel to the top and bottom faces of the piezoelectric ce-

[a)]Portions of this work were presented in "Thin, Low Frequency, Oil-Filled Projectors for UUV Applications," Proceedings of the Oceans 2001 MTS/IEEE Conference, Honolulu, HI, November 2001.

[b)]Electronic mail: tressler@pa.nrl.navy.mil



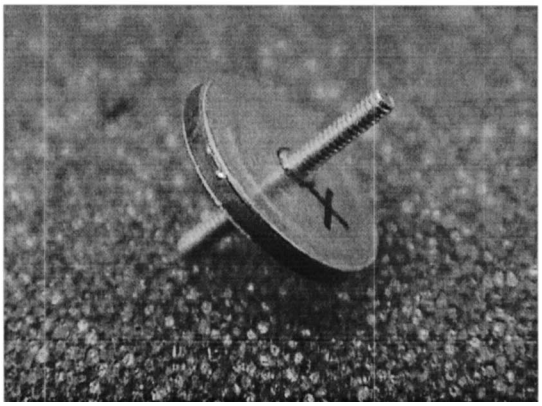


FIG. 1. Cross-sectional sketch of the 'cymbal' electromechanical flextensional driver (top), along with a photograph of an assembled device (bottom). The numerals in the top diagram correspond to the dimensions given in Table I.

ramic disk via an electrically conductive adhesive. Studs with UNF-080 threads are microwelded onto the apex of each of the endcaps. The studs, in conjunction with the endcaps, serve as the electrical conduit from the piezoelectric ceramic to the electrical leads. In addition, by adding additional nuts (i.e., mass loading) to each side of the cymbal, its fundamental resonance frequency can be further reduced. The effect of the number of nuts on the resonance frequency,

TABLE I. Dimensions of the cymbal flextensional driver, where the numbers proceeding the parameter correspond to those in Fig. 1.

Parameter	Dimension (mm)
1—ceramic diameter	12.7
2—ceramic thickness	1.0
3—cap diameter	12.7
4—cap thickness	0.25
5—cavity depth	0.28
6—cavity diameter	10.2
7—cap apex diameter	2.3
8—stud diameter	1.45
9—stud length	6.5

TABLE II. The effect of the number of nuts (i.e., added mass) on the fundamental resonance frequency for a 49-element oil-filled projector. Also shown is the TVR amplitude at said resonance frequency.

Nuts/side	Washers/side	f_r (kHz)	TVR@f_r (dB re: 1 μPa/V @1 m)
1	1	9.9	142
3	1	7.2	136
4	1	6.6	135
5	1	5.8	135

as well as transmitting voltage response at resonance, for a prototype 49-element oil-filled projector (such as seen in Fig. 15) is shown in Table II. The addition and/or subtraction of nuts and washers has been shown to effect the radiation mass of the elements.[12] Rather subtle changes in the mechanical configurations have shown frequency shifts and acoustic output variations. The full control of these parameters has not yet been established, as it remains an issue of future study.

B. Mounting

The cymbal drivers are mounted within a molded sheet of 5 mm thick neoprene rubber.[12,13] Recessed cavities within the neoprene sheet hold the cymbals in place around their flat rim on the circumference of the disk. Each neoprene sheet measures 101 mm by 101 mm and holds 49 cymbals in a 7×7 arrangement, as seen in Fig. 2. The 49 cymbals are connected electrically in parallel using 0.026 mm thick nickel ribbon attached onto the studs by nuts (Fig. 3). The projector consists of six of these 49 cymbal element modules combined together in a 2×3 arrangement. A thin Plexiglas® sheet that has a 9.525 mm diameter hole over each of the cymbals is bonded to the upper and lower surfaces of the neoprene sheets (Fig. 4). The Plexiglas® sheet provides a means of clamping the neoprene rubber modules to the projector housing while assuring that the cymbal drivers remain in the same plane. In addition, the presence of the holes in the Plexiglas® allows the cymbal caps to flex freely without interference. The completed caster oil-filled projector assembly is shown in Fig. 5.

C. Wiring configurations

The six 49 cymbal element modules were, in turn, wired together such that they could be combined electrically either in parallel, in series, or in a parallel–series combination without the need to disassemble the projector for rewiring. A

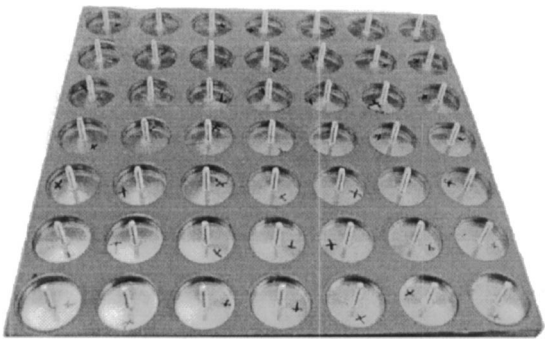


FIG. 2. Cymbal drivers held in place within a sheet of neoprene rubber.

FIG. 3. Electrical lead ribbon attached to the cymbal drivers with washers and nuts. Note that all 49 cymbal elements are electrically in parallel.

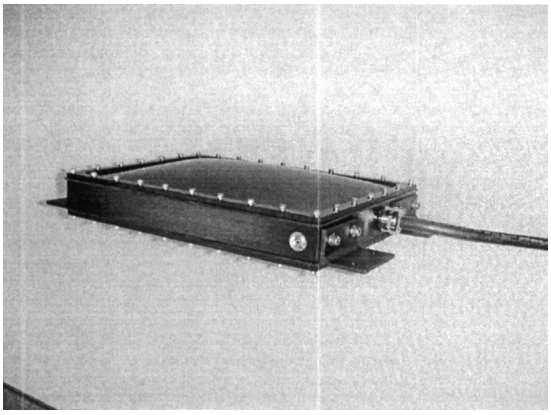


FIG. 5. The completed projector assembly after being filled with oil.

conception of the three wiring configurations is given in Fig. 6. A cable was attached to the projector frame via a liquid-tight connector. This cable contained six pairs of lead wires (one pair from each module), thus allowing for an easy change from parallel to series or to parallel-series wiring configurations.

III. IN-AIR CHARACTERIZATION

Prior to underwater calibration, the projector was characterized in air with no cable attached. The capacitance and dielectric loss factor (tan δ) were measured at 1 kHz for the parallel wiring configuration and are provided in Table III. The mechanical quality factor Q_M and electrical quality factor Q_E were calculated from the equations below[14]

$$Q_M = \frac{f_S}{f_2 - f_1}, \tag{1}$$

where f_S is the series resonance frequency. The frequencies f_2 and f_1 ($f_2 > f_1$) are the frequencies at which the motional contribution to the conductance (G_{mS}) is half its maximum value.[14] The electrical Q_E was calculated from the conductance (G_{mS}) and susceptance (B_S) values at resonance as

$$Q_E = \frac{B_S}{G_{mS}}. \tag{2}$$

Since the Q_M fell between 10 and 50, the effective coupling coefficient k_{eff} was subsequently calculated from the mechanical Q_M and electrical Q_E as[15]

$$k_{\text{eff}} = \sqrt{\frac{1}{Q_M Q_E + 1}}. \tag{3}$$

The low coupling coefficient is a primary contributor to the low electroacoustic efficiency of the projector, as will be seen in the next section.

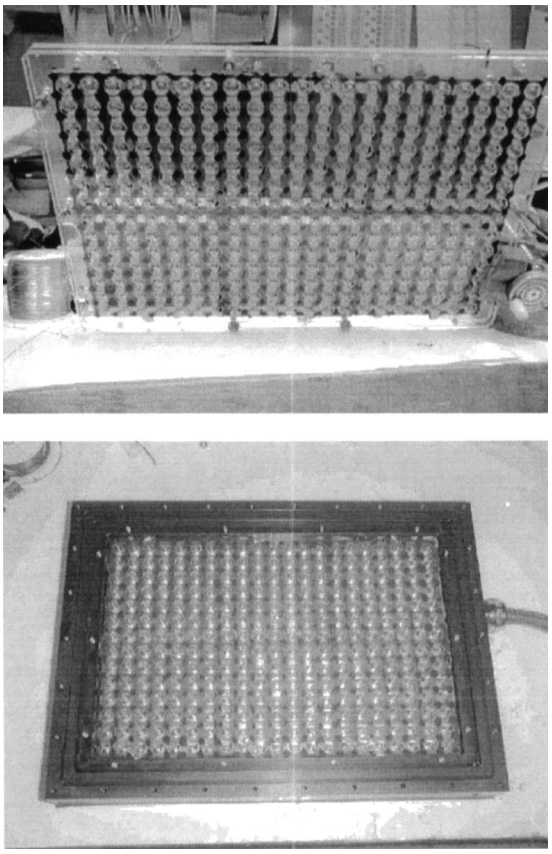


FIG. 4. The six cymbal-filled neoprene rubber modules sandwiched between two Plexiglas® plates (top), and after being inserted into the projector housing (bottom).

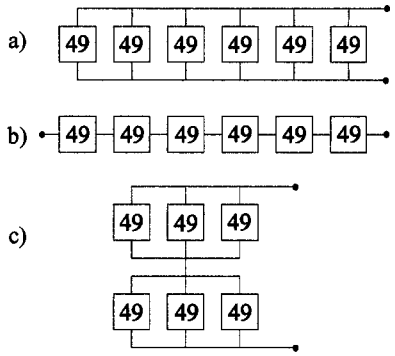


FIG. 6. Wiring concept for the six cymbal-filled neoprene rubber modules comprising the projector, where the six individual modules are (a) in electrical parallel, (b) in electrical series, and (c) in a parallel–series combination.

TABLE III. In-air characteristics of the cymbal-based oil-filled projector for the parallel wiring configuration.

C @ 1 kHz	tan δ @ 1 kHz	f_S	Q_m	Q_e	k_{eff}
848 nF	0.014	6.6 kHz	26	1.6	0.15

IV. UNDERWATER CALIBRATION

A. Measurement technique

The first set of underwater calibration measurements was performed at the Naval Sea Systems Command, Division Crane, Glendora Lake Facility[16] in Sullivan, Indiana in June 2001. The lake temperature was 16 °C. Measurements were taken with the projector held at its two top corners with nylon wire wraps. Data were taken with the projector hanging in both the elevation and azimuthal directions. The receiver hydrophone was positioned at the same depth as the projector (6.1 meters) with a separation distance of 8.3 meters. The applied electrical drive signal was a chirp with a one percent duty cycle. The amplitude of the signal was such that each 49 cymbal element module experienced $\sim 100\,V_{rms}$ across the entire frequency band. The receive signal was gated to eliminate interference from reflections off the surface of the water. All measurements taken at Glendora Lake were conducted at frequencies below 10 kHz.

The projector was recalibrated in September 2002 in the Acoustic Test Facility[17] (Code 8211) of the Naval Sea Systems Command, Division Newport, Newport, Rhode Island. The size of the test tank is 18 meters×12 meters×11.5 meters and no two sides are parallel. The water temperature was 21 °C. Measurements were made from 1 kHz to 100 kHz. When comparing the transmitting power curves measured at the Crane and Newport facilities over the common frequency band (3 kHz–10 kHz), the difference was less than 1 dB.

B. Calibration results

The transmitting voltage response (TVR) and the transmitting current response (TCR) of the projector from 1 kHz to 100 kHz are shown in Fig. 7. The measured TVR data approximate the classic projector characteristic of rising at +12 dB per octave below resonance (stiffness controlled region). The first resonance, at 6.4 kHz, is associated with the (0,1), or fundamental flexural, mode of the cymbal caps.[18,19] The null in the response at 20 kHz is due to the (0,2) vibration mode of the cymbal caps.[18,19] In this mode, portions of the cap are vibrating out of phase with each other resulting in partial sound cancellation and consequently a decrease in the TVR. A second resonance, at ~40 kHz, is attributed to the "breathing" mode of the cymbal element.[19] In this mode, the cymbal caps are vibrating in phase with the piezoelectric ceramic disk.

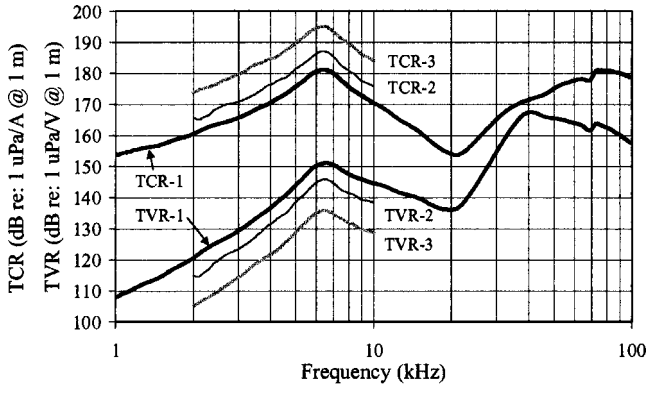


FIG. 7. Transmitting voltage response (TVR) and transmitting current response (TCR) of the projector for the three different wiring configurations: 1=parallel, 2=parallel–series, and 3=series.

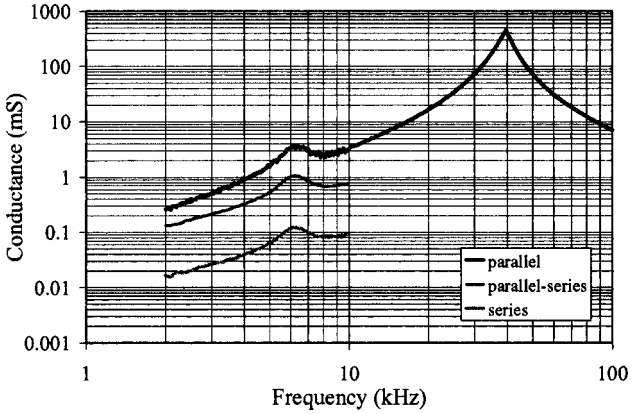


FIG. 8. In-water conductance (G) of the projector for the three different wiring configurations.

The TVR is a function of the projector's directivity index (DI), electroacoustic efficiency (η), and electrical conductance (G) as per Eq. (4),[20]

$$\mathrm{TVR} = 10\log(G) + 170.8 + \mathrm{DI} + \eta_{(\mathrm{dB})}. \tag{4}$$

As will be discussed later, the directivity index and electroacoustic efficiency are essentially the same regardless of whether the six modules are wired together in parallel, in series, or as a parallel–series combination. However, the electrical conductance is strongly dependent upon how the modules are wired together. Figure 8 shows the in-water conductance as a function of frequency for the three different wiring configurations. The data has been corrected to eliminate the effects of the attached cable. From the conductance curves, the in-water Q_M is estimated to be about 2.4, which is approximately an order of magnitude less than its in-air value. The large peak in conductance at ~40 kHz is again attributed to the radial displacement of the piezoelectric ceramic disks being in phase with the axial displacement of the cymbal caps.

Table IV lists the transmitting voltage response, in-water conductance, and electroacoustic efficiency of the projector

TABLE IV. Transmitting voltage response, conductance, and electroacoustic efficiency of the projector measured at its fundamental in-water resonance frequency of 6.4 kHz.

Wiring scheme	TVR (dB re: 1 μPa/V @ 1 m)	Conductance G (mS)	Efficiency (η) (dB)	Efficiency (η) pct.
Parallel	151.12	3.607	−3.68	43
Series	135.97	0.121	−4.08	39
Parallel series	145.91	1.047	−3.51	45

TABLE V. Expected and measured ratios of the conductance (G) at 6.4 kHz for the three different projector wiring configurations. The correlation of this ratio to the difference in the measured TVR for the two wiring configurations being compared is also shown.

	Expected ratio[a]	Measured ratio	$10\log_{10}$ of measured ratio$+\Delta\eta$ (dB)$=\Delta$TVR expected			ΔTVR measured	Error
$G_{(p)}/G_{(s)}$	36	29.81	14.74	0.40	15.14	15.15	0.01
$G_{(p)}/G_{(p-s)}$	4	3.45	5.37	−0.17	5.20	5.21	0.01
$G_{(p-s)}/G_{(s)}$	9	8.65	9.37	0.57	9.94	9.94	0.00

[a]subscript key: (p)=parallel, (s)=series, $(p\text{-}s)$=parallel–series.

at its fundamental resonance frequency of 6.4 kHz for the three wiring schemes. This data is subsequently used in Table V to explain the difference in the TVR as seen in Fig. 7. The first column of data in Table V shows the expected, or theoretical, ratio of the conductance between two wiring configurations. The second column is the measured ratio based on the data provided in Table IV. The sum of 10 times the common logarithm of this measured ratio and the respective difference in efficiency is then compared to the difference in TVR as measured at the resonance frequency. The final column of data shows the difference between the expected and measured difference in TVR. The difference in efficiency from one wiring scheme to another is quite small. Consequently, it is the change in conductance from one wiring configuration to another that primarily accounts for the differences in the TVR curves as seen in Fig. 7. The transmitting current response seen in Fig. 7 likewise approximates the classic projector characteristic of rising at +6 dB per octave below resonance and decreasing at −6 dB per octave above resonance. As opposed to the TVR, the TCR increases in going from a parallel to parallel–series to series wiring scheme. The constitutive equation for the TCR has the same form as Eq. (4), but with TVR replaced by TCR, and G replaced by R, the electrical resistance. Whereas the conductance decreases in going from parallel to parallel–series to series, the resistance increases (Fig. 9). This accounts for the increase in TCR in going from a parallel to parallel–series to series wiring configuration.

The electrical resistance of the projector (after correcting for the cable), along with the electrical impedance phase angle, θ_z, is plotted in Fig. 9. The difference in resistance is consistent with the parallel/series/parallel–series wiring configurations. The data show that the series or parallel–series wiring scheme may be desirable from a drive amplifier viewpoint. The phase angle is independent of wiring configuration and goes towards 0° at 40 kHz, which is indicative of the radial mode of the piezoelectric ceramic disks being in phase with the flexural motion of the cymbal caps.

The transmitting power response (TPR) is related to the directivity index and electroacoustic efficiency as per Eq. (5),[21]

$$\mathrm{TPR}=170.8+\mathrm{DI}+\eta_{(\mathrm{dB})}. \tag{5}$$

Since the directivity index and electroacoustic efficiency are essentially the same for all three wiring configurations, the TPR curves would be expected to coincide. This is in fact what is observed in Fig. 10.

The acoustic source level (SL) is related to the transmitting voltage and current responses as[20]

$$\mathrm{SL}=\mathrm{TVR}+20\log(V_{\mathrm{in}}), \tag{6a}$$

$$\mathrm{SL}=\mathrm{TCR}+20\log(I_{\mathrm{in}}) \tag{6b}$$

where V_{in} and I_{in} are the rms input drive voltage and current, respectively, applied to the projector.

Figure 11(a) shows the voltage and current necessary (for the all-in-parallel case) to achieve an acoustic source level of 180 dB (re: 1 μPa @ 1 m) at a given frequency. The curves were calculated by solving Eqs. (6) for V_{in} and I_{in}, respectively, where the TVR and TCR values are known from Fig. 7 and the SL is fixed at 180 dB. Figure 11(b) plots the product of the voltage and current. Measured data points are presented in the three plots as a validation for the calculations. Note that the input voltage into the PZT-5H piezoelectric ceramic is higher than the maximum recommended

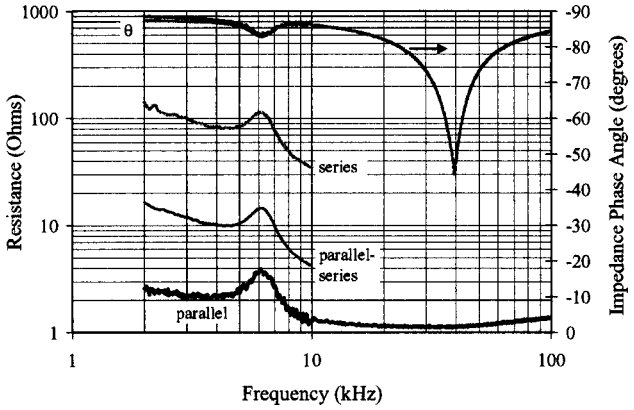


FIG. 9. Electrical resistance and impedance phase angle of the projector for the three different wiring configurations.

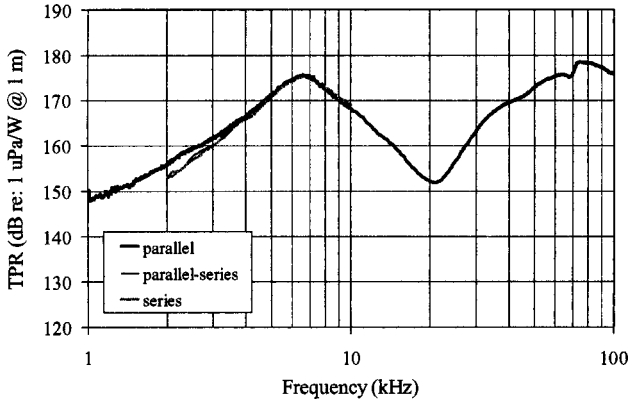


FIG. 10. Transmitting power response (TPR) of the projector for the three different wiring configurations.

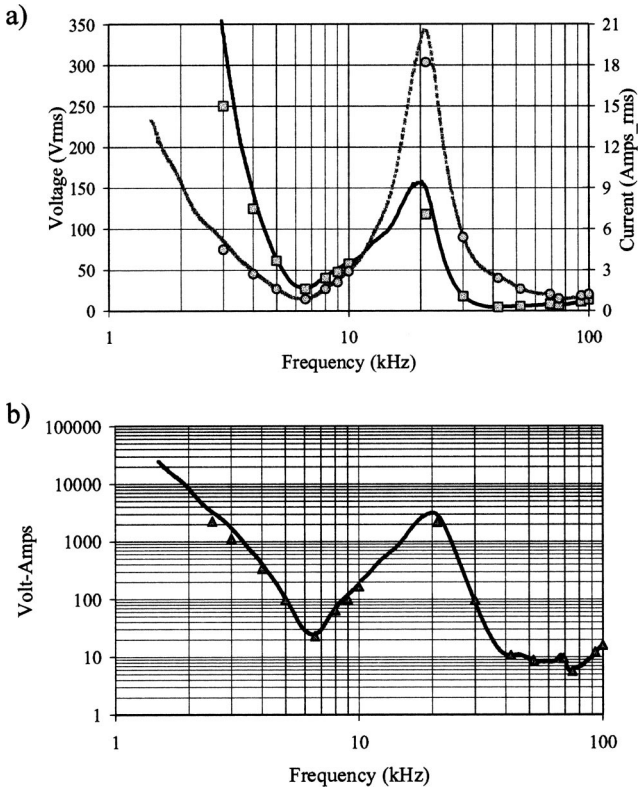


FIG. 11. Calculated (a) input voltage (solid line) and current (dotted line) and (b) input power required to attain an acoustic source level of 180 dB (re: 1 μPa @ 1 m). Measured data points (squares, voltage; circles, current; triangles, power) are shown for comparison.

drive level of 200 V rms.[22] This is because the recommended maximum drive level of 200 V rms is for a duty cycle of 5–10%, whereas in our case a duty cycle of only 1% was used—allowing for a higher input voltage.

A plot of the acoustic source level at the projector resonance frequency as a function of applied voltage is given in Fig. 12. The dashed line gives the expected values and was calculated from Eq. (6a), where the TVR value at resonance was taken from Fig. 7. The source level as interpolated from measured data is shown by the solid line. The data show that the acoustic output is approximately linear with electrical input voltage. This linearity begins to deviate from the calculated expectation when the source level exceeds 190 dB. The reason for this deviation is that the PZT-5H material has a higher hysteresis than other types of piezoelectric ceramic. This indicates that for applications where a source level above 190 dB is needed, perhaps a piezoelectrically harder PZT should be considered.

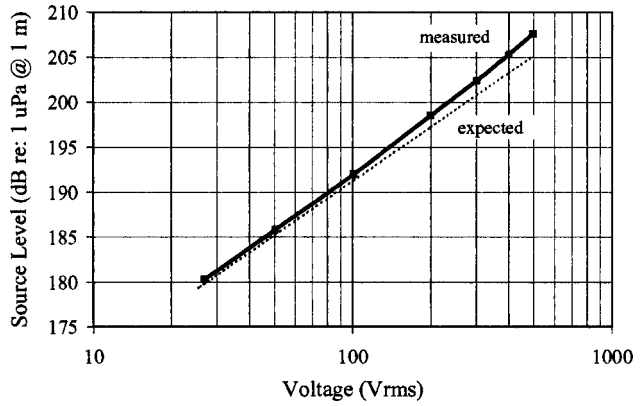


FIG. 12. Acoustic source level achieved for a given input voltage to the projector at its resonance frequency. Measured data points are shown as squares.

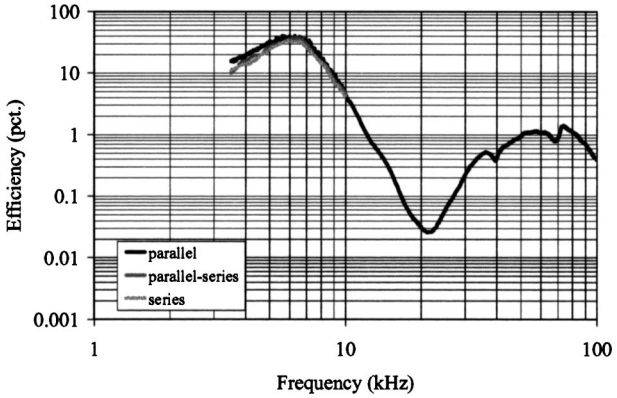


FIG. 13. Calculated electroacoustic efficiency of the projector for the three different wiring configurations.

The electroacoustic efficiency of the projector as a function of frequency is shown in Fig. 13. The efficiency (η) was calculated by solving Eq. (4) for η. The values for TVR and G are from Figs. 7 and 8, respectively. The directivity index (DI) is related to the directivity function (DF) of a transducer as

$$\mathrm{DI} = 10\log(\mathrm{DF}), \tag{7}$$

The directivity function of an unbaffled planar array of equally spaced, uniformly driven sources can be represented by[23]

$$\mathrm{DF} \cong \left[\frac{\lambda^2}{L^2}\int_0^L \left(1-\frac{\rho}{L}\right)\sin\left(\frac{2\pi\rho}{\lambda}\right)d\rho\right]^{-1} \tag{8}$$

which, upon integration, gives

$$\mathrm{DF} = \left[\frac{\lambda^2}{2\pi L_x L_y}\left\{1-\frac{\lambda}{2\pi D}\sin\left(\frac{2\pi D}{\lambda}\right)\right\}\right]^{-1}, \tag{9}$$

where λ is the acoustic wavelength in water, L_x and L_y are the respective acoustic aperture dimensions of the projector in the x and y directions, and D is the diameter of a piston having an area equivalent to $L_x L_y$.

C. Directivity patterns

Measured and predicted directivity patterns at 10 kHz are shown in Fig. 14. Patterns in both the azimuthal [Fig. 14(a)] and elevation [Fig. 14(b)] directions are provided. In Fig. 14(c), the directivity patterns for the three different wiring configurations are superimposed. The plots show that the wiring configuration has no significant effect on the directivity pattern, especially along the acoustic axis.

The predicted beam patterns were calculated from Eq. (10),[24]

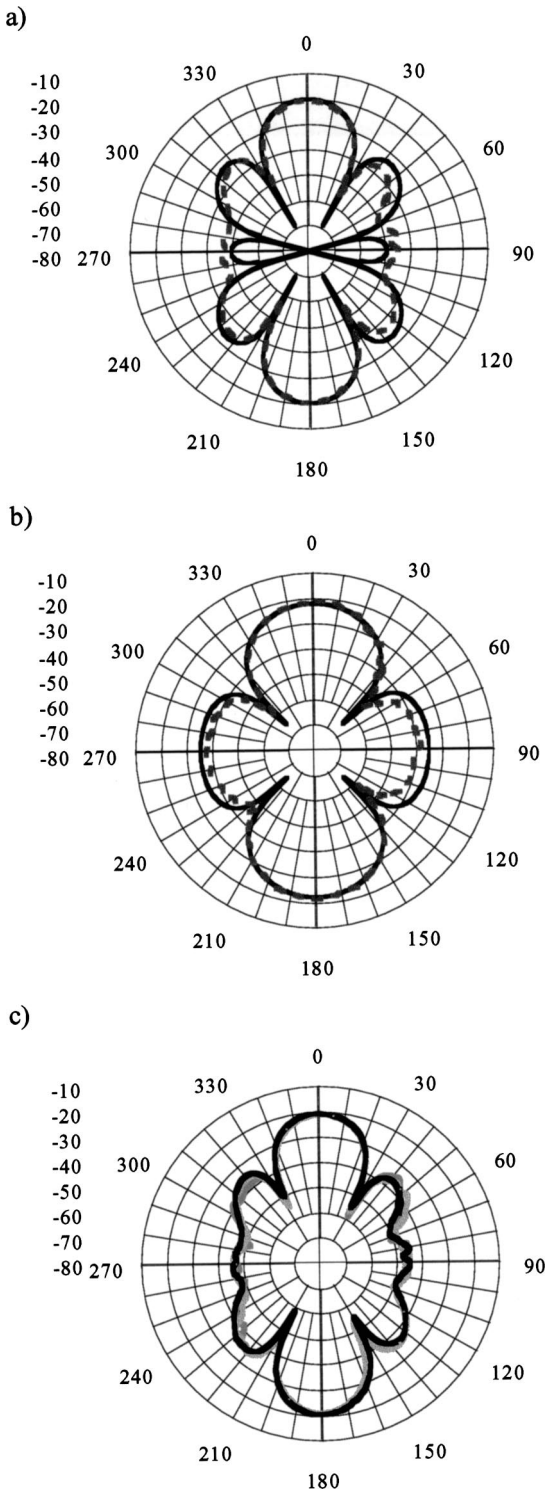


FIG. 14. Calculated (solid line) and measured (dashed line) directivity patterns of the projector measured at 10 kHz with the projector in (a) the azimuthal direction and (b) the elevation direction. In case (c), the measured azimuthal patterns are shown superimposed for the three different wiring configurations.

$$F(\psi)=\left\{\frac{\sin\left(\frac{K\Phi}{2}\right)}{\frac{K\Phi}{2}}\right\}\cdot\left\{\frac{\sin\left(\frac{NKr}{2}\right)}{\sin\left(\frac{Kr}{2}\right)}\right\}, \tag{10}$$

where $K=(2\pi/\lambda)\sin(\psi)$, Φ is the diameter of an individual cymbal element, r is the cymbal center-to-center separation distance, and N is the number of cymbals along the length of the projector aperture in either the azimuthal or elevation direction. The measured data compare quite well to the predicted patterns, especially along the main acoustic axis, which is indicative of a well-behaved projector.

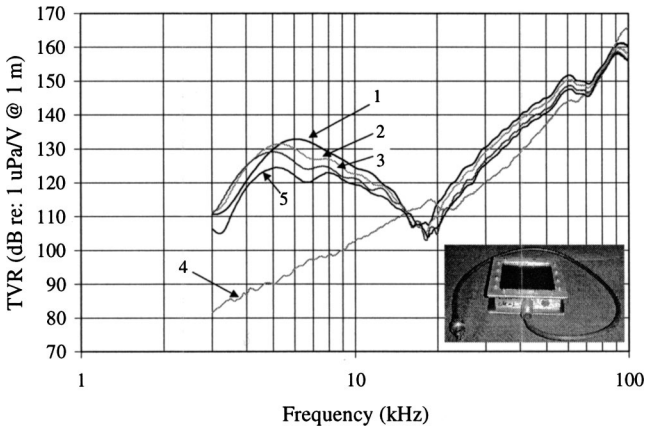


FIG. 15. Transmitting voltage response of a prototype projector (seen in the inset) as a function of hydrostatic pressure, where 1=0.02 MPa, 2=1.38 MPa, 3=2.07 MPa, and 4=5.52 MPa. Data set 5 corresponds to the TVR remeasured at 0.02 MPa, after having been subject to 5.52 MPa. To convert pressure to water depth, x MPa$\sim 100x$ meters.

For an unshaded rectangular aperture, the maximum amplitude of the side lobes should theoretically be 13 dB less than the maximum amplitude of the main lobe,[25] which in fact can be seen in Fig. 14.

D. Performance under hydrostatic pressure

In order to examine the effects of hydrostatic pressure on the projector performance, a 49 cymbal element module prototype was built and evaluated. A picture of the unit is shown as an inset in Fig. 15. The data shown in this figure was measured at the Newport Code 8211 Acoustic Test Facility[17] open water pool with the prototype mounted inside a pressure tank and the hydrophone mounted outside the tank in the pool. The measurements were conducted over the frequency range 3 kHz to 100 kHz at hydrostatic pressures of 0.02 MPa, 1.38 MPa, 2.07 MPa, and 5.52 MPa, with a repeat measurement at 0.02 MPa after being subjected to 5.52 MPa. This type of measurement reveals the relative performance changes as opposed to absolute calibration. Furthermore, the frequency response of this prototype is different from the 294 cymbal element projector used in the previously described investigation (Fig. 7) because of the difference in aperature area.

As the pressure on the cap surface is increased, the cap deforms[26] and the air cavity between the PZT and cymbal cap is compressed. This shallower air cavity accounts for the reduction in the resonance frequency[18] as the pressure is increased from 0.02 MPa to 2.07 MPa. As the hydrostatic pressure is increased beyond 2.07 MPa, the external load exceeds the generative force capability of the cymbal caps, resulting in the cap no longer flexing. Consequently, a resonance is no longer seen in the transmitting response and its magnitude is comparable to that of the PZT disk alone. Above ~15 kHz,

the TVR is governed by the radial mode of the PZT disks, and the hydrostatic pressure does not affect this mode. Note that the upper resonance frequency, at approximately 100 kHz, was not significantly changed by pressure. This upper resonance is associated with the radial mode of the piezoelectric ceramic disk. Bulk PZT ceramics do not suffer significant changes in performance from hydrostatic pressure up through at least 50 MPa.[27]

The reduction in transmitting response cannot be viewed as catastrophic, however, as the resonance characteristics of the projector return when the high hydrostatic pressure is alleviated (curve 5 of Fig. 15). The response is lower than what was measured initially because of permanent deformation in the cymbal caps but the projector did remain operational.

V. CONCLUDING REMARKS

A thin profile, lightweight underwater electroacoustic projector has been designed and experimentally evaluated. The low frequency acoustic output from the projector primarily comes from the low frequency resonance associated with the flexure of the cymbal caps. It has been shown in earlier work[18] that the fundamental resonance frequency of the cymbal driver can be easily manipulated simply by changing its dimensions. As a consequence, the fundamental resonance frequency of the projector, and subsequently the desired usable frequency band, can be tailored to meet the design needs by utilizing appropriately sized cymbal drivers. The higher output at the upper end of the frequency band is associated with the increasing conductance due to the radial mode of the piezoelectric ceramic disk as well as the increasing directivity index (DI) of the projector.

The versatility of this design also allows the cymbal driver elements to be addressed electrically either individually or in predetermined apertures (i.e., ''nested''). This means that (a) multiple apertures may be realized within the same radiating area, (b) each individual cymbal element can be addressed to manipulate beam forming, and (c) through predetermined selection, a sparse array can be realized. Such a wiring scheme allows the cymbal elements to be combined electrically either in parallel, in series, or as a combination of both without the need to take apart the projector for rewiring. Ultimately, this device design will allow for multipurpose SONAR within the same packaging.

Although designed for and best suited for littoral environments, the oil-filled cymbal-based projector has shown that it can operate without significant degradation in performance at water depths not exceeding 200 meters.

ACKNOWLEDGMENTS

This work was supported by ONR. The authors would like to acknowledge the contributions of Joe Klunder of SFA, Inc. as well as Dehua Huang, Kim Benjamin, Walter Boober, Rene LaFleur, and Hugo Mendoza of NAVSEA-Newport, Kirk Robinson and Mel Jackaway of the NAVSEA-Crane Glendora Lake test facility, and Pat Arvin, Scott Small, and Phil Meadows of NAVSEA-Crane.

[1] R. Scott and M. Hewish, ''Remote hunting: key to littoral waters,'' Jane's International Defense Review, 1999, pp. 48–54.
[2] S. J. Tomaszeski, ''Naval transformations and the oceanographic community,'' Sea Technol. **2004**, 10–13 (2004).
[3] T. R. Howarth and R. Y. Ting, ''Electroacoustic evaluations of 1–3 piezocomposite SonoPanel™ materials,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 886–894 (2000).
[4] J.-N. Decarpigny, B. Hamonic, and O. B. Wilson, Jr., ''The design of low-frequency underwater acoustic projectors: present status and future trends,'' IEEE J. Ocean. Eng. **16**, 107–122 (1991).
[5] D. Boucher, ''New solutions for low frequency sonar projectors,'' in *Transducers for Sonics and Ultrasonics*, edited by M. D. McCollum, B. F. Hamonic, and O. B. Wilson (Technomic Publishing, Lancaster, PA, 1993), Chap. 2, pp. 17–37.
[6] R. A. Nelson, Jr. and L. H. Royster, ''Development of a mathematical model for the class V flextensional underwater acoustic transducer,'' J. Acoust. Soc. Am. **49**, 1609–1620 (1971).
[7] R. S. Woollett, ''Basic problems caused by depth and sound constraints in low-frequency underwater transducers,'' J. Acoust. Soc. Am. **68**, 1031–1037 (1980).
[8] R. E. Newnham and A. Dogan, ''Metal-electroactive ceramic composite transducer,'' U.S. Patent No. 5,729,077, 1998.
[9] A. Dogan, K. Uchino, and R. E. Newnham, ''Composite piezoelectric transducer with truncated conical endcaps 'cymbal','' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 597–605 (1997).
[10] U.S. Military Standards on Piezoelectric Ceramic Materials and Measurements Guidelines for Sonar Transducers, MIL-STD-1376B(SH) (1995).
[11] Channel Industries, Inc., *Piezoelectric Ceramics*, product catalog provided by Channel Industries, Inc., Santa Barbara, CA.
[12] W. L. Carney, J. F. Tressler, and T. R. Howarth, ''Thin, low frequency, oil-filled projectors for UUV applications,'' *Oceans 2001 MTS/IEEE Conference Proceedings Vol. 3* (IEEE Service Center, Piscataway, NJ, 2002), pp. 1577–1580.
[13] T. R. Howarth, W. L. Carney, J. L. Merryfield, P. L. Arvin, P. W. Meadows, S. L. Small, and J. F. Tressler, ''Mount for underwater acoustic projector,'' U.S. Patent Application (submitted), U.S. Navy Case No. 84,310.
[14] D. Stansfield, *Underwater Electroacoustic Transducers* (Bath University Press, Bath, UK, 1991), Chap. 4-3, pp. 78–83.
[15] M. B. Moffett and W. J. Marshall, Jr., ''The importance of coupling factor for underwater acoustic projectors,'' NUWC-NPT Technical Document 10,691, 1994.
[16] http://www.crane.navy.mil/acoustics/glendoralake/glendora.asp
[17] http://www.npt.nuwc.navy.mil/ATF8211/index.html
[18] J. F. Tressler, W. Cao, K. Uchino, and R. E. Newnham, ''Finite element analysis of the cymbal-type flextensional transducer,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 1363–1369 (1998).
[19] K. E. Jenne, ''Acoustic cymbal transducers—design, hydrostatic pressure compensation, and acoustic performance,'' M.S. thesis, Naval Postgraduate School, Monterey, CA, 2004, pp. 35–36.
[20] ITC, *Application Equations for Underwater Sound Transducers* (pamphlet) (International Transducer Corporation, Santa Barbara, CA, 1993).
[21] Edo Corporation, *Sonar & Transducer Computer*, 3rd ed. (Edo Acoustics Division, Salt Lake City, UT, 1993).
[22] Piezo Kinetics, Inc., *PKI Material Datasheet*, provided by Piezo Kinetics, Inc., Bellefonte, PA. It can also be accessed via the internet at http://www.piezokinetics.com/_downloads/datasheet.pdf
[23] W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1991), Chap. 11, pp. 332–336.
[24] D. Stansfield, *Underwater Electroacoustic Transducers* (Bath University Press, Bath, UK, 1991), Chap. 6-4, pp. 164–170.
[25] D. Stansfield, *Underwater Electroacoustic Transducers* (Bath University Press, Bath, UK, 1991), Chap. 9-5, p. 262.
[26] J. F. Tressler, ''Capped ceramic underwater sound projector, the 'cymbal','' Ph.D. dissertation, The Pennsylvania State University, University Park, PA, 1997, pp. 214–219.
[27] T. R. Gururaja, A. Safari, R. E. Newnham, and L. E. Cross, ''Piezoelectric ceramic-polymer composites for transducer applications,'' in *Electronic Ceramics: Properties, Devices and Applications*, edited by L. M. Levinson (Marcel Dekker, New York, 1988), Chap. 2.3, pp. 92–128.

Analysis of screw pitch effects on the performance of bolt-clamped Langevin-type transducers

Kazunari Adachi, Toru Takahashi, and Hiroshi Hasegawa
Cooperative Research Center, Yamagata University, 4-3-16 Jonan, Yonezawa, Yamagata, 992-8510, Japan

(Received 5 February 2004; revised 15 May 2004; accepted 9 June 2004)

Bolt-clamped Langevin-type transducers (BLTs) are common vibration sources in high-power ultrasonic applications such as ultrasonic plastic joining. In this paper, the authors propose a low-aspect-ratio BLT shape based on numerical solutions of a complex elastic contact problem concerning the bearing stress (prestress) imposed on the interfaces between the parts by clamping with the screw bolt. The prestress distribution at the interface has significant influence on the mechanical quality factor (Q) of the BLT. It is found that the screw pitch of the clamping bolt heavily affects the prestress distribution in the simulation using the finite element method. The newly developed BLTs with a high resonance frequency of approximately 80 kHz has a relatively wide radiating face and sufficient volume ratio of the piezoelectric elements that convert electrical energy into mechanical energy. The average of their measured Q values exceeds 1000 despite their high resonance frequency when they are driven at a voltage higher than 17 V rms. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1777852]

PACS numbers: 43.38.Fx, 43.35.Zc [AJZ] Pages: 1544–1548

I. INTRODUCTION

Bolt-clamped Langevin-type transducers (BLTs) are widely used in the field of high-power ultrasonics as high-amplitude vibration sources. These types of transducers are particularly indispensable to such important industrial applications as ultrasonic plastic joining, cleaning, and machining. Figure 1 shows a typical configuration of this type of transducer. In the structure, two piezoelectric ceramic disks (commonly those of polarized lead zirconate titanate or PZT) with phosphorous or beryllium copper alloy electrodes are sandwiched by two metal blocks and clamped with a screw bolt. The clamping exerts static bearing stress or prestress on the disks so as to compensate for low tensile strength of the ceramic material. The tensile strength of a piezoelectric ceramic material is typically around 50 MPa though they bear compressive stress higher than 500 MPa.

In order to design a BLT with a high mechanical quality factor (Q)—in other words, a low-mechanical-loss BLT—one of the most crucial points is the distribution of the static bearing stress imposed on the interfaces between the parts. Sufficient margin of the static stress above the vibratory one must be kept for maintaining stable contact at the interfaces between the parts in high-amplitude operation of the transducer at its resonance frequency.[1,2] Otherwise, excessive heat generation due to repetitive impacts at the interfaces decreases the efficiency of its electromechanical transduction, lowers the Q, and the thermal stress may cause fractures in the hard but brittle ceramic disks. Furthermore, since the internal dissipation of the materials themselves are usually very low for BLTs of practical use, the only dominant cause of degrading the Q is this heat generation at the interfaces. Nevertheless, estimation of the bearing stress distributions involves intricate elastic contact problems with piezoelectricity to be taken into account.

On the other hand, the use of high-resonance-frequency BLTs has recently been becoming more desirable in some fields of this technology because of increasing demand for fine finish of works, high production efficiency, and low-noise working environment. A higher resonance frequency means a shorter length of the transducer. However, for keeping its high-power ultrasound transduction capability, the volume of the ceramic disks that convert electrical energy into mechanical energy cannot be reduced. This technical requirement inevitably leads to a low-aspect-ratio or ''stumpy'' shape of the BLT as illustrated in Fig. 1 and makes rather difficult the metal block design for maintaining the stable and uniform contact all over the interfaces.

Many researchers have already investigated the influence of the prestress on the characteristics of the transducers.[1–8] Nevertheless, very few papers[1,7,8] have clearly stated on its distribution at the interfaces. Adachi and co-workers have been striving to find out the keys to high-performance BLT design and establish its procedure for many years. Using the finite element analysis system specially developed for this purpose, they have already found that the metal block shape significantly affects the bearing stress distribution.[7] In this paper, by means of the same numerical analysis, the authors elucidate the influence of the screw pitch of the clamping bolt on the bearing stress distribution which eventually determines the Q of the transducer. On the basis of the analysis, they produced BLTs whose resonance frequency is some 80 kHz.

II. NUMERICAL INVESTIGATION

Figure 2 shows the BLT structure for investigating the screw pitch effect on the static bearing stress distribution numerically, where two piezoelectric ceramic disks and two metal blocks are assumed to be assembled by clamping with a hollow screw bolt. The dimensions were chosen to realize a resonance frequency higher than 80 kHz, while maintaining the volume of piezoelectric disks and the area of the radiating face (the end surface) as large as possible. In addi-

 © 2004 Acoustical Society of America

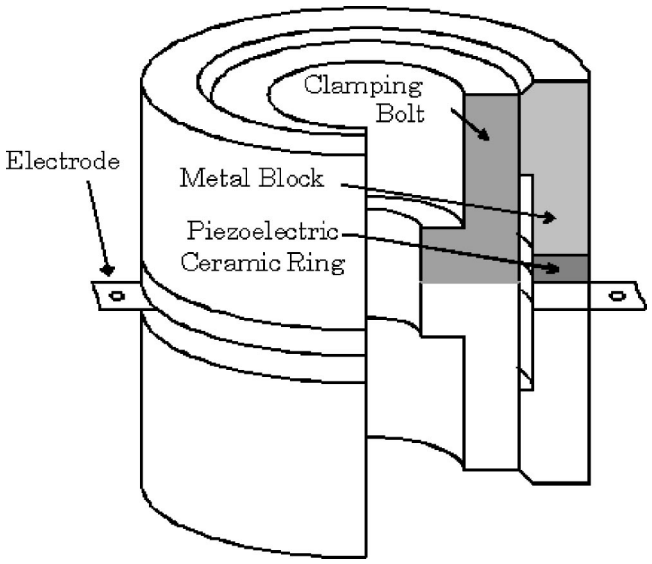


FIG. 1. A typical bolt-clamped Langevin-type transducer (BLT).

tion, the hollow cylindrical shape helps to suppress Poisson's effect and drastically reduce the lateral component of the vibratory stress at the interface between the metal block and the piezoelectric ceramic disk in resonance vibration, though the lateral movement itself is not negligible.

For simplicity, the thin electrodes were not represented and the symmetry with respect to both its axis and the interface between the two ceramic disks was also assumed in all cases of the simulation. Thus, only the quarter part of the transducer was considered here. Energy dissipation was totally ignored in the calculation. The shape of the screw thread was realistically modeled with small finite elements as shown in Fig. 3(b) to grasp the pitch effect as precisely as possible, while its simple representation by single nodes shown in Fig. 3(a) had been used in previous research[7,8] conducted by the authors with the use of the same finite-element analysis system.

The authors adopted two idealistic frictional contact models at the thread in the calculation of the static bearing stress between the metal block and the ceramic disk. In the

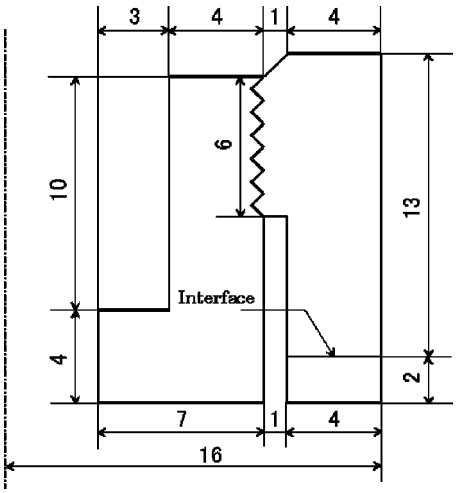


FIG. 2. Geometry of the BLT for which the screw pitch effect on the static bearing stress imposed on the interface between the parts was investigated.

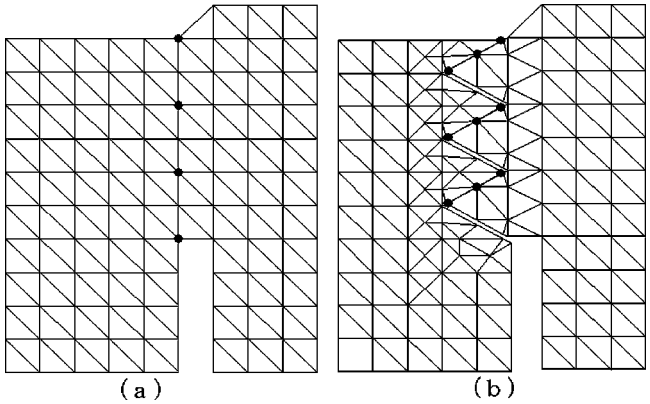


FIG. 3. Screw thread representations in the finite-element modeling: (a) simple representation by single nodes; (b) realistic one.

first model, it is assumed that the pair of nodes to be coupled at the thread get perfectly fixed to each other as the result of the clamping and share the same position as shown in Fig. 4(a). The second one allows their relative movement parallel to the thread surface without friction in the clamping process as illustrated in Fig. 4(b). Actual frictional condition at the thread is undoubtedly an intermediate between the two extreme cases.

At the interface between the metal block and the piezoelectric ceramic disk, only the no-lateral-movement frictional contact model is adopted since the static bearing stress component perpendicular to the interfaces is apparently overwhelming compared with the lateral one and the friction between the parts avoids their relative movement. For the half part of the BLT, the relative axial displacement between the male and female screws for clamping was assumed to be 0.18 mm, which is equivalent to approximately 65 degree rotation of each metal block for the case of 1 mm screw pitch when tightening it.

The vibratory stress in resonance at the interface between the metal block and the piezoelectric ceramic disk was also calculated for the case of high-amplitude operation where the particle displacement in the axial direction is 10 microns (0-p) at the outer rim of the BLT's end surface. In the modal vibration analysis, each pair of nodes coupled in the clamping-process analysis were assumed to be combined into a single node owing to the friction, since the lateral

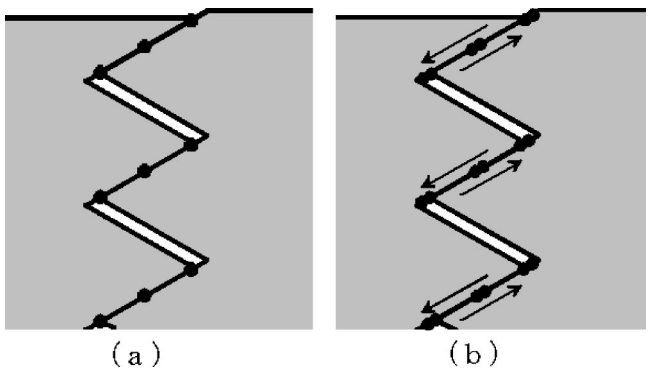


FIG. 4. Two idealistic frictional contact models adopted in the static bearing stress analysis: (a) no-lateral-movement model; (b) free-lateral-movement model.

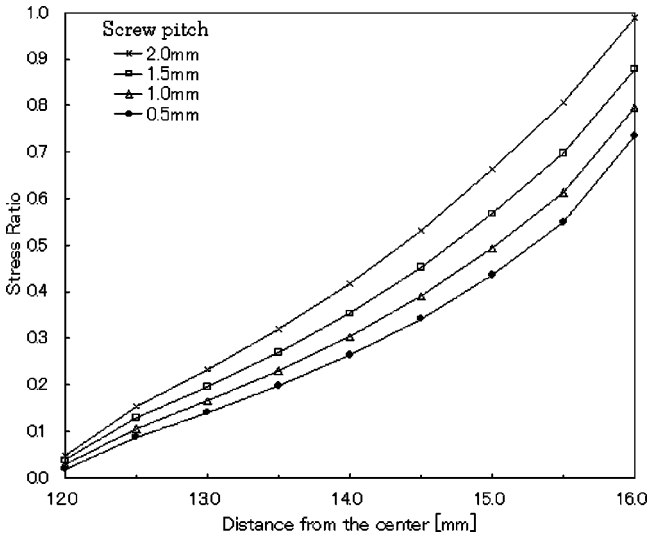


FIG. 5. Calculated stress ratio distribution at the interface between the metal block and the ceramic disk using the no-lateral-movement model for calculation of the static bearing stress distribution. The parameter is the screw pitch of the bolt.

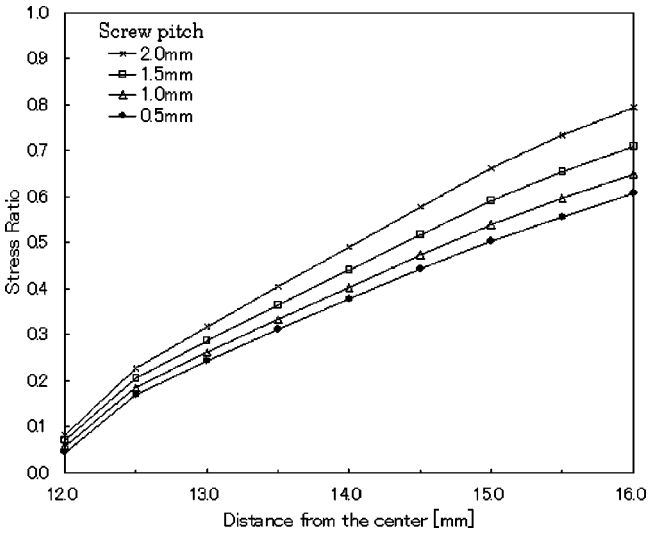


FIG. 6. Same as in Fig. 5 but using the free-lateral-movement model for calculation of the static bearing stress distribution.

component of the vibratory stress is quite small compared with the vertical component of the static bearing stress if the clamping is firm enough.

A value called the ''stress ratio'' was introduced to interpret the results of numerical analysis on the static and vibratory stress at the interface between the metal block and the piezoelectric ceramic disk. It is defined as the ratio of the vertical component of the vibratory stress calculated for an amplitude of 10 microns in resonance to that of the static bearing stress exerted by the clamping at the interface. When the ratio is less than unity all over the interface, the contact between the metal block and the piezoelectric ceramic disk is always stable even in the operation of a very high amplitude of 10 microns. The stable contact leads to low energy dissipation at the interface and can contribute to realize a high Q of the transducer as mentioned in the preceding section. In actual cases, however, the ratio should be much less than unity in order to secure the stability of contact and eventually the high Q of the BLT since there is inevitably some uncertainty of contact conditions at the interface such as the surface roughness that heavily affects the friction between the parts.

Figures 5 and 6 show the stress ratio distributions at the interface between the metal (duralumin) block and the ceramic disk (Fuji Ceramics C-21), which were calculated with the use of the two frictional contact models at the screw thread in the clamping process as illustrated in Fig. 4, respectively, for four different screw pitches, namely 0.5, 1.0, 1.5, and 2.0 mm. In both frictional contact cases, we find the same tendency that the finer screw pitch gives the lower stress ratio distribution at the interface and can be expected to realize the higher Q of the BLT, though the sign of curvature of the lines in the graphs is reversed. Since the real situations must be settled in between the two cases, it is not preposterous to infer that the screw pitch actually has this effect.

III. EXPERIMENTAL VERIFICATION

The numerical results described in the preceding section should be experimentally verified. The authors, therefore, made four BLTs for each screw-pitch type—the total number of the test pieces manufactured for this purpose is thus 16—and measured their resonance frequencies, Q and electromechanical coupling factors using HP-4192A impedance analyzer controlled by a personal computer through its GP–IB interface. Before assembling, the metal blocks and piezoelectric ceramic disks of each BLT had been polished at their faces contacting each other so that the average surface roughness (Ra) there was maintained at 100 nm or lower so as to make the contact between them as uniform as possible. The voltage applied to the BLTs was 1 V rms at this stage.

TABLE I. All the raw data of the measured characteristics of BLTs and the comparison with their calculated values. The screw pitches of the clamping bolts are 0.5 mm, 1.0 mm, 1.5 mm, and 2.0 mm. The symbols f_0, Q, and k_{vn} are the resonance frequency, mechanical quality factor, and electromechanical coupling factor, respectively, of the BLT. The driving voltage is 1 V rms.

	Calculated	No. 1	No. 2	No. 3	No. 4
		Screw pitch:0.5 mm			
f_0 [kHz]	81.45	79.86	79.72	80.02	79.90
Q		1141	1092	879.3	429.6
k_{vn}	0.099	0.116	0.119	0.121	0.126
		Screw pitch:1.0 mm			
f_0 [kHz]	81.41	78.73	79.08	79.22	79.37
Q		354.6	454.5	660.2	809.9
k_{vn}	0.101	0.122	0.122	0.131	0.132
		Screw pitch:1.5 mm			
f_0 [kHz]	81.27	77.67	78.71	79.11	78.81
Q	–	202.3	378.4	530.5	406.2
k_{vn}	0.103	0.125	0.122	0.126	0.122
		Screw pitch:2.0 mm			
f_0 [kHz]	81.04	79.51	79.36	79.36	79.11
Q		262.4	547.3	566.9	530.9
k_{vn}	0.104	0.111	0.118	0.117	0.123

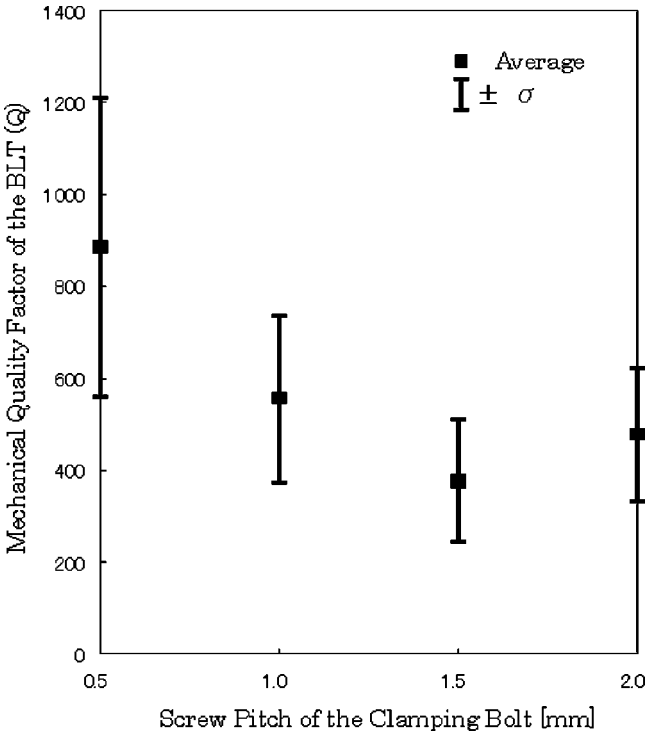


FIG. 7. The mechanical quality factor of the BLT versus the screw pitch. Each error bar indicates the range of a standard deviation from the average. The driving voltage is 1 V rms.

All the BLTs are placed on a soft polyurethane-foam bed during the measurement to minimize the effects of mechanical support.

The measured raw data are tabulated in Table I. From these results, we find the characteristics calculated by the authors' finite-element analysis system fairly match the measured ones. Figure 7 shows the Q values plotted versus the screw pitch where each error bar indicates the range of one standard deviation from the average value of each four data. The figure well demonstrates the effect of fine screw pitch on the Q value as expected in the preceding section except for the case of 2.0 mm screw pitch. In spite of their high resonance frequencies which reach approximately 80 kHz, the Q values of some 1000 are realized by the two BLTs with the clamping bolt of the screw pitch of 0.5 mm. The average Q of the BLTs with this type of screw bolt exceeds 800.

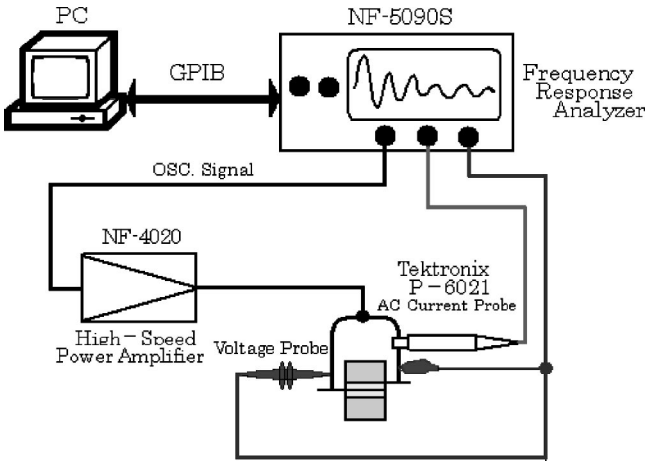


FIG. 8. Experimental setup for measuring the resonance frequency and Q of the BLTs with a high driving voltage.

TABLE II. The average resonance frequency and Q value measured for each screw pitch of the clamping bolt of the BLT with the driving voltage of 17.9 V rms.

Screw pitch	0.5	1.0	1.5	2.0
f_0 (Average) [kHz]	80.02	79.29	79.03	79.58
Q (Average)	1100	596.0	400.0	457.3

However, the characteristics of the BLTs for high-power use, especially the Q should be measured with a much higher driving voltage so as to prove the screw pitch effect. The driving voltage of just 1 V rms is so small that it can excite the vibration of an amplitude only as much as 0.025 microns at the end surface of the BLT. This amplitude is 1/50 to 1/20 of a practically required level at the frequency of 80 kHz.[8] The higher driving voltage may decrease the Q of the BLTs owing to instable contact at the interface.

Hence, the authors measured the resonance frequencies and the Q values of the BLTs with the driving voltage of 17.9 V rms using NF Electronic Instruments NF-5090S Frequency Response Analyzer, NF-4020 high-speed power amplifier, and Tektronix P-6021 AC current probe as illustrated in Fig. 8. Table II shows the averages of those values for each screw pitch of the clamping bolt. The average Q of the transducers reaches 1100 when the screw pitch is 0.5 mm. In Fig. 9, we can confirm the same tendency of the Q as a function of the screw pitch as shown in Fig. 7 with this relatively high driving voltage.

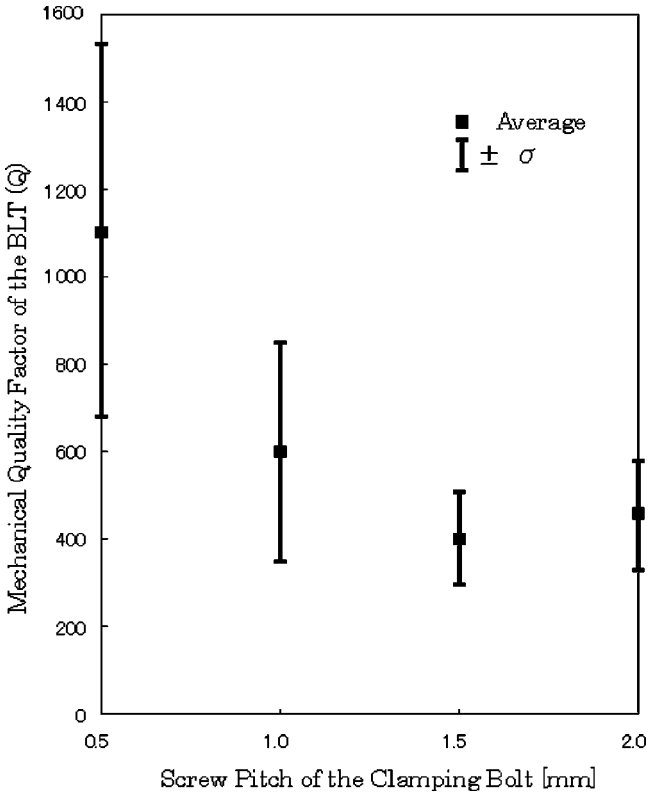


FIG. 9. Same as in Fig. 7 but the driving voltage is 17.9 V rms.

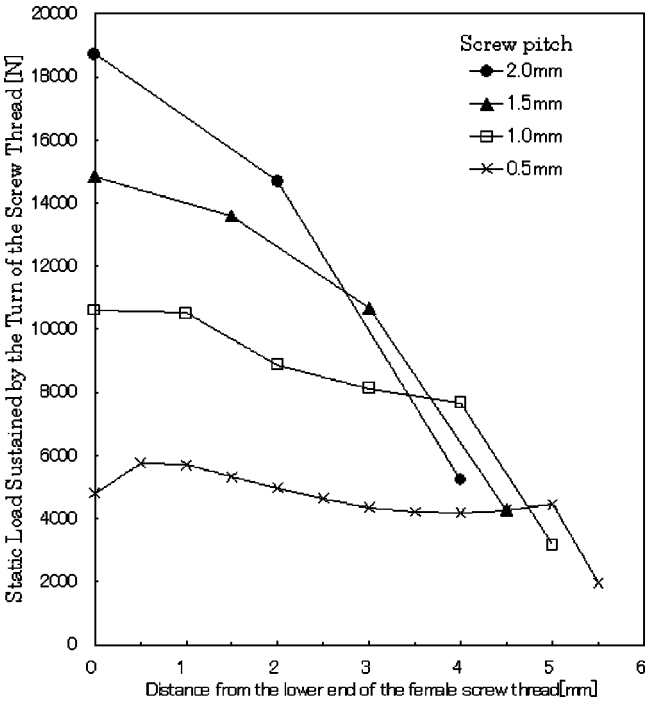


FIG. 10. The calculated distribution of the static load sustained by the clamping bolt along its length with screw pitch as a parameter.

IV. DISCUSSION

In the numerical simulation, the finer is the pitch of the screw bolt, the more uniform is the distribution of the static load along its length as shown in Fig. 10. When the screw pitch is the largest 2 mm, most part of the static load is sustained by only the two turns of the screw thread closest to the interface. In such a case, Saint Venant's principle in elasticity suggests quite a nonuniform prestress distribution at the interface. The screw pitch effect on the stress ratio distribution can, therefore, be attributed to this phenomenon, but further investigation may be still necessary for elucidation of the cause. More accurate modeling of frictional contact at the screw thread might be required for the purpose.

There is another problem to be considered from the practical point of view. The standard deviation of the Q tends to be larger as the screw pitch becomes finer both in the low and high amplitude operations of the BLTs. The most possible reason for this dispersion of data is relative accuracy of the shape of the screw thread. It gets more difficult to fabricate the screw thread quite precisely when the pitch becomes finer. The slight deviation from the standard geometry of the screw thread may cause significant fluctuation of the contact at the interface and eventually influence the Q of the BLT. Hence, the errors in the Q shown in Figs. 7 and 9 can be attributed to the lack of precision of the screw thread fabrication.

In addition, plastic deformation at the screw thread in the clamping process may not be negligible. If it occurs, the decrease in the prestress can be significant owing to the relaxation of the static load sustained by the thread. However, estimation of the plastic deformation effect for this specific case would be prohibitively difficult without bold approximations based on the insight into the phenomenon because of its nonlinear and moving-boundary nature.

V. CONCLUSION

With the use of the method of finite-element simulation proposed by Adachi, the authors investigated the influence of the screw pitch of the clamping bolt on the stress ratio distribution at the interface between the metal block and the piezoelectric ceramic disk of the BLT whose resonance frequency exceeds 80 kHz. The results of experimental verification well indicate appropriateness of the modeling that the authors adopted in the numerical analysis. From the numerical and experimental results, it may be concluded that the finest screw pitch, namely 0.5 mm results in the lowest value of the stress ratio at the interface and eventually realize the highest average Q of some 1100 in operation at a driving voltage of 17 V rms. Consequently, the method can be quite helpful for designing BLTs used in high-power ultrasonics particularly those with high resonance frequencies whose aspect ratios tend to be low.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Kazuyuki Yoneya, vice director of research and development department in Seidensha Electronics Co. Ltd., Toshiyuki Nishino, executive officer of JAE Yamagata, Ltd., Nobunaga Shibuya in Honda Electronics Co. Ltd. for their cooperation in manufacturing many test pieces.

[1] S. Ueha and Eiji Mori, "Resonance condition and clamping-bolt diameter of Bolt-clamped Langevin type transducer," J. Acoust. Soc. Jpn. **1978**, 635–640 (1978) [in Japanese].
[2] S. Ueha and Eiji Mori, "Effects of surface condition of components and of temperature upon performance of a Bolt-clamped Langevin type longitudinal transducer," J. Acoust. Soc. Jpn. **1979**, 469–476 (1979) [in Japanese].
[3] H. H. A. Krueger and D. Berlincourt, "Effects of high static stress on the piezoelectric properties of transducer materials," J. Acoust. Soc. Am. **33**, 1339–1344 (1961).
[4] R. Y. Nishi, "Effects of one-dimensional pressure on the properties of several transducer ceramics," J. Acoust. Soc. Am. **40**, 486–495 (1966).
[5] H. H. A. Krueger, "Stress sensitivity of piezoelectric ceramics: Part 1. Sensitivity to compressive stress parellel to the polar axis," J. Acoust. Soc. Am. **42**, 636–645 (1967).
[6] T. Inoue and S. Takahashi, "Measurement for piezoelectric ceramic properties applied compressive stress parallel to the polar axis," Trans. Inst. Chem. Eng., Part A **64**, 1180–1187 (1986).
[7] K. Adachi, M. Tsuji, and H. Kato, "Elastic contact problem of the piezoelectric material in the structure of bolt-clamped Langevin-type transducer," J. Acoust. Soc. Am. **105**, 1651–1656 (1999).
[8] K. Adachi and H. Hasegawa, "Development of a bolt-clamped Langevin-type transducer for operation at a high frequency of 80 kHz," 2001 IEEE Ultrasonics Symposium Proceedings, 2001, Vol. 1, pp. 661–664.

A wide-angle on-surface radiation condition applied to scattering by spheroids

David C. Calvo[a)]
Naval Research Laboratory, Washington, DC 20375

(Received 20 February 2004; revised 8 June 2004; accepted 9 June 2004)

A rational approximation of the pseudo-differential Dirichlet-to-Neumann operator for a circular cylinder is made using zero placement. For an acoustically hard circular cylinder, this zero corresponds to the dominant creeping-wave singularity in the residue series for the time-harmonic free-space surface Green's function. Based on this operator, a wide-angle on-surface radiation condition (OSRC) is developed first for two-dimensional slender convex bodies. For the object scattering problem, this OSRC is accurate for wide-scattering angles relative to the surface normal. The notion of creeping waves is then used to develop a wide-angle, three-dimensional OSRC applicable to smooth convex bodies. Application to both hard and soft, prolate and oblate spheroids is presented, and excellent agreement is found in comparison with a more costly technique. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1777874]

PACS numbers: 43.40.Fz, 43.20.Fn [ANN] Pages: 1549–1558

I. INTRODUCTION

The on-surface radiation condition (OSRC) approach for solving scattering problems was introduced by Kriegsmann *et al.*[1] by shrinking the extent of a computational domain, truncated by a radiation boundary condition, all the way down to the surface of the scattering obstacle. The OSRC operator, which may be viewed as an approximation of the exact Dirichlet-to-Neumann (D-t-N) impedance operator, was then used to determine the scattered field outside the boundary using the Helmholtz integral equation.[2,3] Since its inception, the accuracy of the OSRC has been improved by including derivatives of the surface curvature in the operator and applicability has been extended to three-dimensional bodies.[4–9]

Despite these improvements, the accuracy of OSRCs generally deteriorates when the incident field grazes the obstacle over several wavelengths, the best example being end-on incidence of a plane wave with a hard, slender body. This inaccuracy stems from a nonuniformity in the expansion underlying the OSRC for wide propagation angles of the scattered field relative to the surface normal. In a recent study, Calvo *et al.* have improved accuracy in this difficult case by introducing a rational approximation of the exact pseudo-differential D-t-N operator for a circular cylinder.[10,11] Taking into account other approximations used in past OSRC studies to handle noncircular bodies, excellent results for grazing angles of incidence in both high and low frequency regimes were obtained.

Even though a solution obtained using an OSRC (especially with the above improvements) can be very accurate and inexpensive (cost scales only linearly with frequency in the two-dimensional problem), most scattering computations in engineering applications are based directly on the Helmholtz integral equation—the boundary element method being a common choice. Therefore, an OSRC can be very valuable if used in conjunction with an iterative integral-equation scheme to compute an initial guess or as a preconditioning operator. There is evidence that using OSRCs in this way can lead to faster convergence of a Krylov subspace iterative method (Antoine and Darbas, personal communication). Such an approach could provide impressive gains in computational speed for mid-to-high-frequency problems featuring complicated diffractive geometries. This still remains an expensive task if one desires an accurate solution using boundary elements.

Given this motivation to further enhance OSRCs, the objective of this study is twofold. First, in Secs. II and III, the OSRC presented in Calvo *et al.* is substantially economized by taking into account the high-frequency asymptotic behavior of the exact pseudo-differential D-t-N operator for complex values of the scattering-angle parameter. Then, with some relatively simple replacements in the two-dimensional OSRC, extension is made to three-dimensional convex bodies in Sec. IV allowing for a general angle of incidence. Although, as formulated, the OSRC is not limited to bodies of revolution, application is made to prolate and oblate spheroids in Sec. V. Excellent agreement is found with results computed using the boundary element method.

II. DIRICHLET-TO-NEUMANN OPERATOR

A. Pseudo-differential representation

As a starting point, a formal derivation of the pseudo-differential D-t-N operator (map) for an obstacle with a circular boundary is presented. Working in polar coordinates with origin in the center of the circle and assuming that no sources exist outside the boundary, the radiated (or scattered) field $p(r,\theta)$ satisfies

$$\frac{\partial^2 p}{\partial r^2}+\frac{1}{r}\frac{\partial p}{\partial r}+\frac{1}{r^2}\frac{\partial^2 p}{\partial \theta^2}+k^2 p=0, \tag{1}$$

where k is the wave number. Assuming the time-harmonic

a)Electronic mail: calvo@aegir.nrl.navy.mil

dependence $e^{-i\omega t}$ and expressing the field in terms of outgoing modes, the pressure is given by

$$p(r,\theta)=\sum_{m=0}^{\infty}\phi_m(r)\psi_m(\theta), \tag{2}$$

where

$$\frac{\partial^2\psi_m}{\partial\theta^2}+m^2\psi_m=0, \tag{3}$$

$$\frac{\partial^2\phi_m}{\partial r^2}+\frac{1}{r}\frac{\partial\phi_m}{\partial r}+\left(k^2-\frac{m^2}{r^2}\right)\phi_m=0, \tag{4}$$

and

$$\phi_m(r)=H_m^{(1)}(kr). \tag{5}$$

In the separation of variables solution, the operator $-\partial^2/\partial\theta^2$ acting on the mode $\psi_m(\theta)$ is equivalent to multiplication by the scalar m^2. Substituting this operator into the index of the Hankel function, one obtains

$$p(r,\theta)=H^{(1)}_{\sqrt{-\partial^2/\partial\theta^2}}(kr)\sum_{m=0}^{\infty}\psi_m(\theta). \tag{6}$$

Differentiating Eq. (6) with respect to r leads to the operator representation of the D-t-N map

$$\frac{\partial p}{\partial r}=kLp, \tag{7}$$

where

$$L\left(\frac{\partial^2}{\partial\theta^2},\frac{1}{kr}\right)=\frac{1}{k}\frac{(\partial/\partial r)H^{(1)}_{\sqrt{-\partial^2/\partial\theta^2}}(kr)}{H^{(1)}_{\sqrt{-\partial^2/\partial\theta^2}}(kr)}, \tag{8}$$

valid on and outside the boundary.

Two parameters appear in the D-t-N operator: the scattering angle, represented by the index, and the normalized radius of curvature appearing in the argument. In a previous study,[11] a rational approximation of Eq. (8) over a range of scattering angles was computed in which the coefficients in the approximation depended on curvature. The coefficients were obtained by a least-squares fit over a real interval of index values for a set of curvature values. Although this fitting method gave excellent results down to $kr=1$, it requires a four-term rational approximation (and therefore four tridiagonal matrix solves per evaluation of the D-t-N map). In addition, the numerically determined coefficients must be entered into the computer code used to implement the scheme. In the next section, a simpler and convenient rational approximation is constructed by using zero placement and matching in the complex plane of the index parameter.

B. Simplified rational approximation

It is useful to first make a connection between Eq. (8) and previous work on narrow-angle OSRCs. Applying Eq. (8) to a circle of radius a and expanding in the high-frequency limit, one obtains

$$L\sim i-\frac{1}{2ka}+\frac{i}{(ka)^2}\left(\frac{1}{8}-\frac{\tilde{\nu}^2}{2}\right)+O\left(\frac{1}{(ka)^3}\right), \tag{9}$$

where $\tilde{\nu}^2=-\partial^2/\partial\theta^2$. This expression matches the narrow-angle OSRC for a circular cylinder used in previous studies to $O[1/(ka)^2]$.[4–9]

Although for narrow scattering angles ($\tilde{\nu}\ll ka$) Eq. (9) is a good approximation of the D-t-N operator, it becomes nonuniform for large scattering angles ($\tilde{\nu}\approx ka$) and does not detect the complex zero of Eq. (8) nearest to $\tilde{\nu}=ka$:

$$\tilde{\nu}_*\sim ka+(0.4043+0.7003i)(ka)^{1/3}+(-0.0727+0.1260i)(ka)^{-1/3}+\cdots. \tag{10}$$

This zero corresponds to the dominant pole in the residue series for scattering by a hard circular cylinder[12] and it is therefore not surprising that past OSRCs have had difficulty treating scattering and diffraction by hard slender objects in which the diffracted (creeping wave) component of the solution makes a substantial contribution. The approach then is to construct an approximation that is uniformly valid near $\tilde{\nu}=ka$.

A simple rational approximation of Eq. (8), $L_A\cong L$ is proposed as

$$L_A=L_0(ka)\frac{(1+\alpha V^2+\beta V^4)}{1+\gamma V^2}, \tag{11}$$

where

$$V=\frac{\tilde{\nu}}{\tilde{\nu}_*} \tag{12}$$

and

$$L_0(ka)=i-\frac{1}{2ka}+\frac{i}{8(ka)^2}+\cdots. \tag{13}$$

In addition to Eq. (9) for $\tilde{\nu}\ll ka$, the asymptotic behavior of Eq. (8) near $\tilde{\nu}_*$ must be taken into account:

$$L\sim C(ka)(\tilde{\nu}-\tilde{\nu}_*)\quad(\tilde{\nu}\to\tilde{\nu}_*), \tag{14}$$

where[12]

$$C(ka)=\frac{[1-(\tilde{\nu}_*/ka)^2]}{[1+(0.1348+0.2334i)(ka)^{-2/3}]}+O[(ka)^{-4/3}]. \tag{15}$$

The constraints used to determine α, β, and γ are zero placement [Eq. (10)], matching slopes at the zero after expanding Eq. (11) about $V=1$, and matching the narrow-angle behavior of the D-t-N operator [Eq. (9)] by Taylor expanding Eq. (11) about $V=0$:

$$1+\alpha+\beta=0, \tag{16a}$$

$$\frac{2L_0(\alpha+2\beta)}{1+\gamma}=C\tilde{\nu}_*, \tag{16b}$$

$$L_0\frac{(\alpha-\gamma)}{\tilde{\nu}_*^2}=-\frac{i}{2(ka)^2}. \tag{16c}$$

Solving for the unknown coefficients, one obtains

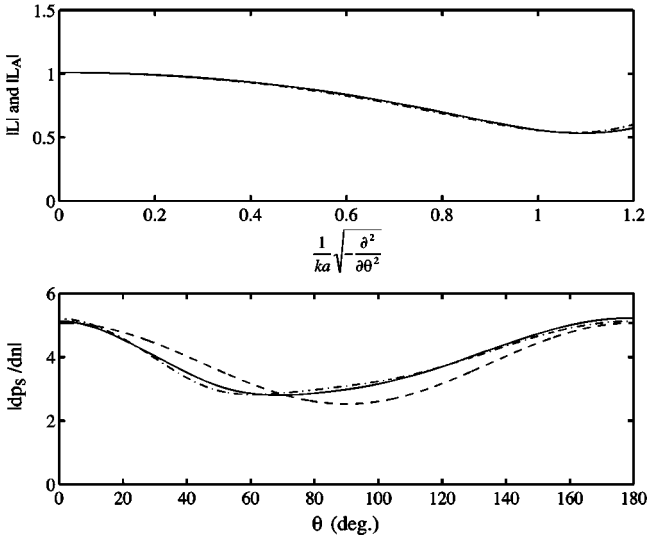


FIG. 1. Comparison of exact [Eq. (8)] (solid) and approximate [Eq. (11)] (dash-dot) D-t-N operators for $ka=5$ (top figure). For the same frequency, the surface value of the normal derivative of the scattered field, computed using the narrow-angle (dashed) and wide-angle (dash-dot) OSRC, is compared with the exact result (solid) for scattering of a plane wave by a soft circular cylinder (zero degrees is the back of the cylinder).

$$\alpha=-\frac{[2+C\tilde{\nu}_*/2L_0+iC\tilde{\nu}_*^3/4L_0^2(ka)^2]}{1+C\tilde{\nu}_*/2L_0},$$

$$\beta=\frac{1+iC\tilde{\nu}_*^3/4L_0^2(ka)^2}{1+C\tilde{\nu}_*/2L_0},$$

$$\gamma=-\frac{[2+C\tilde{\nu}_*/2L_0-i\tilde{\nu}_*^2/2L_0(ka)^2]}{1+C\tilde{\nu}_*/2L_0}. \tag{17}$$

Making the replacement $V^2=-(1/\tilde{\nu}_*^2)\partial^2/\partial\theta^2$ in Eq. (11), a *wide-angle* OSRC operator for a circular cylinder is obtained:

$$L_A=L_0(ka)\frac{(1-(\alpha/\tilde{\nu}_*^2)\partial^2/\partial\theta^2+(\beta/\tilde{\nu}_*^4)\partial^4/\partial\theta^4)}{1-(\gamma/\tilde{\nu}_*^2)\partial^2/\partial\theta^2}. \tag{18}$$

The quality of the fit on the real line is displayed in Figs. 1 and 2 for $ka=5$ and $ka=30$, respectively. Also displayed is the normal derivative of the scattered field on the surface of a soft circular cylinder, found using both wide-angle [Eq. (18)] and narrow-angle [Eq. (9)] OSRCs, compared with exact results. The incident field p_i is a plane wave of unit amplitude, and periodic boundary conditions have been used in constructing the discrete differentiation operator $\partial^2/\partial\theta^2$ [note that in Eq. (7) p is now the scattered field p_s, and the boundary condition is $p_s=-p_i$]. The increased accuracy at 90°, where the incident field grazes the cylinder, is evident in both cases.

Scattering by a rigid circular cylinder is considered next for the same conditions. Results are shown in Fig. 3. Comparison of the surface value of the scattered pressure field again shows the superior accuracy of the wide-angle OSRC. Although the accuracy of Eq. (18) begins to deteriorate below $ka=5$, agreement can be maintained down to $ka=1$ if one uses the four-term rational approximation presented in the previous study.[11]

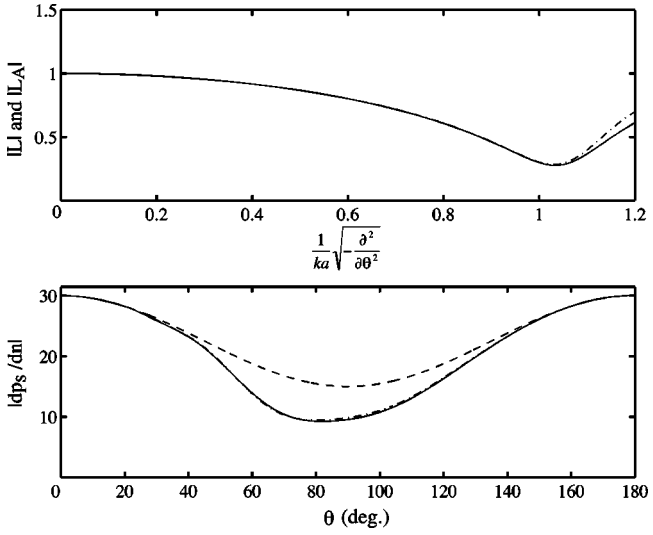


FIG. 2. Same as Fig. 1, but $ka=30$.

III. WIDE-ANGLE OSRC FOR TWO-DIMENSIONAL CONVEX BODIES

Working in analogy with the geometrical theory of diffraction, scattering by a noncircular body can be considered if one approximates the local surface arc length by the arc length s of an osculating circle of radius equal to the local radius of curvature.[1] The following replacements are therefore made in Eqs. (7) and (18),

$$\frac{\partial}{\partial r}\rightarrow\frac{\partial}{\partial n},\quad a\rightarrow\frac{1}{\kappa},\quad \frac{1}{a^2}\frac{\partial^2}{\partial\theta^2}\rightarrow\frac{\partial^2}{\partial s^2}, \tag{19}$$

where κ is the local curvature and n is the outward normal to the surface. Accordingly, the replacements

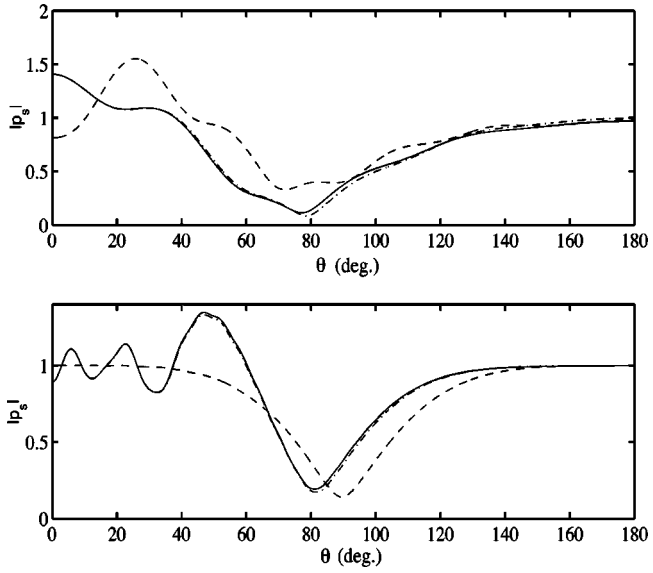


FIG. 3. Scattered surface pressure results using the narrow-angle (dashed) and wide-angle (dash-dot) OSRCs in comparison with the exact result (solid) for scattering of a plane wave by a rigid circular cylinder. Top figure has wavenumber $ka=5$; bottom figure has $ka=30$ (zero degrees is the back of the cylinder).

$$\frac{1}{\tilde{\nu}_*^2}\frac{\partial^2}{\partial\theta^2}\to\frac{1}{k^2}\left(1+(0.4043+0.7003i)\frac{\kappa^{2/3}}{k^{2/3}}+(-0.0727+0.1260i)\frac{\kappa^{4/3}}{k^{4/3}}+\cdots\right)^{-2}\frac{\partial^2}{\partial s^2}, \tag{20}$$

$$\frac{1}{\tilde{\nu}_*^4}\frac{\partial^4}{\partial\theta^4}\to\frac{1}{k^4}\left(1+(0.4043+0.7003i)\frac{\kappa^{2/3}}{k^{2/3}}+(-0.0727+0.1260i)\frac{\kappa^{4/3}}{k^{4/3}}+\cdots\right)^{-4}\frac{\partial^4}{\partial s^4}, \tag{21}$$

and

$$\alpha(ka)\to\alpha(\kappa/k),\quad \beta(ka)\to\beta(\kappa/k),$$
$$\gamma(ka)\to\gamma(\kappa/k),\quad L_0(ka)\to L_0(\kappa/k) \tag{22}$$

are made in Eq. (18) to arrive at the wide-angle OSRC operator for two-dimensional bodies with variable curvature.

As a test case, both wide-angle and narrow-angle OSRCs are applied to scattering by a rigid elliptical cylinder ensonified end-on with an incident plane wave of unit amplitude and wavenumber $k=20$ to compute the far-field directivity $|D(\theta)|$,[11]

$$p_s\sim D(\theta)\frac{\exp(ikr)}{\sqrt{r}}\quad (r\to\infty). \tag{23}$$

The numerical approximation of the derivatives with respect to arc length accounted for the periodic boundary condition as before. The local curvature was computed analytically (a simple numerical approach for computing the curvature is presented in Ref. 9). The results in Fig. 4 are compared with a reference solution computed using the T-matrix method.[13] For the end-on case shown there is an obvious difference in accuracy between the two OSRC methods. Though not shown, for broadside angles of incidence the discrepancy between the narrow- and wide-angle OSRC results is not as great, but the wide-angle OSRC gives better results overall.

IV. EXTENSION TO THREE-DIMENSIONAL BODIES

Although the procedure used to derive the D-t-N map [Eq. (8)] could be adapted to three-dimensional bodies by working with special functions appropriate to particular coordinate systems, it is more transparent to work with the Helmholtz equation in the intrinsic coordinate system (n,s_1,s_2), consider both the narrow and wide-angle scattering limits, and then incorporate these behaviors into a rational approximation as before. Accordingly, the acoustic field near the surface of a general smooth convex body satisfies

$$\left\{\frac{\partial^2}{\partial n^2}+\left(\frac{\kappa_1}{1+n\kappa_1}+\frac{\kappa_2}{1+n\kappa_2}\right)\frac{\partial}{\partial n}+\frac{1}{(1+n\kappa_1)}\frac{\partial}{\partial s_1}\left(\frac{1}{1+n\kappa_1}\frac{\partial}{\partial s_1}\right)+\frac{1}{(1+n\kappa_2)}\frac{\partial}{\partial s_2}\left(\frac{1}{1+n\kappa_2}\frac{\partial}{\partial s_2}\right)+k^2\right\}p=0, \tag{24}$$

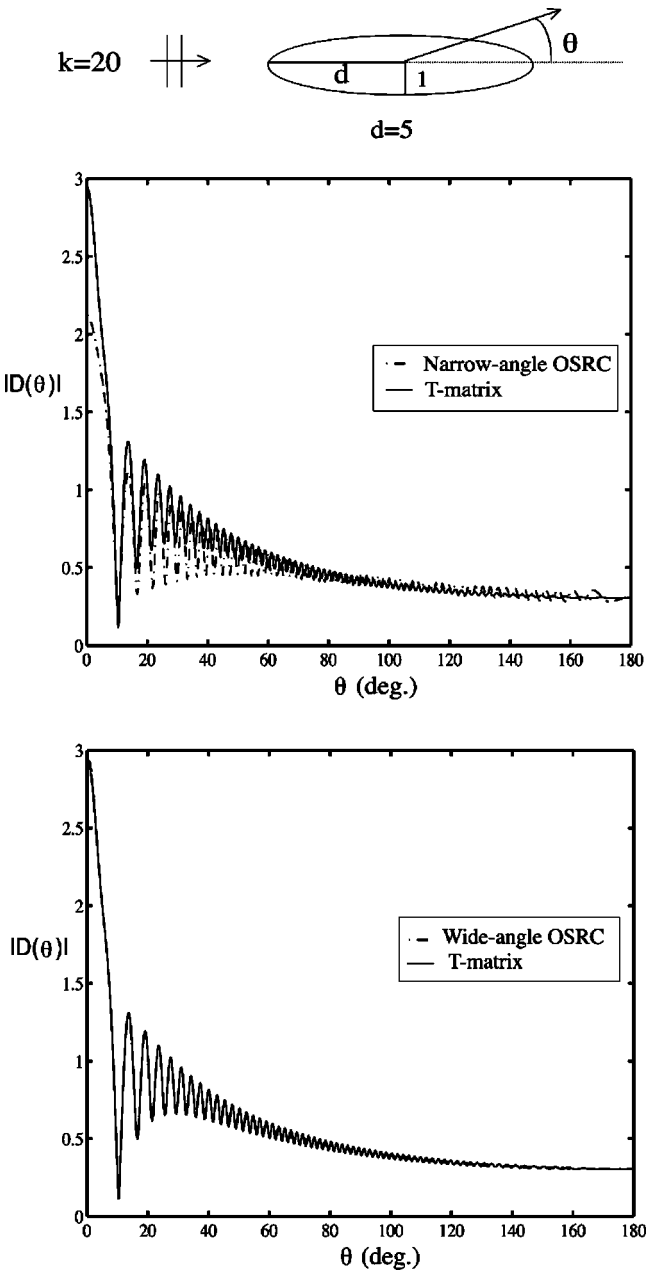


FIG. 4. Far-field directivity patterns for acoustic scattering by a hard elliptical cylinder using the narrow-angle OSRC (top plot) and the wide-angle OSRC (bottom plot).

where s_1 and s_2 are the arc lengths along the principal curvatures κ_1 and κ_2 and n is the outward normal to the surface as depicted, for example, in Fig. 5. Assuming the body is smooth, and neglecting derivatives of the curvatures with respect to arc length, one obtains for $\kappa_1+\kappa_2\ll k$

$$\left\{\frac{\partial^2}{\partial n^2}+(\kappa_1+\kappa_2)\frac{\partial}{\partial n}+\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}+k^2\right)\right\}p=0. \tag{25}$$

In analogy with the derivation of parabolic equations for propagation problems, this two-way wave equation can be factored into a product of incoming and outgoing operators.[14] After first completing the square and neglecting the higher-order term,

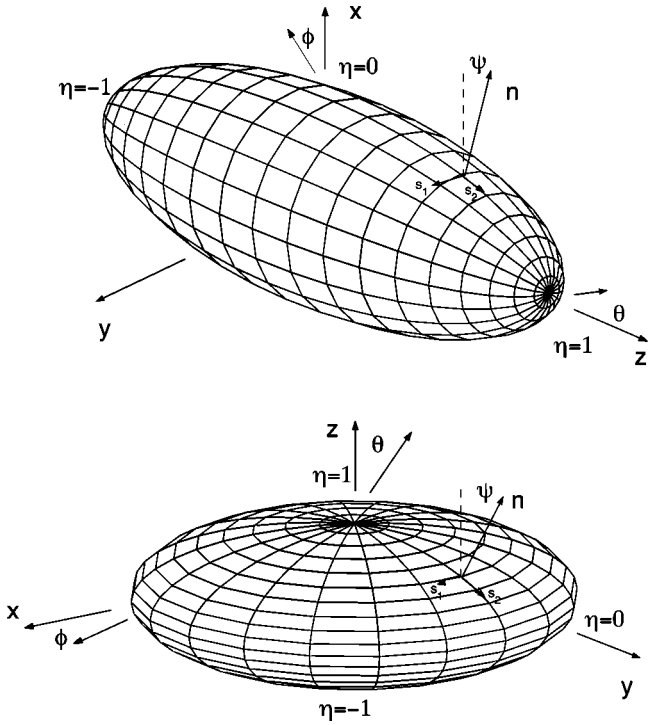


FIG. 5. Prolate (top) and oblate (bottom) spheroids. See Appendices for details.

$$\left\{\left(\frac{\partial}{\partial n}+\frac{\kappa_1+\kappa_2}{2}\right)^2+\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}+k^2\right)\right\}p=0, \tag{26}$$

the factored equation reads

$$\left\{\left(\frac{\partial}{\partial n}+\frac{\kappa_1+\kappa_2}{2}\right)+i\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}+k^2\right)^{1/2}\right\}$$
$$\times\left\{\left(\frac{\partial}{\partial n}+\frac{\kappa_1+\kappa_2}{2}\right)-i\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}+k^2\right)^{1/2}\right\}p=0. \tag{27}$$

The scattered (outgoing) field satisfies

$$\frac{\partial p}{\partial n}=-\frac{\kappa_1+\kappa_2}{2}p+i\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}+k^2\right)^{1/2}p. \tag{28}$$

A. Narrow-angle scattering

For narrow scattering angles, a Taylor expansion of the square-root operator can be made to yield the following three-dimensional OSRC that can be found in Ref. 6:

$$\frac{\partial p}{\partial n}=-\frac{\kappa_1+\kappa_2}{2}p+ikp+\frac{i}{2k}\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}\right)p. \tag{29}$$

While it would appear that an OSRC accurate for wide scattering angles could be constructed by approximating the square-root operator by a rational function, care has to be taken in doing this as the incident and scattered fields are essentially both grazing the light-shadow boundary in a complicated manner and therefore the scattered field is only weakly radiating. As can be seen in Fig. 2 of Calvo *et al.*[11] as well as by comparing the singularities of the D-t-N operator [Eq. (8)] with those of the square-root operator, curvature enters into the wide-angle OSRC in a significant way for large scattering angles. This point is taken up in the following.

B. Wide-angle scattering

Near the light-shadow border where the total pressure wave propagates along the surface in a weakly radiating manner, it can be shown that the field normal to the surface varies on a scale[15]

$$d=O[(k^2\max(\kappa_1,\kappa_2))^{-1/3}]. \tag{30}$$

Using the above to order terms in Eq. (24) (with $\kappa_1+\kappa_2 \ll k$), normalizing $\tilde{n}=kn$, and introducing the symbols

$$\nu_1^2=-\frac{\partial^2}{\partial s_1^2},\quad \nu_2^2=-\frac{\partial^2}{\partial s_2^2}, \tag{31}$$

with $\nu^2=\nu_1^2+\nu_2^2$, one arrives at the Airy equation[16]

$$\left(\frac{\partial^2}{\partial\tilde{n}^2}+1-\frac{\nu^2}{k^2}+2\frac{\kappa_e}{k}\tilde{n}\right)p(\tilde{n};\nu)=0, \tag{32}$$

where

$$\kappa_e=\kappa_1\cos^2\phi+\kappa_2\sin^2\phi. \tag{33}$$

In Eqs. (32) and (33) it has been assumed that the field near the surface propagates predominantly as a plane wave at an angle ϕ with respect to the κ_1 principal curvature direction:

$$\frac{\nu_1}{\nu}=\cos\phi,\quad \frac{\nu_2}{\nu}=\sin\phi. \tag{34}$$

Given the outgoing-wave boundary condition, an appropriate solution of Eq. (32) is given in terms of the traveling-wave Airy function ($w_1=\mathrm{Ai}+i\mathrm{Bi}$),

$$p(\tilde{n};\nu)=w_1(\tilde{d}^2(\nu^2/k^2-1)-\tilde{n}/\tilde{d}), \tag{35}$$

where

$$\tilde{d}\equiv\frac{k}{(2k^2\kappa_e)^{1/3}}. \tag{36}$$

The D-t-N operator on the surface is therefore

$$\frac{\partial p/\partial\tilde{n}}{p}=-\frac{1}{\tilde{d}}\frac{w_1'(\tilde{d}^2(\nu^2/k^2-1))}{w_1(\tilde{d}^2(\nu^2/k^2-1))}, \tag{37}$$

and has a zero at

$$\tilde{\nu}_*=\frac{\nu_*}{\kappa_e}=\frac{k}{\kappa_e}+\frac{1}{2^{1/3}}\left(\frac{k}{\kappa_e}\right)^{1/3}\beta e^{i\pi/3}\ldots \tag{38}$$

[where $w'(\beta e^{i\pi/3})=0$, $\beta=1.0188$], which matches Eq. (10) to $O[(ka)^{1/3}]$ when one makes the replacement $a=1/\kappa_e$. Expanding about this zero and using Eq. (32), one obtains

$$\frac{1}{k}\frac{\partial p/\partial n}{p}=\left(1-\frac{\nu_*^2}{k^2}\right)(\tilde{\nu}-\tilde{\nu}_*), \tag{39}$$

which matches Eq. (14) to leading order. Therefore, provided that the replacements

$$\frac{\partial^2}{\partial s^2} \to \frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}, \quad \kappa \to \kappa_e, \tag{40}$$

are used in Eqs. (20)–(22), and

$$a \to \frac{1}{\kappa_e}, \quad \tilde{\nu}_*(ka) \to \tilde{\nu}_*(\kappa_e/k), \tag{41}$$

are used in computing α, β, γ, and L_0, the wide-angle OSRC can be extended to smooth three-dimensional bodies. Calculation of derivatives in the (s_1, s_2) coordinates can be made, for the sake of illustration in this study, by specifying to particular global coordinate geometries which is done in Sec. V. An expression for the optimal effective curvature for global use is proposed first.

C. Effective curvature

An expression that captures the effective curvature needed for accuracy near grazing angles of incidence, while providing a smooth transition to the narrow-angle limit, is the following:

$$\kappa_e = |(\hat{k} \times \hat{n}) \cdot \hat{s}_1|^2 \kappa_2 + |(\hat{k} \times \hat{n}) \cdot \hat{s}_2|^2 \kappa_1 + |\hat{k} \cdot \hat{n}|^2 \frac{(\kappa_1 + \kappa_2)}{2}, \tag{42}$$

where the hat-quantities are unit vectors ($\hat{k}$ is the direction of the incident field). Equation (42) reduces to Eq. (33) in the grazing regime, while it recovers the mean curvature in Eq. (29) in the near-normal incidence regime where narrow-angle scattering predominates.

V. APPLICATION TO SPHEROIDS

A. Discretization

Although the OSRC can in principle be applied to any convex body, as a demanding test of accuracy it is applied here to scattering by prolate (long and slender) and oblate (short and fat) impenetrable spheroids as shown in Fig. 5. Spheroids are good tests of the OSRC because of the predominance of creeping waves for end-on incidence and the large local variations of the principal radii of curvature for large aspect ratios. The details of the prolate and oblate coordinates systems,[17] and the expressions for the derivatives used in Eq. (40) for these systems, are discussed in the Appendices. The discretization uses N uniformly spaced points between the poles (on a longitudinal path) and M circumferential points (on a latitudinal path) as indicated by the grid in Fig. 5. Points were numbered in a ''spiral'' pattern starting near a pole.

Because the surface Laplacian features a singular point at the poles of the spheroid, care must be taken in the discretization. For this study, the longitudinal discrete points nearest the poles were located a longitudinal half-step away from the poles, therefore bypassing a need to evaluate at the singular point. A similar approach was used with success by Murch.[8] Periodicity is imposed in the lateral and longitudinal directions. Tests indicated good convergence as the step size was reduced. For the results presented in the figures, the computational grid was refined to the point where further changes are hard to perceive in the plot. Approximately 8–12 points per wavelength were used to obtain a converged solution.

Only one application of the D-t-N map [Eq. (7) with r replaced by n] needs to be made. For the scattering problem $p \to p_s$, and either $p_s = -p_i$ (pressure release) or $\partial p_s/\partial n = -\partial p_i/\partial n$ (rigid) boundary conditions apply.

B. Examples

Comparison will focus on the total surface pressure (for hard obstacles) or its normal derivative (for soft obstacles) and the far-field directivity which for three dimensions can be computed using

$$p_s(r, \theta, \phi) = \frac{D(\theta, \phi) e^{ikr}}{r} \quad (r \to \infty), \tag{43}$$

$$p_s(\mathbf{x}) = \int_S \int \left(p_{\text{tot}}(\mathbf{x}') \frac{\partial g}{\partial n'}(\mathbf{x}, \mathbf{x}') - \frac{\partial p_{\text{tot}}}{\partial n'}(\mathbf{x}') g(\mathbf{x}, \mathbf{x}') \right) dS(\mathbf{x}'), \tag{44}$$

$$g(\mathbf{x}, \mathbf{x}') = \frac{\exp(ikr)}{4\pi r}, \quad r = |\mathbf{x} - \mathbf{x}'|, \tag{45}$$

where $\mathbf{x}' \in S$. Combining the previous relations, using the law of cosines and $|\mathbf{x}'| \ll |\mathbf{x}|$ one obtains

$$D(\theta, \phi) = -4\pi \int_S \int \left(ik \hat{\mathbf{n}}' \cdot \hat{\mathbf{x}} p_{\text{tot}}(\mathbf{x}') + \frac{\partial p_{\text{tot}}}{\partial n'}(\mathbf{x}') \right) \times \exp(-ik \hat{\mathbf{x}} \cdot \mathbf{x}') dS(\mathbf{x}'), \tag{46}$$

where $\hat{\mathbf{x}}$ is directed towards the observation point and S is the surface of the body.

As a first example, a rigid sphere (aspect ratio $A = 1$) is ensonified by a plane wave. Figure 6 compares the wide-angle and narrow-angle results with a reference solution computed using the boundary element method.[18] An outstanding result is the superior focusing of the total surface pressure near the back of the sphere. This is naturally because of the improved treatment of diffractive effects near the light-shadow border. Better agreement for the far-field directivity is also obtained using the wide-angle OSRC. To test variation with frequency, the backscattered directivity is plotted in Fig. 7.

A rigid prolate spheroid of aspect ratio $A = 5$ (semi-minor axis has unit length) is considered next in Fig. 8 for an end-on ($+z$ direction in Fig. 5) incident plane wave of wavenumber $k = 20$. The wide-angle OSRC result is in better agreement with the BEM result overall. Adjusting the angle of incidence to the broadside direction (the $+x$ direction in Fig. 5), results for $A = 3$ and $k = 15$ are displayed in Fig. 9. This is a truly three-dimensional example; the former examples are axisymmetric. Again, agreement is better with the wide-angle OSRC although the narrow-angle OSRC does a

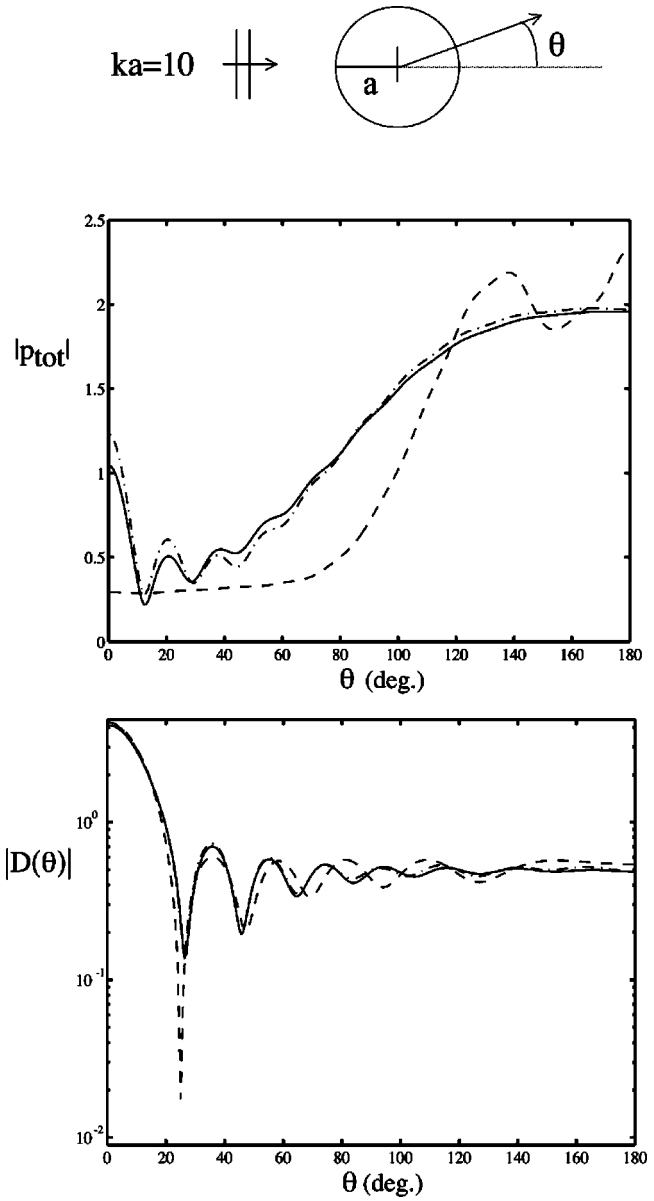


FIG. 6. Scattering of an incident plane wave of wavenumber $ka=10$ by a rigid sphere. Total surface pressure (top plot) and far-field directivity (bottom plot) are displayed. Narrow-angle (dashed) and wide-angle (dash-dot) OSRC results are compared with BEM results (solid).

good job owing to the lesser contribution of creeping waves. Next, the accuracy of the D-t-N map is directly examined for the same spheroid and wavenumber but the wavevector $\hat{k}$ is tilted 45° towards the z axis in the $x-z$ plane. Excellent agreement is seen with the BEM results in Fig. 10. Finally, for the same conditions, but now a soft (pressure release) boundary condition, a directivity comparison is made with the BEM result in Fig. 11.

Turning finally to oblate spheroids, two examples are considered for an aspect ratio $A=2$ and wavenumber $k=10$. Good results are obtained for both end-on (Fig. 12) and broadside (Fig. 13) incidence.

C. Cost comparison

For the examples considered, the cost of the wide-angle OSRC method was compared with a standard (non-

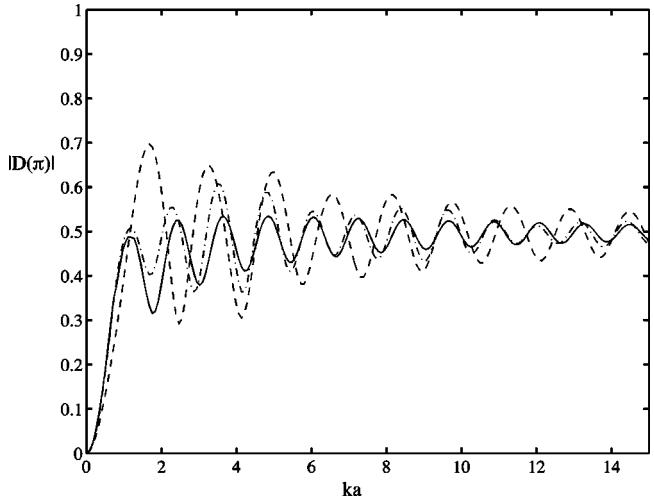


FIG. 7. Backscattering from a rigid sphere as a function of frequency. Narrow-angle (dashed) and wide-angle (dash-dot) OSRC results are compared with the exact series solution result (solid).

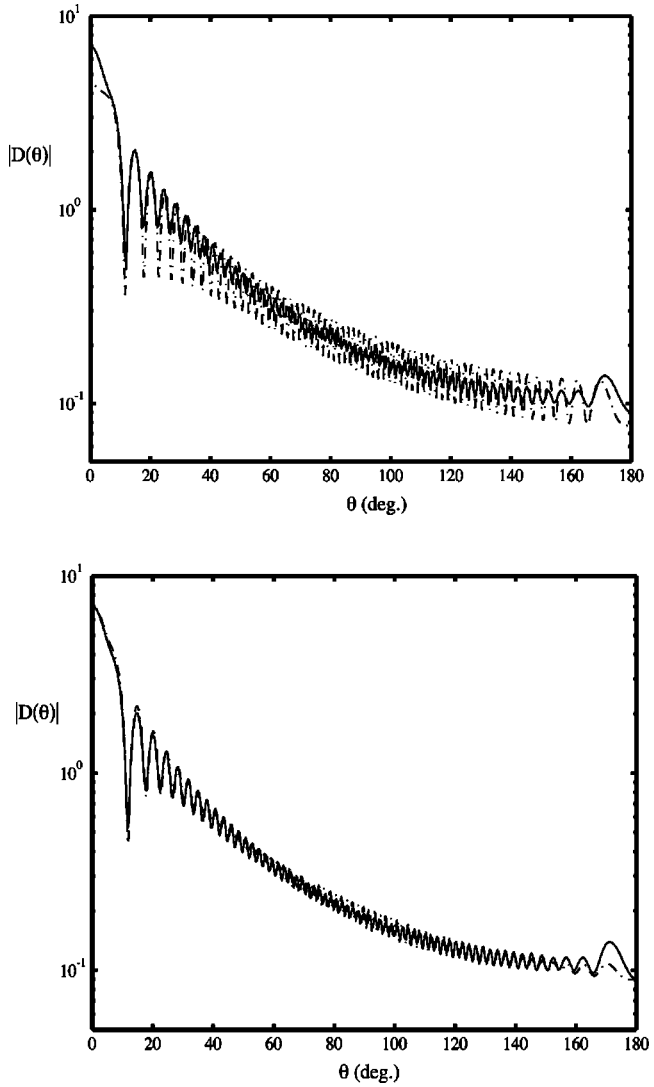


FIG. 8. Directivity for scattering of a plane wave with $k=20$, incident in the $+z$ direction (end-on), by a rigid prolate spheroid of aspect ratio $A=5$. BEM result (solid) is compared with narrow-angle OSRC (top plot) and wide-angle OSRC (bottom plot) results.

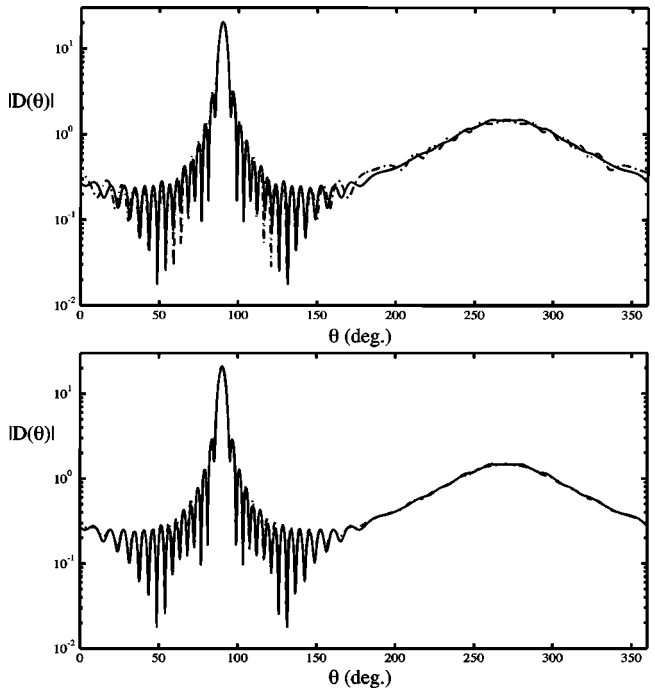


FIG. 9. Far-field directivity (in the x–z plane) for scattering of a plane wave with $k=15$, incident in the $+x$ direction (broadside), by a rigid prolate spheroid of aspect ratio $A=3$. BEM result (solid) is compared with narrow-angle OSRC (top plot) and wide-angle OSRC (bottom plot) results.

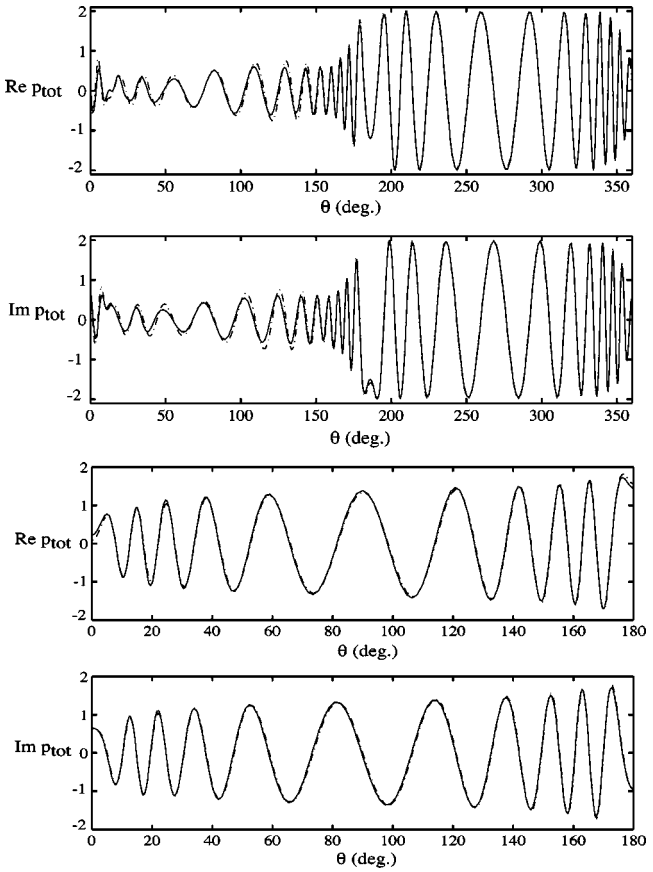


FIG. 10. Real and imaginary parts of the total surface pressure ($p_{\text{tot}}=p_i+p_s$) for scattering by a plane wave ($k=15$) incident obliquely ($\varphi=0$, $\theta=45°$) on a rigid prolate spheroid ($A=3$). Solid lines are BEM reference solutions compared with wide-angle OSRC results. Top two figures: pressure along back and front surfaces of spheroid (in the $y=0$ plane); bottom two figures: pressure along side of spheroid (in the $x=0$ plane).

axisymmetric) BEM method used to compute the reference solutions.[18] For oblique incidence on an axisymmetric geometry, one can alternatively decompose the incident field into circumferential modes and then solve a sequence of axisymmetric boundary-element problems[19] for each mode. The estimated cost of this technique was also compared with the basic wide-angle OSRC cost. To obtain an even more competitive OSRC cost, the same manner of solution was also applied to the OSRC. Hence, a sequence of reduced matrix bandwidth problems was solved in that case.

For two-dimensional problems, in which the discrete surface Laplacian is simply a tridiagonal matrix, computational cost of the OSRC method scales linearly with the wavenumber whereas the costs of most standard methods scale with the square of the wavenumber. In the three-dimensional case, scaling comparison depends on the orientation of the incident wave with a major axis of rotation of a body. For nearly end-on incidence, few circumferential grid points are needed for a converged solution and hence cost scales much like the two-dimensional problem. For general bodies, however, the matrix bandwidth will be larger for an equally spaced mesh in both surface directions. The OSRC is still expected to be more cost effective because of sparsity and the lack of having to compute singular integrals in the (nonlocal) integral equation method.

Examples showed the wide-angle OSRC to cost less in absolute terms as can be seen in Table I. The normalized times shown include not only the time to compute the unknown surface quantities, but also a computation of the far-field pressure over 1000 field points. The comparisons between the third and fifth columns are in fact conservative since the BEM calculation made use of the plane of symmetry and hence the number of points in a latitude section in the computation is actually $M/2$ whereas the OSRC method generally uses M points. Also, the BEM codes are machine-optimized executables whereas the OSRC code was

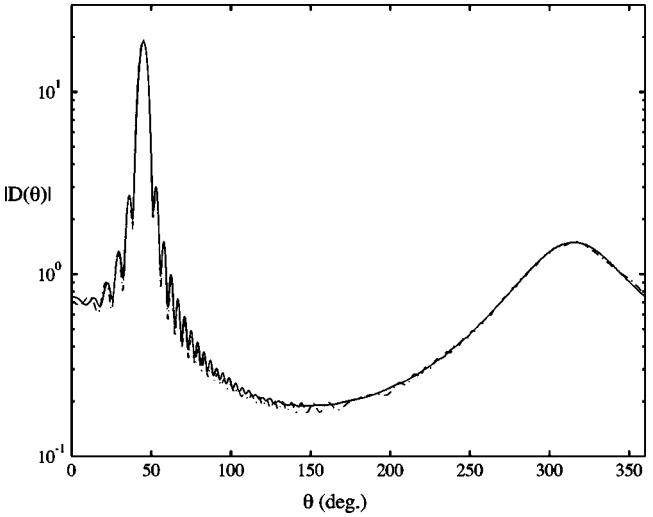


FIG. 11. Far-field directivity for scattering by a soft prolate spheroid of aspect ratio $A=3$ and wavenumber $k=15$. Plane wave is incident obliquely ($\varphi=0$, $\theta=45°$). BEM result (solid) is compared with wide-angle OSRC result (dash-dot).

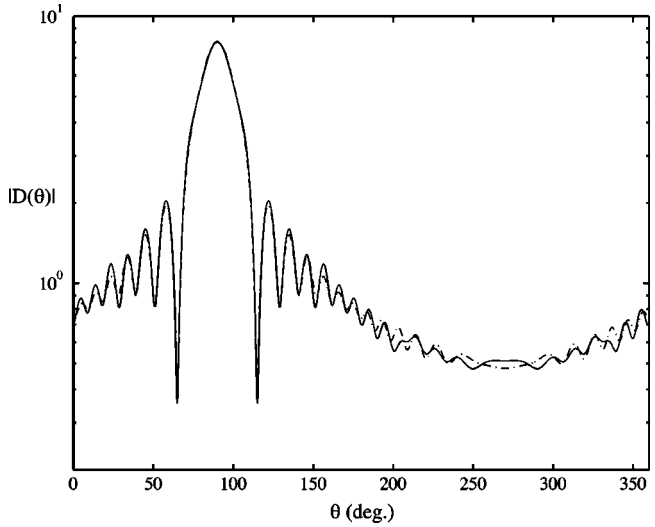


FIG. 12. Far-field directivity for scattering by a rigid oblate spheroid of aspect ratio $A=2$ and frequency $k=10$. Plane wave is incident end-on (in the $+x$-direction). BEM result (solid) is compared with wide-angle OSRC result (dash-dot).

programmed simply using a *Matlab* script environment. It is clear that for higher frequency problems where more points are needed per wavelength, the BEM cost goes up significantly. Exhaustive comparison with other methods (the fast multipole method, for example) is not taken up in this study.

VI. DISCUSSION

The wide-angle OSRC, here presented in a simplified form, gave better results than the narrow-angle OSRC for about the same amount of computational effort for acoustically hard cases (i.e., one solution of $Ax=b$ is needed for both the narrow and wide-angle OSRC types for this boundary condition). Given the speed of the algorithm, utility might be found for iterative/inverse problems such as shape determination or for computing scattering in the time-domain by Fourier synthesis.

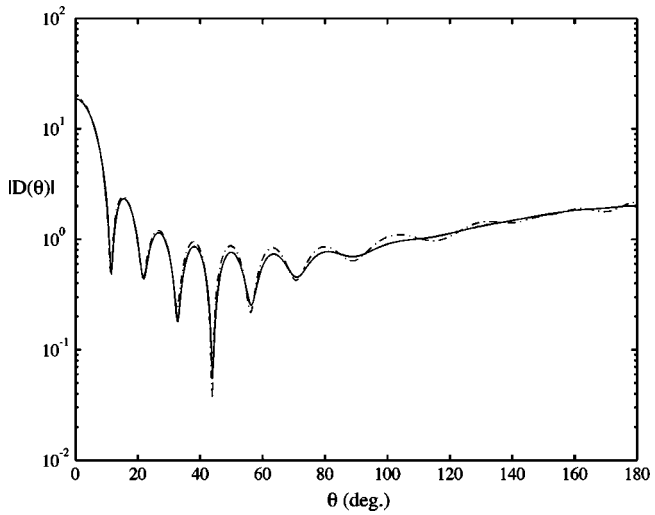


FIG. 13. Far-field directivity (in the $x-z$ plane) for scattering by a rigid oblate spheroid of aspect ratio $A=2$ and frequency $k=10$. Plane wave is incident broadside (in the $+z$ direction). BEM result (solid) is compared with wide-angle OSRC (dash-dot) result.

TABLE I. Normalized computational cost comparison among methods applied to scattering of an obliquely incident plane wave. Displayed are results using a standard BEM code,[18] an axisymmetric BEM code applied to M circumferential modes (estimate), the wide-angle OSRC code, and an axisymmetric version of the wide-angle OSRC code applied to M circumferential modes. The reference time is 0.53 min using an R10000 processor. See text for discussion.

N	M	Standard BEM	Axisymmetric BEM	OSRC	Axisymmetric OSRC
35	32	4.33	1.36	1.02	1.00
65	64	58.5	6.55	2.96	2.62
25	248	151	2.83	2.62	1.68
251	32	332	109	10.1	7.28
451	32	1672	589	14.9	13.4
81	184	1309	20.8	13.0	4.85
91	184	1728	28.3	14.9	5.60
121	184	6326	58.5	15.9	7.57

Numerical application to non-axisymmetric bodies can best be made using a finite-element formulation of the wide-angle OSRC which will be reported in further work. A finite-element formulation of a narrow-angle OSRC has been carried out by Antoine.[9] The weak-form framework is helpful for examining discontinuities in surface slope (infinite curvature) which have not been treated here. As a practical guideline for applying Eq. (11) to bodies with corners and very flat surfaces, best accuracy can be obtained by rounding corners such that the minimum k/κ is 5 and giving slight curvature to flat surfaces so that the maximum value does not exceed 1500. If one uses the four-term rational approximation described by Calvo *et al.*, these restrictions can be relaxed further, although some limiting may still be necessary.

Finally, the wide-angle OSRC is a Dirichlet-to-Neumann map which represents the surface impedance. Therefore, application to radiation problems can be made, and in this respect there is a connection with the work of DiPerna and Feit[20] for the thin elastic plate. Preliminary work on locally excited radiation on spheroidal bodies gives good agreement with solutions computed using the boundary element method. In addition, application of the present work to problems involving scattering by penetrable bodies can be made.

APPENDIX A: PROLATE SPHEROID

The geometry depicted in the top of Fig. 5 is described by

$$x=\sqrt{1-\eta^2}\cos\phi, \quad y=\sqrt{1-\eta^2}\sin\phi, \quad z=A\eta, \quad \text{(A1)}$$

where $\eta=\cos u$, $0\leqslant u\leqslant\pi$, and $0\leqslant\phi<2\pi$. The aspect ratio is therefore given by A. The physical angle θ is related to u by

$$\tan\theta=\frac{\tan u}{A}, \quad \text{(A2)}$$

and the angle the unit normal makes with the x axis, ψ, is given by

$$\tan\psi=\frac{\cot u}{A}. \quad \text{(A3)}$$

In both Eqs. (A2) and (A3), $0\leqslant u\leqslant\pi/2$, but in view of symmetry, one can extend the angles to $0\leqslant\theta\leqslant\pi$ and $-\pi/$

$2 \leqslant \psi \leqslant \pi/2$. The principal curvatures κ_1 and κ_2 are given by

$$\kappa_1 = \frac{A}{(\cos^2 u + A^2 \sin^2 u)^{1/2}}, \quad \kappa_2 = \frac{A}{(\cos^2 u + A^2 \sin^2 u)^{3/2}}. \tag{A4}$$

In terms of u, the derivatives needed for the surface Laplacian are given by

$$\frac{\partial^2}{\partial s_1^2} = \frac{1}{A^2 - (A^2-1)\cos^2 u} \cot u \frac{\partial}{\partial u} + \frac{1}{\sin^2 u} \frac{\partial^2}{\partial \phi^2}, \tag{A5}$$

$$\frac{\partial^2}{\partial s_2^2} = \frac{1}{A^2 - (A^2-1)\cos^2 u} \frac{\partial^2}{\partial u^2}. \tag{A6}$$

The unit normal to the surface is given by

$$\hat{n} = \cos\psi(\cos\phi \hat{\mathrm{x}} + \sin\phi \hat{\mathrm{y}}) + \sin\psi \hat{\mathrm{z}}. \tag{A7}$$

In the limit $A \to 1$, the above expressions reduce to the familiar ones found in spherical coordinates.

APPENDIX B: OBLATE SPHEROID

The geometry depicted in bottom of Fig. 5 is described by

$$x = A\sqrt{1-\eta^2}\cos\phi, \quad y = A\sqrt{1-\eta^2}\sin\phi, \quad z = \eta, \tag{B1}$$

where $\eta = \cos u$, $0 \leqslant u \leqslant \pi$, and $0 \leqslant \phi < 2\pi$. The physical angle θ is related to u by

$$\tan\theta = A \tan u, \tag{B2}$$

and the angle the unit normal makes with the z axis, ψ, is given by

$$\tan\psi = \frac{\tan u}{A}. \tag{B3}$$

Again, in both Eqs. (B2) and (B3), $0 \leqslant u \leqslant \pi/2$, but in view of symmetry, one can extend the angles to $0 \leqslant \theta \leqslant \pi$ and $0 \leqslant \psi \leqslant \pi$. The principal curvatures κ_1 and κ_2 in this case are given by

$$\kappa_1 = \frac{(\sin^2 u + A^2 \cos^2 u)^{1/2}}{(A \sin^2 u + A^3 \cos^2 u)},$$

$$\kappa_2 = \frac{A}{(\sin^2 u + A^4 \cos^2 u)^{1/2}(\sin^2 u + A \cos^2 u)}. \tag{B4}$$

In terms of u, the derivatives appearing in the surface Laplacian are given by

$$\frac{\partial^2}{\partial s_1^2} = \frac{1}{1 + (A^2-1)\cos^2 u} \cot u \frac{\partial}{\partial u} + \frac{1}{A^2 \sin^2 u} \frac{\partial^2}{\partial \phi^2}, \tag{B5}$$

$$\frac{\partial^2}{\partial s_2^2} = \frac{1}{1 + (A^2-1)\cos^2 u} \frac{\partial^2}{\partial u^2}. \tag{B6}$$

The unit normal to the surface is given by

$$\hat{n} = \sin\psi(\cos\phi \hat{\mathrm{x}} + \sin\phi \hat{\mathrm{y}}) + \cos\psi \hat{\mathrm{z}}. \tag{B7}$$

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research. The author thanks Mike Collins for discussions on this topic and Lou Dragonette and Dalcio Dacol for comments on the manuscript.

[1] G. A. Kriegsmann, A. Taflove, and K. R. Umashankar, "A new formulation of electromagnetic wave scattering using an on-surface radiation boundary condition approach," IEEE Trans. Antennas Propag. **35**, 153–161 (1987).
[2] J. B. Keller and D. Givoli, "Exact non-reflecting boundary conditions," J. Comput. Phys. **82**, 172–192 (1989).
[3] A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210 (1971).
[4] D. S. Jones, "Surface radiation conditions," IMA J. Appl. Math. **41**, 21–30 (1988).
[5] D. S. Jones and G. A. Kriegsmann, "Note on surface radiation conditions," SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **50**, 559–568 (1990).
[6] D. S. Jones, "An improved surface radiation condition," IMA J. Appl. Math. **48**, 163–193 (1992).
[7] M. Teymur, "A note on higher-order surface radiation conditions," IMA J. Appl. Math. **57**, 137–163 (1996).
[8] R. D. Murch, "The on-surface radiation condition applied to three-dimensional convex objects," IEEE Trans. Antennas Propag. **41**, 651–658 (1993).
[9] X. Antoine, "Fast approximate computation of a time-harmonic scattered field using the on-surface radiation condition method," IMA J. Appl. Math. **66**, 83–111 (2001).
[10] D. C. Calvo, M. D. Collins, and D. K. Dacol, "A higher-order on-surface radiation condition derived from an analytic approximation of a Dirichlet-to-Neumann map," J. Acoust. Soc. Am. **110**, 2700 (2001).
[11] D. C. Calvo, M. D. Collins, and D. K. Dacol, "A higher-order on-surface radiation condition derived from an analytic approximation of a Dirichlet-to-Neumann Map," IEEE Trans. Antennas Propag. **51**, 1607–1614 (2003).
[12] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice–Hall, Englewood Cliffs, NJ, 1973), pp. 691–718.
[13] P. C. Waterman, "New formulation of acoustic scattering," J. Acoust. Soc. Am. **45**, 1417–1429 (1969).
[14] M. D. Collins, "Higher-order Padé approximations for accurate and stable elastic parabolic equations with applications to interface wave propagation," J. Acoust. Soc. Am. **89**, 1050–1057 (1991).
[15] V. M. Babich and V. S. Buldyrev, *Short-Wavelength Diffraction Theory: Asymptotic Methods*, Springer Series on Wave Phenomena, Vol. 4 (Springer Verlag, Berlin, 1991).
[16] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1989), p. 477.
[17] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[18] T. W. Wu, *Boundary Element Acoustics* (WIT, Boston, 2000).
[19] S. M. Kirkup, *The Boundary Element Method in Acoustics* (Integrated Sound Software, Hebden Bridge, UK, 1998).
[20] D. T. DiPerna and D. Feit, "An approximate Green's function for a locally excited fluid-loaded thin elastic plate," J. Acoust. Soc. Am. **114**, 194–199 (2003).

Implicit methods of solution to integral formulations in boundary element method based nearfield acoustic holography

Nicolas Valdivia[a] and Earl G. Williams
Code 7130, Naval Research Laboratory, Washington, DC 20375

(Received 8 September 2003; revised 17 May 2004; accepted 10 June 2004)

Two integral equation methods are considered which can be inverted to provide the surface velocity and/or pressure given a measurement of the pressure on an imaginary surface in the nearfield of a vibrating or scattering body. This problem is central to nearfield acoustical holography (NAH). The integral equations are discretized using boundary element methods (BEMs). The integral equation methods considered are (1) an indirect formulation method based on the single layer integral equation and (2) a direct formulation method based on a system of equations derived from the Helmholtz–Kirchhoff integral equation. The formation of integral equations from the mentioned methods will not involve the explicit inversion of matrices, but instead will require this inversion to be done implicitly. Since these methods back-track the sound field from the measurement surface to the surface of the source/vibrator they are ill-posed in nature and Tikhonov regularization are used to stabilize the reconstruction. Problems are noted with the use of higher order shape functions in the BEM discretization which are detrimental to the inversion, and it is shown that the breakdown of the method at the interior resonance frequencies can be mitigated successfully. These two approaches are compared with the conventional direct method using measured experimental data from a point-driven rectangular plate. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777854]

PACS numbers: 43.40.Sk, 43.20.Rz, 43.40.Yq [MO] Pages: 1559–1572

I. INTRODUCTION

The reconstruction of the acoustic field (pressure, normal velocity, and intensity) plays an important role in the early stage of effective noise control. A powerful technique used for this reconstruction is called nearfield acoustical holography (NAH).[1] NAH has had a tremendous success analyzing sources with geometries which conform closely to one of the separable geometries of the acoustic wave equation; for example, planar, cylindrical, and spherical geometries. In these analyses, the pressure field radiated or scattered from an object is measured on an imaginary surface outside (for exterior problems) or inside (for interior problems) the source. NAH uses this pressure hologram to reconstruct the acoustic field on the body of the source, an ill-posed inverse problem. The solution of the inverse problem in these geometries relies on the expansion of the pressure field in terms of a complete set of eigenfunctions along with a knowledge of the analytical form of the Neumann or Dirichlet Green's function. The reconstructions are very efficient, requiring only fractions of a second of computation time per frequency. Sources with boundaries which vary appreciably in shape from one of these separable geometries can be attacked using boundary element methods (BEMs) which yields a numerical form of the appropriate Green's function, along with the singular value decomposition (SVD) which provides orthonormal eigenfunctions for the geometry. With these eigenfunctions, the methodology of the inversion is identical to that used for the separable geometry. One pays, however, for the expanded versatility with a considerable increase in computation time. This generalization of NAH is called IBEM—inverse boundary element method.

Inverse boundary element methods are broken into two classes of formulations: the direct formulations which spring from the Helmholtz–Kirchhoff integral equation (HIE) and the indirect formulations which manipulate the HIE and solve for a fictitious source distribution. The first papers in IBEM-NAH used the direct formulation and appeared in 1988.[2–4] An extensive comparison of different approaches to direct formulation IBEM was given in 1992.[5] Complicated underwater axisymmetric vibrators were studied using the direct formulation and conformal measurement sets in the early 1990s.[6–8] Precise and automated approaches to regularization[9,10] of the inversion first appeared in 1996 for the direct formulation,[11] and general discussions of regularization useful for NAH and IBEM-NAH appeared later.[12–16] Other researches of note have appeared on the direct formulation IBEM,[17–19] and on methods using spherical harmonic expansions in an effort to improve accuracy.[20]

The second popular approach to NAH using IBEM is the indirect formulation in which fictitious source distributions, which are the kernels of single and/or double layer integral representations, are used as an intermediate step between the measured pressure and the reconstructed velocity or pressure fields.[21–24] We call these methods indirect *implicit* formulations since the creation of the transfer function between the surface velocity (or pressure) and the measured pressure does not require a matrix inversion. The last two cited papers, in fact, make use of this property and thus are particularly attractive for very large matrices since iterative approaches can be used, although formulation of effective regularization methods is a major issue here. A larger body of work, which

[a]Electronic mail: valdivia@pa.nrl.navy.mil

are *explicit* in nature since they require computation and inversion of full matrices, use a variational principle in the indirect formulation and start with the double layer integral representation.[25–31] This variation approach is sometimes referred to as IVBEM (indirect variational boundary element method) and it is being used in several major commercial computer codes.

We present in this paper two implicit approaches which have not been investigated in the literature, a direct formulation based on the solution of a system of integral equations and an indirect formulation based on the single layer. The latter is attractive for its theoretical and computational simplicity. It does not require computation of a hyper-singular kernel needed for the variational approaches discussed above and does not require any intermediate matrix inversions. We compare these approaches to the traditional direct formulation to illustrate their accuracy.

II. BOUNDARY INTEGRAL REPRESENTATION

Let D be a domain in $\mathbb{R}^3$, interior to the boundary surface Γ where we assume that Γ is allowed to have edges and corners. Similarly we will denote as D^+ the region outside of D that shares the same boundary Γ. For a time-harmonic ($e^{-i\omega t}$) disturbance of frequency ω the sound pressure p satisfies the homogeneous Helmholtz equation in D (or D^+),

$$\Delta p + k^2 p = 0, \tag{1}$$

where $k=\omega/c$ is the wave number and c is the constant for the speed of sound. A solution p that satisfies Eq. (1) in D^+ also needs to satisfy the Sommerfeld radiation condition (see Ref. 32).

In this work we will consider exterior NAH, where the acoustical sensors are placed on a surface Γ_0 outside the domain D. These are used to measure the pressure p and the fundamental problem is to recover the acoustic field (pressure, normal velocity, and normal intensity) on Γ. The solution for NAH by the *direct formulation* is based on the Helmholtz–Kirchhoff integral equation (HIE), for $\mathbf{x}=(x_1,x_2,x_3)\in D^+$

$$p(\mathbf{x}) = (Dp)(\mathbf{x}) - i\omega\rho(Sv)(\mathbf{x}), \tag{2}$$

which is a solution to Eq. (1) (see Ref. 32). In Eq. (2) we use an integral operator notation similar to[33,32]

$$(S\varphi)(\mathbf{x}) := \int_\Gamma \Phi(\mathbf{x},\mathbf{y})\varphi(\mathbf{y})dS(\mathbf{y}), \quad \mathbf{x}\in\mathbb{R}^3, \tag{3}$$

$$(D\varphi)(\mathbf{x}) := \int_\Gamma \frac{\partial\Phi(\mathbf{x},\mathbf{y})}{\partial\mathbf{n}(\mathbf{y})}\varphi(\mathbf{y})dS(\mathbf{y}), \quad \mathbf{x}\in D^+, \tag{4}$$

where Φ is the *free space Green's function* to the Helmholtz equation

$$\Phi(\mathbf{x},\mathbf{y}) = \frac{\exp(ik|\mathbf{x}-\mathbf{y}|)}{4\pi|\mathbf{x}-\mathbf{y}|}. \tag{5}$$

In Eq. (2), ρ is the mean fluid density and in Eq. (4), $\mathbf{n}$ is the unit interior normal to Γ. As described in Ref. 33, Theorem 2.13, there is a jump relation in Eq. (4) when $\mathbf{x}\in\Gamma$,

$$(D\varphi)^+(\mathbf{x}) := \int_\Gamma \frac{\partial\Phi(\mathbf{x},\mathbf{y})}{\partial\mathbf{n}(\mathbf{y})}\varphi(\mathbf{y})dS(\mathbf{y}) + \Omega(\mathbf{x})\varphi(\mathbf{x}), \tag{6}$$

where $\Omega(\mathbf{x})$ is the solid angle coefficient given by the integral formula

$$\Omega(\mathbf{x}) = -\int_\Gamma \frac{\partial}{\partial\mathbf{n}(\mathbf{y})}\left(\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\right)dS(\mathbf{y}), \quad \mathbf{x}\in\Gamma. \tag{7}$$

$\Omega(\mathbf{x})$ will be $\frac{1}{2}$ for all $\mathbf{x}\in\Gamma$, when Γ is a smooth surface.

The *indirect formulations* are based on representations to the solution of Eq. (1) like the single source (or single layer) representation[21–24]

$$p(\mathbf{x}) = (S\varphi)(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^3, \tag{8}$$

where φ is the density function, and the double source (or double layer) representation[25–31]

$$p(\mathbf{x}) = (D\varphi)(\mathbf{x}), \quad \mathbf{x}\in D^+. \tag{9}$$

The double source representation is used to avoid the failure of HIE when D is a thin shape body or a regular body with appendages.

The traditional approach for solving NAH (both for direct and indirect formulations) is based on the integral equations

$$p(\mathbf{x}) = \int_\Gamma \mathcal{G}(\mathbf{x},\mathbf{y})p(\mathbf{y})dS(\mathbf{y}), \quad \mathbf{x}\in\Gamma_0, \tag{10}$$

and

$$p(\mathbf{x}) = \int_\Gamma \mathcal{G}'(\mathbf{x},\mathbf{y})v(\mathbf{y})dS(\mathbf{y}), \quad \mathbf{x}\in\Gamma_0, \tag{11}$$

where the functions $\mathcal{G}$ and $\mathcal{G}'$ are respectively the *Dirichlet Green's function* and *Neumann Green's function* for Eq. (1). These integral equations are similar to the Rayleigh's integral formulas for planar NAH (see Ref. 1), since an explicit relationship between p on the measurement surface Γ_0 and p, v on Γ is given. In general, the knowledge of an explicit formula for these functions will be helpful, but can only be found for some restricted geometries (see Refs. 34–36). In the acoustical literature, to overcome this difficulty for general surfaces, effective numerical approximations have been applied. The common practice (see Refs. 5, 11, 14, and 29) will be to combine the direct formulation in Eq. (2) with Eq. (6) to produce a numerical approximation of the operators given by Eqs. (10) and (11). We will refer to this particular approach as the *direct-explicit* (DE) approach, and provide a more detailed discussion in the next section. Similarly in Refs. 25–31 the indirect formulation in Eq. (9) is used with a variational approach to create a numerical approximations of Eqs. (10) and (11). The variational approach is used to overcome some of the difficulties related to the hypersingular integral equation created by the normal derivative of the integral operator in Eq. (9). This approach is known as the inverse variational boundary elements (IVBEMs). We will not discuss IVBEM any further, since previous comparisons with the DE approach are available in the cited literature.

As explained above, the traditional approach uses either direct or indirect formulations to produce a numerical approximation of the transfer functions in Eqs. (10) and (11). There are special approaches that will not require these numerical approximations of the transfer functions, and for that reason will be called *implicit* approaches. In this work our interest will focus on two particular implicit approaches: the *direct-implicit* (DM) approach and the *indirect-implicit* (IM) approach. The DM approach uses the direct formulation in Eq. (2) combined with Eq. (6) to construct the *Helmholtz–Kirchhoff system* (suggested in Ref. 37). The solution to this system will provide the reconstruction of both pressure and normal velocity. The explicit formula for this system will be given in Sec. III. Finally, the IM approach uses the indirect formulation in Eq. (8) (as suggested in Refs. 23, 22, and 24) to solve

$$p(\mathbf{x}) = (S\varphi)(\mathbf{x}), \quad \mathbf{x} \in \Gamma_0, \tag{12}$$

for the density φ. Then use φ to obtain

$$v(\mathbf{x}) = \frac{1}{i\omega\rho}(K\varphi)^+(\mathbf{x}), \quad \mathbf{x} \in \Gamma, \tag{13}$$

where K is the normal derivative of the integral operator $(S\varphi)(\mathbf{x})$ in Eq. (3), when $\mathbf{x} \in \Gamma$. The formula of this normal derivative is given in Ref. 33, Theorem 2.19,

$$(K\varphi)^+(\mathbf{x}) := \int_\Gamma \frac{\partial \Phi(\mathbf{x},\mathbf{y})}{\partial n(\mathbf{x})} \varphi(\mathbf{y}) dS(\mathbf{y}) - \Omega(\mathbf{x})\varphi(\mathbf{x}). \tag{14}$$

The pressure over Γ will be obtained using the density φ on the integral Eq. (8) evaluated at $\mathbf{x} \in \Gamma$.

III. NUMERICAL IMPLEMENTATION

As exposed in the previous sections the NAH technique will be reduced to the solution of integral equations. The numerical solution of an integral equation is based on its discretization, which is a reduction into a linear matrix system where numerical methods can be applied. Boundary element methods (BEMs) will be used as the discretization method in this work since they have been successfully used in the area of acoustic radiation and scattering (see Refs. 5, 17, 11, 38, 39, and 28) for three-dimensional surfaces.

A. BEM discretization of integral equations

The boundary elements used for approximating the surface integral are schematically shown in Fig. 1. N_v triangular elements are used in this study for the construction of meshes. The global Cartesian coordinates $\mathbf{y}$ on any point of an element (denoted as Δ_j, $j=1,...,N_v$) are assumed to be related to the nodes $\mathbf{y}_m$, $m=1,...,q$, by

$$\mathbf{y}(\boldsymbol{\xi}) = \sum_{m=1}^{q} N_m(\boldsymbol{\xi})\mathbf{y}_m, \tag{15}$$

in which $\boldsymbol{\xi}=(\xi_1,\xi_2)$, $q=3$ and $N_m(\boldsymbol{\xi})$ are linear shape functions defined as

$$\begin{aligned} &N_1(\xi_1,\xi_2) = u, \quad N_2(\xi_1,\xi_2) = \xi_2, \\ &N_3(\xi_1,\xi_2) = \xi_1, \quad u = 1-\xi_1-\xi_2, \end{aligned} \tag{16}$$

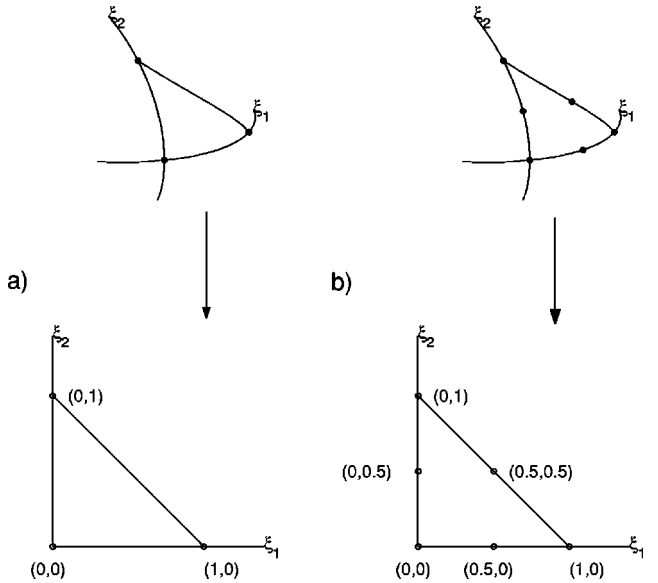


FIG. 1. Boundary elements used in iso-parametric transformation. (a) Linear elements. (b) Quadratic elements. The lower part of the figure shows the unit triangles.

or quadratic shape functions with $q=6$

$$\begin{aligned} &N_1(\xi_1,\xi_2) = u(2u-1), \quad N_2(\xi_1,\xi_2) = \xi_2(2\xi_2-1), \\ &N_3(\xi_1,\xi_2) = \xi_1(2\xi_1-1), \quad N_4(\xi_1,\xi_2) = 4\xi_2 u, \\ &N_5(\xi_1,\xi_2) = 4\xi_1\xi_2, \quad N_6(\xi_1,\xi_2) = 4\xi_1 u. \end{aligned} \tag{17}$$

Equation (15) is an isoparametric transformation in which a surface element is mapped into a plane equilateral unit triangle (as in the lower part of Fig. 1). Next, the boundary variable φ given in Eqs. (3), (4), (6), and (14) will be represented on each element j according to

$$\varphi(\boldsymbol{\xi}) = \sum_{m=1}^{q} N_m(\boldsymbol{\xi})\varphi_{(m,j)}, \tag{18}$$

where $\varphi_{(m,j)}$ are the values of φ at node m of the element j. For $\mathbf{x}_i \in D^+$, we have that for the integral in Eqs. (3) and (4)

$$(S\varphi)(\mathbf{x}_i) \approx \sum_{j=1}^{N_v}\sum_{m=1}^{q} S^i_{(m,j)}\varphi_{(m,j)}, \tag{19}$$

$$(D\varphi)(\mathbf{x}_i) \approx \sum_{j=1}^{N_v}\sum_{m=1}^{q} D^i_{(m,j)}\varphi_{(m,j)}, \tag{20}$$

where

$$\begin{aligned} S^i_{(m,j)} = &\int_0^1 \int_0^{1-\xi_2} \Phi(\mathbf{x}_i, \mathbf{y}(\xi_1,\xi_2)) N_m(\xi_1,\xi_2) \\ &\times J(\xi_1,\xi_2) d\xi_1 d\xi_2, \end{aligned} \tag{21}$$

$$\begin{aligned} D^i_{(m,j)} = &\int_0^1 \int_0^{1-\xi_2} \frac{\partial \Phi(\mathbf{x}_i, \mathbf{y}(\xi_1,\xi_2))}{\partial \mathbf{n}(\mathbf{y}(\xi_1,\xi_2))} N_m(\xi_1,\xi_2) \\ &\times J(\xi_1,\xi_2) d\xi_1 d\xi_2, \end{aligned} \tag{22}$$

in which $J(\xi_1,\xi_2)$ is the Jacobian of the coordinate transformation. This Jacobian is known exactly for some surfaces, however for general surfaces it is more convenient to ap-

proximate it numerically using Eq. (15) as suggested in Ref. 40. Integrals $S^i_{(m,j)}$, $D^i_{(m,j)}$ are approximated numerically using the adaptive integration scheme given in Ref. 41, Chap. 9, with the formula (T_2:5.1) from Ref. 42, Chap. 8.

For $\mathbf{x}_i \in \Gamma$ we have that the integral on Eq. (3) will be approximated as in Eq. (19), and the integrals on Eqs. (6) and (14) will be

$$(D\varphi)^+(\mathbf{x}_i) \approx \sum_{j=1}^{N_v} \sum_{m=1}^{q} D^i_{(m,j)} \varphi_{(m,j)} + C(\mathbf{x}_i)\varphi(\mathbf{x}_i), \quad (23)$$

$$(K\varphi)^+(\mathbf{x}_i) \approx \sum_{j=1}^{N_v} \sum_{m=1}^{q} K^i_{(m,j)} \varphi_{(m,j)} - C(\mathbf{x}_i)\varphi(\mathbf{x}_i), \quad (24)$$

where

$$K^i_{(m,j)} = \int_0^1 \int_0^{1-\xi_2} \frac{\partial \Phi(\mathbf{x}_i, \mathbf{y}(\xi_1,\xi_2))}{\partial \mathbf{n}(\mathbf{x}_i)} N_m(\xi_1,\xi_2) \times J(\xi_1,\xi_2) d\xi_1\, d\xi_2. \quad (25)$$

Integrals $S^i_{(m,j)}$, $D^i_{(m,j)}$, and $K^i_{(m,j)}$ will be approximated using the adaptive integration scheme used for the previous case when $\mathbf{x}_i \notin \Delta_j$. When $\mathbf{x}_i \in \Delta_j$, then $S^i_{(m,j)}$, $D^i_{(m,j)}$ and $K^i_{(m,j)}$ are approximated numerically using a Duffy transformation and Gaussian quadrature as suggested in Ref. 43.

In practice for the NAH technique we will take M pressure measurements on Γ_0 and will want to recover N pressure or normal velocity points on Γ. In general we will find that $M \geqslant N$, but there is no major theoretical contradiction if $M < N$. We will denote as $\mathbf{p}$ the column vector with M pressure measurements on Γ_0. The N points of the recovered normal velocity and pressure on Γ will be given respectively by the column vectors $\mathbf{v}^s$, $\mathbf{p}^s$. For $\mathbf{x}_i \in D^+$, the sums in Eqs. (19) and (20) can be reduced for the N points on Γ to produce the $M \times N$ complex matrices $[\mathbf{S}]$ and $[\mathbf{D}]$. Similarly, for $\mathbf{x}_i \in \Gamma$ we will obtain the $N \times N$ complex matrices $[\mathbf{S}^s]$, $[\mathbf{D}^+]$, $[\mathbf{K}^+]$ by a reduction of the sums in Eqs. (19), (23), and (24), respectively.

The direct formulation [in Eq. (2)] combined with Eq. (6) will give the explicit expression of the Green's transfer functions for the DE approach,

$$[\mathbf{G}_d]\mathbf{p}^s = \mathbf{p}, \quad (26)$$

$$[\mathbf{G}_n]\mathbf{v}^s = \mathbf{p}, \quad (27)$$

where $[\mathbf{G}_d]$, $[\mathbf{G}_n]$ are $M \times N$ complex matrices that respectively are discretizations of the integral operators in Eqs. (10) and (11). These matrices have the explicit representation

$$[\mathbf{G}_d] = [\mathbf{D}] - [\mathbf{S}][\mathbf{S}^s]^{-1}([\mathbf{D}^+] - [\mathbf{I}]), \quad (28)$$

$$[\mathbf{G}_n] = i\omega\rho([\mathbf{D}]([\mathbf{D}^+] - [\mathbf{I}])^{-1}[\mathbf{S}^s] - [\mathbf{S}]), \quad (29)$$

where $[\mathbf{I}]$ is the $N \times N$ identity matrix. Notice that the creation of the matrices $[\mathbf{G}_d]$, $[\mathbf{G}_n]$ will be constrained to the existence of $[\mathbf{S}^s]^{-1}$ and $([\mathbf{D}^+] - [\mathbf{I}])^{-1}$. This problem will be addressed in Sec. IV. In the following paragraphs we present implicit approaches that do not rely on the formation of the matrices $[\mathbf{G}_d]$ and $[\mathbf{G}_n]$.

The Helmholtz–Kirchhoff system will be created using the direct formulation in Eq. (2) combined with Eq. (6) to produce

$$\begin{pmatrix} -i\omega\rho[\mathbf{S}] & [\mathbf{D}] \\ -i\omega\rho[\mathbf{S}^s] & ([\mathbf{D}^+] - [\mathbf{I}]) \end{pmatrix} \begin{pmatrix} \mathbf{v}^s \\ \mathbf{p}^s \end{pmatrix} = \begin{pmatrix} \mathbf{p} \\ 0 \end{pmatrix}. \quad (30)$$

This system will be solved for $\mathbf{v}^s$ and $\mathbf{p}^s$, and is used for the DM approach.

For the IM approach, the indirect formulation in Eq. (12) is used to solve

$$[\mathbf{S}]\boldsymbol{\varphi} = \mathbf{p} \quad (31)$$

for the column vector of N entries $\boldsymbol{\varphi}$. Then the normal velocity is obtained by Eq. (13) and the discretization of Eq. (14)

$$\mathbf{v}^s = \frac{1}{i\omega\rho}[\mathbf{K}^+]\boldsymbol{\varphi}. \quad (32)$$

Similarly the pressure $\mathbf{p}^s$ will be obtained using

$$\mathbf{p}^s = [\mathbf{S}^s]\boldsymbol{\varphi}. \quad (33)$$

B. Numerical regularization

For the experimental problem, the pressure is a measured quantity that will contain errors and we will denote it as $\tilde{\mathbf{p}}$. If the elements of the perturbation $\mathbf{e} = \mathbf{p} - \tilde{\mathbf{p}}$ are unbiased and uncorrelated with covariance matrix $\sigma_0^2[\mathbf{I}]$, then $E(\|\mathbf{e}\|_2^2) = M\sigma_0^2$, where $\|\cdot\|_2$ is the two-norm. It is well known that the linear systems in Eqs. (26), (27), (30), and (31) are ill-posed, i.e., the presence of errors in $\tilde{\mathbf{p}}$ will be amplified in the solutions $\mathbf{v}^s$, $\mathbf{p}^s$ or $\mathbf{h}$, and in most of the cases the recovery will be useless. For that reason, special methods for the solution of these linear systems are required.

Instead of considering these special methods of numerical solution for each ill-posed matrix system in Eqs. (28)–(31), we consider the solution to the generic ill-posed linear matrix system

$$[\mathbf{A}]\mathbf{z} = \tilde{\mathbf{b}}. \quad (34)$$

Here $[\mathbf{A}]$, $\mathbf{z}$ and $\tilde{\mathbf{b}}$ will respectively represent the ill-posed matrix, the solution to the linear system and the measurement vector of Eqs. (28)–(31). Let $\bar{M} \times \bar{N}$ be the dimensions of the matrix $[\mathbf{A}]$. Then $\mathbf{z}$, $\tilde{\mathbf{b}}$ are column vectors of $\bar{N}$, $\bar{M}$ entries respectively. For Eqs. (28), (29), and (31), $\bar{M} = M$ and $\bar{N} = N$. For Eq. (30), $\bar{M} = N + M$ and $\bar{N} = 2N$.

The ill-posedness of the linear system on Eq. (34) can be explained using the singular value decomposition (SVD). The SVD of $[\mathbf{A}]$ will give a decomposition of the form

$$[\mathbf{A}] = [\mathbf{U}][\Sigma][\mathbf{V}]^H, \quad (35)$$

where $[\mathbf{U}]$ and $[\mathbf{V}]$ are unitary matrices of dimensions $\bar{M} \times \bar{M}$ and $\bar{N} \times \bar{N}$, respectively, and $[\mathbf{V}]^H$ is the conjugate transpose of $[\mathbf{V}]$. $[\Sigma]$ is a diagonal matrix with values $\sigma_1 \geqslant \cdots \geqslant \sigma_{N^*} \geqslant 0$ with $N^* \equiv \mathrm{rank}([\mathbf{A}])$. The values σ_i are called *singular values*.

For the acoustic problem $[\mathbf{V}]_i$, $[\mathbf{U}]_i$ (the ith columns of $[\mathbf{V}]$, $[\mathbf{U}]$), $i = 1,...,N$, will be an approximation to the acoustic field by basic acoustic waves or *mode shapes* (see Ref. 28). In particular when $[\mathbf{A}] = [\mathbf{G}_n]$ [as in Eq. (27)], $[\mathbf{U}]_i$ will be the modes of the measured pressure on Γ_0 and $[\mathbf{V}]_i$ will be the modes of the normal velocity on Γ. These modes are

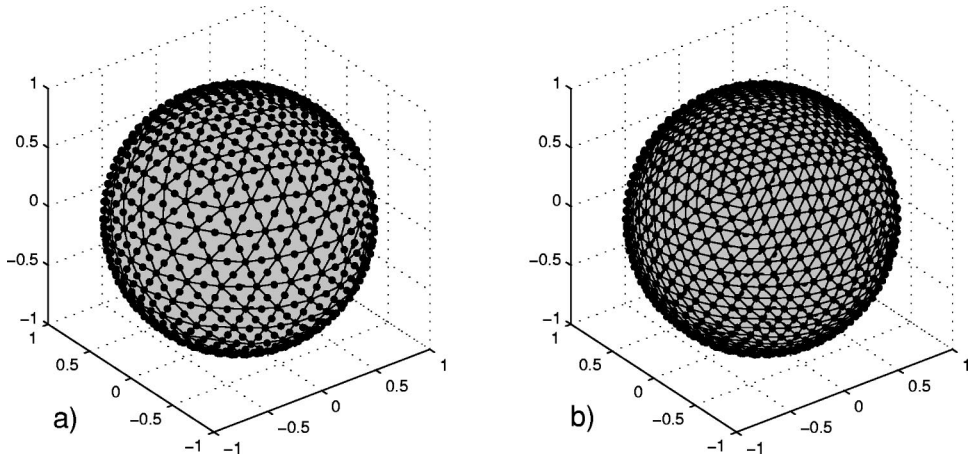


FIG. 2. Sphere of radius 1, center at the origin and 1026 points. The surface is decomposed onto (a) 512 quadratic elements and (b) 2048 linear elements.

organized in such a way that the first modes are related with the nonevanescent waves and the last modes will be related with the evanescent waves. The numerical solution of Eq. (34) by conventional methods $\mathbf{z}_{LS}$ can be represented using the SVD as

$$\mathbf{z}_{LS}=[\mathbf{V}]\mathrm{diag}\left(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_{N*}}\right)[\mathbf{U}]^H\tilde{\mathbf{b}}. \tag{36}$$

Notice that this solution will include modes related to both nonevanescent and evanescent waves. It is well known that the modes related to evanescent waves will produce the fine details on the reconstruction, but at the same time will amplify the noise. Special methods for the solution to this inverse acoustic problem will need to include enough of these modes in order to obtain a certain resolution, and at the same time exclude some of these modes that are contaminated with noise.

The special methods for the solution to Eq. (34) are called *regularization methods* and can be implemented using the SVD as

$$\tilde{\mathbf{z}}_\alpha=[\mathbf{V}][F^\alpha]\mathrm{diag}\left(\frac{1}{\sigma_1},\ldots,\frac{1}{\sigma_{N*}}\right)[\mathbf{U}]^H\tilde{\mathbf{b}}, \tag{37}$$

where α is known as the regularization parameter and $[F^\alpha]$ is a matrix with diagonal entries called the *filter factor* (see Refs. 9 and 10). The parameter α will control this inclusion of modes related to evanescent waves into the solution $\tilde{\mathbf{z}}_\alpha$. The best known regularization method is Tikhonov and will be implemented using the filter factor

$$[F^\alpha]=\mathrm{diag}\left(\frac{\sigma_1^2}{\sigma_1^2+\alpha},\ldots,\frac{\sigma_{N*}^2}{\sigma_{N*}^2+\alpha}\right). \tag{38}$$

The topic of regularization methods for linear matrix systems has been extensively studied in the last few decades. There are many other approaches depending on the particular problem, where we can mention the use of Tikhonov regularization with a high-pass filter (see Ref. 15) which uses the physical behavior of the evanescent waves to produce an optimal filter. It is not the goal of this work to give a detailed study of regularization methods for this acoustical problem, but instead to provide a fair comparison between the reconstruction of the acoustic field given by the implicit approaches (DM and IM) and the reconstruction given by the traditional DE approach. For that reason, all of our numerical experiments in Secs. IV and V will use Tikhonov regularization (which is the oldest and best known regularization method) with the Morozov discrepancy principle (as used in Ref. 15) to obtain the optimal parameter α.

IV. PROBLEMS WITH THE NUMERICAL SOLUTION

As mentioned previously, the numerical solution to an ill-posed matrix system will require the use of regularization methods, which is itself a subject of current research, but there are other problems that will need our particular attention: the effect of the higher order shape functions and problematic frequencies.

For the numerical examples in this section, we will consider the exterior problem of reconstructing $\mathbf{v}^s$ on the surface Γ, a sphere of radius 1 with 1026 points as shown in Fig. 2, using pressure measurements $\mathbf{p}$ over Γ_0, a sphere of radius 1.01 with 1026 points distributed as in Γ. The vector $\tilde{\mathbf{p}}$ is created by adding to $\mathbf{p}$ the vector $\mathbf{e}$ with normal random distributed entries and 35 dB SNR.

The figures that will be shown in this section over the sphere Γ depend on the spherical coordinates (r,ϕ,θ), and hence will be transformed using interpolation into the plane (ϕ,θ) for better visualization purposes.

A. Effect of shape functions

The surface Γ can be decomposed using 2048 linear triangular elements as shown in Fig. 2(a) and 512 quadratic triangular elements as shown in Fig. 2(b). It is a common practice for the forward problem to use the quadratic triangular decomposition over Γ [and so quadratic shape functions on Eq. (18)] in order to improve the accuracy of the solution to the acoustical problem. This will not be a good idea for the inverse problem, where increasing the resolution will pay the penalty of reducing the smoothness of the solution, as we will show.

Let the quantities W_m^l, $m=1,\ldots,3$, be defined as

$$W_m^l=\int_0^1\int_0^{1-\xi_2}N_m(\xi_1,\xi_2)J(\xi_1,\xi_2)d\xi_1\,d\xi_2, \tag{39}$$

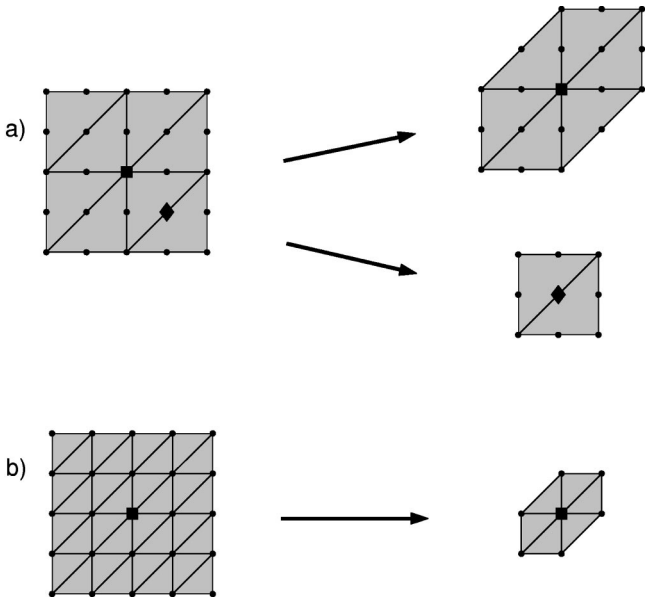


FIG. 3. Decomposition of a plane using triangular (a) quadratic elements and (b) linear elements. The ''square'' point is the vertex of a triangle and intersects six contiguous triangles for both (a) and (b). The ''diamond'' point in (a) is the midpoint of a triangle for quadratic elements and intersects two contiguous triangles.

for the linear shape functions N_m, given in Eq. (16). Similarly define the quantities W_m^q, $m=1,...,6$, as

$$W_m^q = \int_0^1 \int_0^{1-\xi_2} N_m(\xi_1,\xi_2) J(\xi_1,\xi_2) d\xi_1\, d\xi_2, \tag{40}$$

for the quadratic shape functions N_m, given in Eq. (17). Equations (39) and (40) can be thought of as approximations to $S^i_{(m,j)}$, $D^i_{(m,j)}$, and $K^i_{(m,j)}$ when the field point $\mathbf{x}_i$ is far from the element Δ_j. For our further exposition W_m^l and W_m^q, instead of being used as an approximation, will be used to estimate the Column normalized weights (CNWs) for the matrices $[\mathbf{S}]$, $[\mathbf{D}]$, $[\mathbf{S}^s]$, $[\mathbf{D}^+]$, and $[\mathbf{K}^+]$.

The following explanation of the CNW will be given for the matrix $[\mathbf{S}]$, but will also apply to $[\mathbf{D}]$, $[\mathbf{S}^s]$, $[\mathbf{D}^+]$, and $[\mathbf{K}^+]$. Notice that $[\mathbf{S}]$ is obtained by a reduction of the sum of $S^i_{(m,j)}$ in Eq. (19) and this reduction depends on the triangular elements decomposition used for the surface Γ. Figure 3 shows the decomposition of a plane using linear and quadratic triangular elements. Observe in Fig. 3(a) that the ''square'' point (point marked with a square) intersects six triangles. This means that when the sums in Eq. (19) are reduced, then the column of the resultant matrix $[\mathbf{S}]$ that corresponds to the ''square'' point will be the sum of the quantities $S^i_{(m,j)}$ for j's and m's corresponding to the six triangles intersecting the ''square'' point. For that reason, a column of the matrix $[\mathbf{S}]$ that corresponds to the ''square'' point will be associated with the CNW $6W_m^q$. Using the same argument for the ''diamond'' point in Fig. 3(a), the column of $[\mathbf{S}]$ that corresponds to the ''diamond'' point will be associated with the CNW $2W_m^q$. The ''square'' and ''diamond'' points in the quadratic triangle decomposition of a plane [see Fig. 3(a)] represent, respectively, a vertex and a midpoint of a triangle that belong to the triangular decomposition of Γ. For that reason, when quadratic triangular elements are used, the columns of matrices $[\mathbf{S}]$ that correspond to a vertex point ($m=1,2,3$) will be associated with the CNW $6W_m^q$, and the columns corresponding to a midpoint ($m=4,5,6$) will be associated with the CNW $2W_m^q$. A similar reasoning will imply that, for linear triangular elements, all the columns of $[\mathbf{S}]$ are associated with the CNW $6W_m^l$ [see Fig. 3(b)].

Notice that this unbalance of the CNW will be found in the entries of the matrix $[\mathbf{S}]$ that are outside the diagonal (since the field points $\mathbf{x}_i$ in this region are relatively far from the element j). The worst case is when the surface is planar, since the Jacobian on Eq. (40) is constant, then for $m=1$, 2, 3 we have that $W_m^q=0$. This will imply that $S^i_{(m,j)}\approx 0$ when $m=1$, 2, 3, and the matrix $[\mathbf{S}]$ will have entries outside the diagonal that are close to 0 for columns that correspond to vertex points of quadratic elements. We compute the matrix $[\mathbf{S}]$ for the spheres Γ and Γ_0 (explained in the beginning of this section), and compute the SVD. As shown in Fig. 4, the unbalance on the columns of $[\mathbf{S}]$ will produce ''zero'' dots on the plots of the columns of matrix $[\mathbf{V}]$. We can see in Fig. 5 that this effect is not encountered for linear elements where the CNW is balanced. A similar effect will be expected to happen to the SVD of $[\mathbf{G}_n]$.

The DE approach will be used to reconstruct the normal velocity $\mathbf{v}^s$ for 50 Hz given by a point source located at $(0,-0.8,0)$. The reconstruction using quadratic elements and linear elements is shown in Fig. 6. The relative error for the reconstructed normal velocity $\mathbf{v}_r^s$ (given by $\|\mathbf{v}^s-\mathbf{v}_r^s\|_2/\|\mathbf{v}^s\|_2 \times 100\%$) for the linear elements is 8% while the relative error for quadratic elements is 28%. In both cases the regularization parameter was chosen to be the optimal (since the exact solution is known). Notice that the increase in error for the quadratic elements results from the use of ''zero'' dot modes for the SVD of $[\mathbf{G}_n]$ (as in the SVD of $[\mathbf{S}]$ in Fig. 4). For the forward problem, where all the modes are used for the inversion, the use of these quadratic modes will have no effect in the reconstructed solution (as seen in the literature for the forward problem like Ref. 39). In the inverse problem, where just a partial set of these modes is required, then the solution will be affected (as shown by our numerical experiment).

B. Problematic frequencies

There are certain frequencies where the formulations discussed before will fail to provide a unique solution for the forward problem. This uniqueness is a purely mathematical problem arising from the boundary integral formulation rather than from the nature of the physical problem. There are several modified integral formulations to overcome the nonuniqueness problem, but in practice the amount of computational complexity of these formulations will be a high penalty to pay.

In this section we will clarify a common misconception from the available acoustical literature in regards to the solution of the exterior problem by the indirect formulation in Eq. (8). For the exterior problem, when the wave number k coincides with a Dirichlet eigenvalue, it is known theoretically that the solution to the integral equation in Eq. (13) is

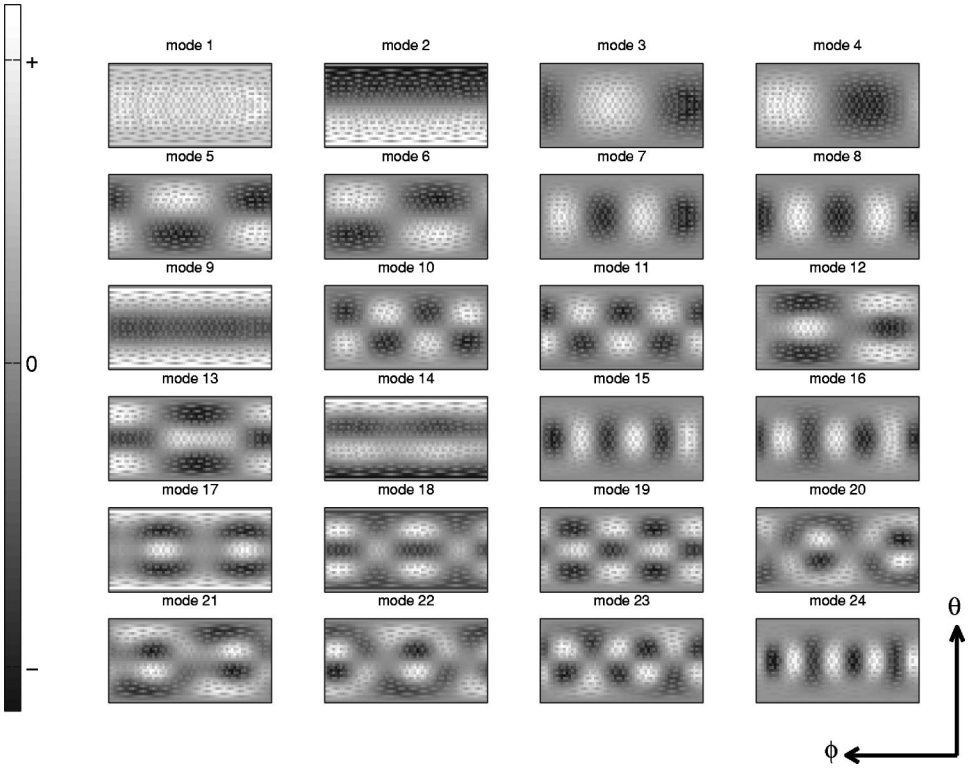


FIG. 4. Plot of the first 24 columns of the matrix [**V**] that is obtained from the SVD of [**S**]. The matrix [**S**] is calculated for the spherical surfaces Γ_0 and Γ using quadratic elements. Notice the existence of "zero" dots over the plot of these columns. The frequency is 50 Hz.

nonunique (see Ref. 44). From this fact, then, it is concluded by the Fredholm–alternative theorem that there is no solution to the integral equation in Eq. (13). This implication is correct, but does not mean that the indirect representation in Eq. (8) "breaks down" for this wave number. Newer results have been published, and in particular we mention Ref. 24, Theorem 2.1, where it is proven that there exists a density φ such that the indirect formulation on Eq. (8) holds, although this density is nonunique. This nonuniqueness is an issue that can be worked around for inverse problems, where we can always obtain a unique solution with the minimal norm (see Ref. 9, Chap. 2). Then a solution using the indirect formulation in Eq. (8) for these problematic frequencies is obtained in a minimal norm sense. We will show that the minimal norm solution will reduce the accuracy of the reconstruction, which will be a serious problem for the forward problem, but

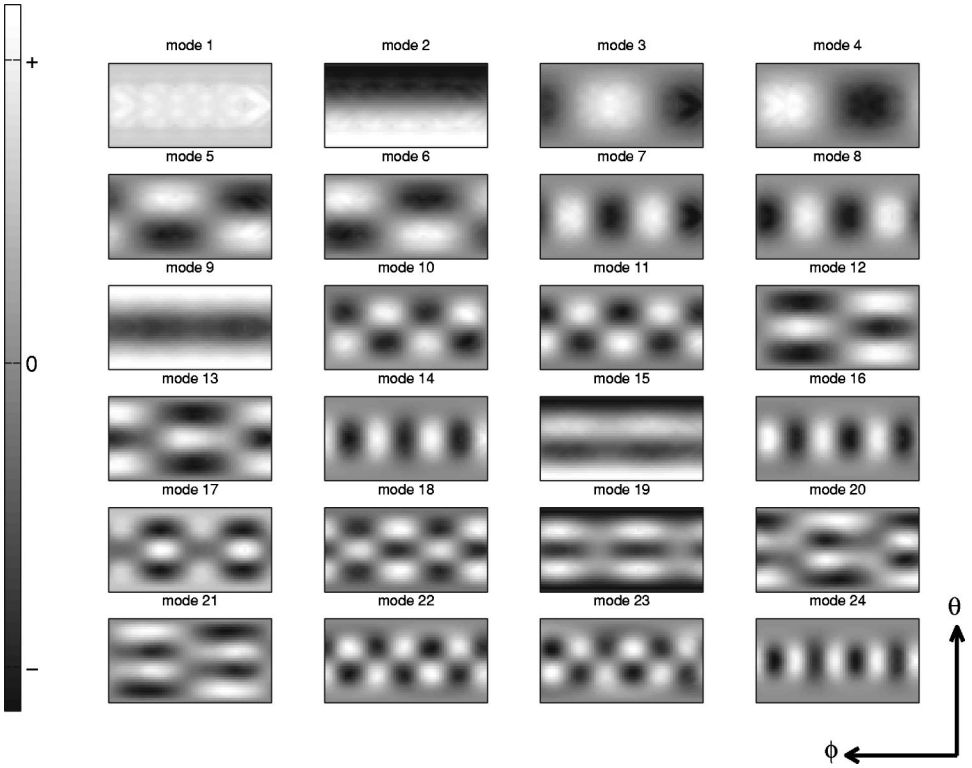


FIG. 5. Plot of the first 24 columns of the matrix [**V**] that is obtained from the SVD of [**S**]. The matrix [**S**] is calculated for the spherical surfaces Γ_0 and Γ using linear elements. The frequency is 50 Hz.

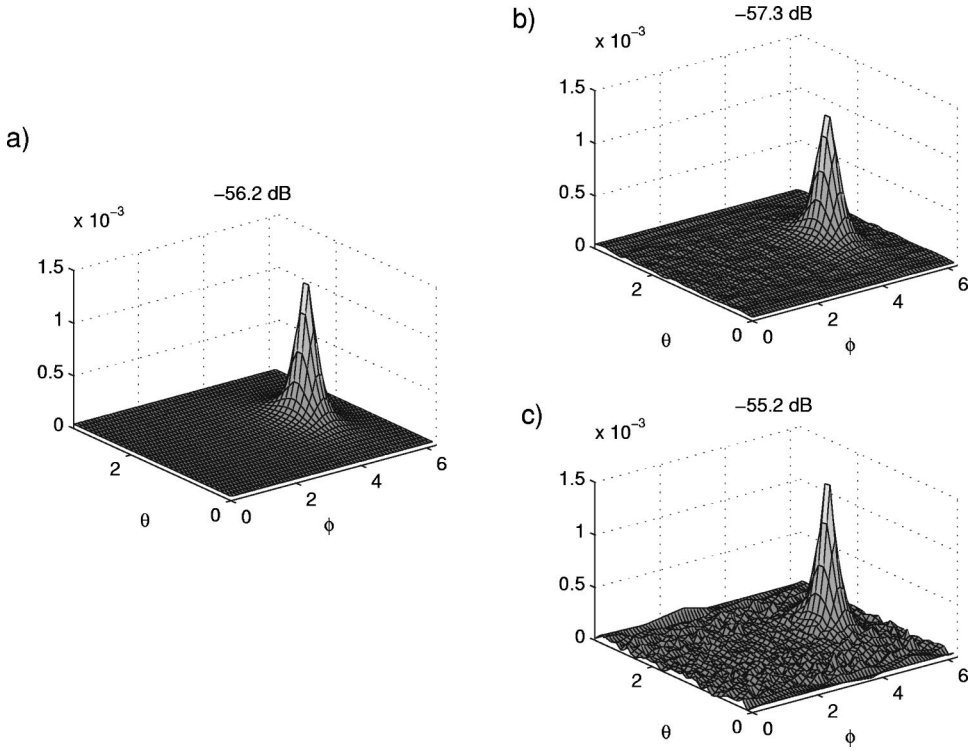


FIG. 6. (a) Exact normal velocity, (b) recovered normal velocity using linear elements, and (c) recovered normal velocity using quadratic elements. The frequency is 50 Hz.

for our inverse problem will encounter acceptable errors.

For our analysis we will use the condition number of a square matrix $cond([\mathbf{A}])$, which is defined as the ratio between the largest and smallest singular values, i.e, σ_1/σ_{N*}. This quantity will give a fair idea of the problems associated with the inversion of a matrix. In Fig. 7(a) we plot the condition number of matrices $[\mathbf{S}]$, $[\mathbf{S}^s]$, and $[\mathbf{K}^+]$ (calculated using linear elements) over different wave numbers. The horizontal axis is the wave number and the vertical axis is the condition number of the respective matrices. Notice that the spikes in the plot correspond to the zeros of the spherical Bessel functions (see Ref. 45, Table 10.6), which correspond to the Dirichlet eigenvalues for the sphere with radius 1. Numerically these spikes imply the existence of singular values close to ''numerical zero,'' which means that we cannot invert directly the matrices $[\mathbf{S}]$, $[\mathbf{S}^s]$ and $[\mathbf{K}^+]$. Similar plots will be obtained for matrices $[\mathbf{D}]$, $[\mathbf{D}^+]$. In particular in Fig. 7(b) we have a similar plot to Fig. 7(a) (with the same vertical scale) over the wave number k_1, which is the first zero of the spherical Bessel function j_1. k_1 is approximately

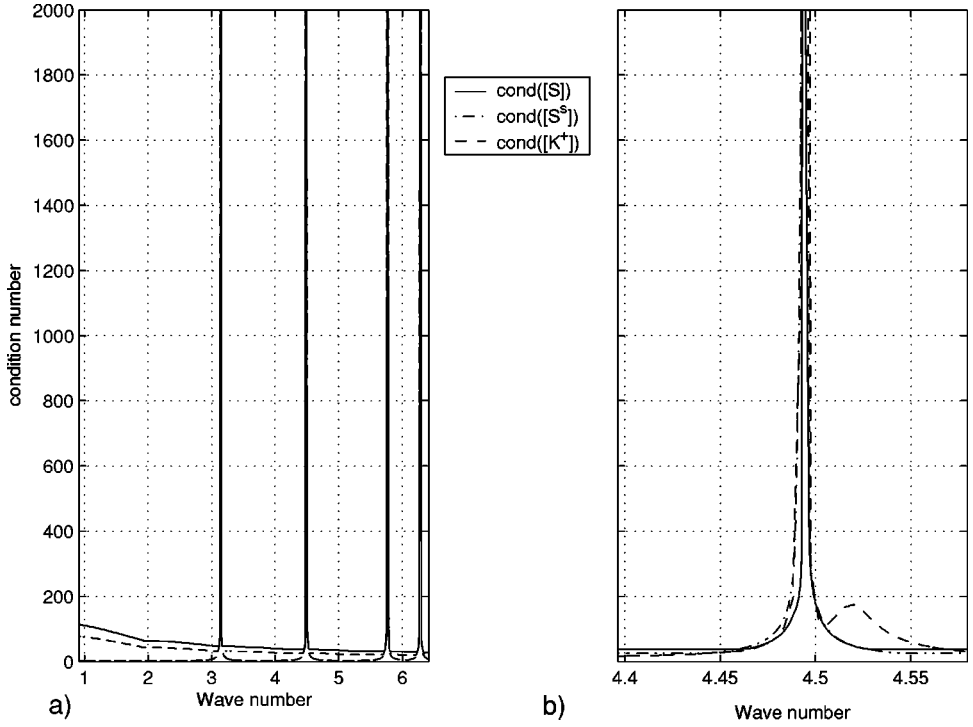


FIG. 7. Plot of condition number for matrices $[\mathbf{S}]$, $[\mathbf{S}^s]$, and $[\mathbf{K}^+]$. The vertical axis is the condition number and the horizontal axis the wave number. In (a) the plot is between wave numbers 1 and 6, and in (b) between wave numbers 4.4 and 4.6.

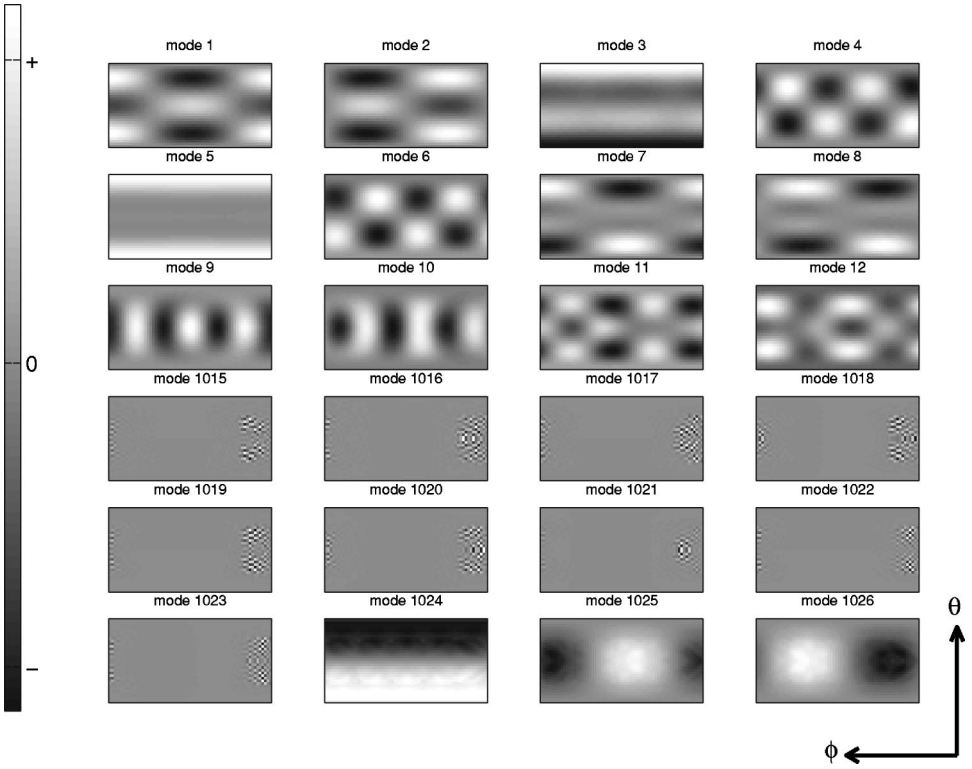


FIG. 8. Plot of columns of [**V**] for the matrix [**S**] calculated at wave number k_1 which corresponds to the first zero of the spherical Bessel function j_1. Notice that the last three columns contain the modes associated with j_1 that will not be used for the reconstruction.

4.492 409 4 and is the second spike from left to right in Fig. 7(a).

The existence of singular values close to ''numerical zero'' will represent a major problem for the explicit creation of the Green's matrices for the DE approach, since we cannot invert $[\mathbf{S}^s]$, $([\mathbf{D}^+]-[\mathbf{I}])$. It is possible to do a pseudo-inversion using the SVD (see Ref. 10), but this will be costly. The use of implicit methods as the IM approach or DM approach will not require the explicit inversion of matrices and for that reason can be applied as explained in Sec. III A. We will only compare the reconstructions from the IM approach with the exact solutions, as we expect similar results for the DM approach.

Figure 8 shows the plots of some columns of [**V**] for [**S**]. Notice that the modes corresponding to the zeros of j_1 [which are the spherical harmonics $Y^1_{-1}(\theta,\phi)$, $Y^1_0(\theta,\phi)$, and $Y^1_1(\theta,\phi)$] will be at the last columns of [**V**]. When

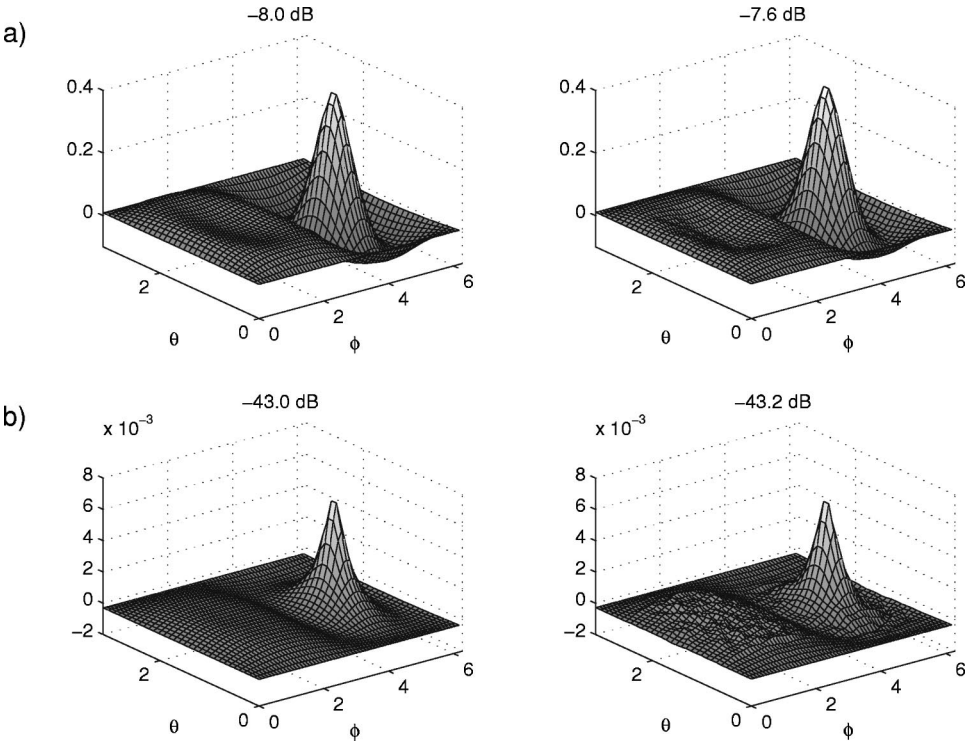


FIG. 9. At wave number k_1, (a) exact (left) and reconstructed (right) pressure and (b) exact (left) and reconstructed (right) normal velocity.

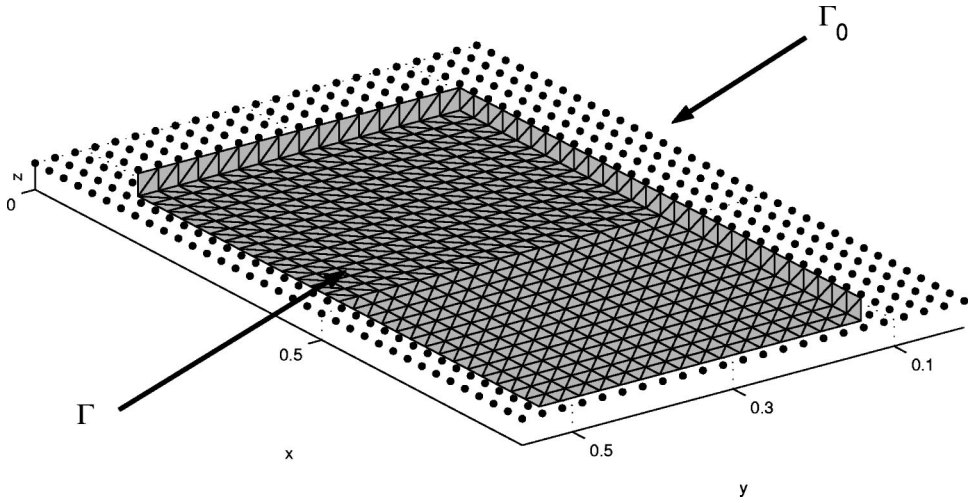


FIG. 10. The parallelepiped surface Γ and array Γ_0 used for the physical experiment. The surface Γ is decomposed into 1968 linear triangular elements and 986 points. The measurement surface Γ_0 have a total of 805 points (35 in the *x* direction and 23 in the *y* direction).

Tikhonov regularization is applied, this means that our solution will eliminate the Fourier coefficients pertaining to those modes. Using the same Γ and Γ_0 from the previous subsection, we show in Fig. 9 the reconstruction of the normal velocity at k_1 for a point source located at $(0,-0.8,0)$. The vector $\tilde{\mathbf{p}}$ is created by the addition of a vector $\mathbf{e}$ with normal random distributed entries and 35 dB SNR. As mentioned in the previous section, we use Tikhonov regularization, but since the exact solution is known, we choose the optimal regularization parameter. Although the relative error is approximately of 30% for the normal velocity and 20% for the pressure, the principal features of the exact velocity and pressure are recovered.

As shown in Fig. 7, there is a spacing between the wave numbers *k* that coincides with the Dirichlet eigenvalues of Γ (or problematic frequencies). This spacing will become smaller as we increase the frequency. In general it is well known[33] that there is always a separation between these problematic frequencies (even if the separation is smaller for higher frequencies). For acoustic problems, where we work up to the kHz regime, the probability of choosing a random positive number that will be a problematic frequency is very small. Moreover, we need to construct our linear system in Eqs. (26), (27), (30), or (31) for the exact problematic frequency (at least with several digits of accuracy) to encounter ''zero'' singular values in the SVD of $[\mathbf{S}]$, as was the case shown in Fig. 8. If that is the case, we have shown through our numerical example that our reconstruction will suffer a moderate loss of accuracy (which is always expected on the inverse problem) due to the loss of the modes corresponding to the ''numerical zero'' singular values. Notice that the proposed implicit methods, applied directly as explained in Sec. III A for these problematic frequencies, will take care of this complication without additional work. The DE approach will

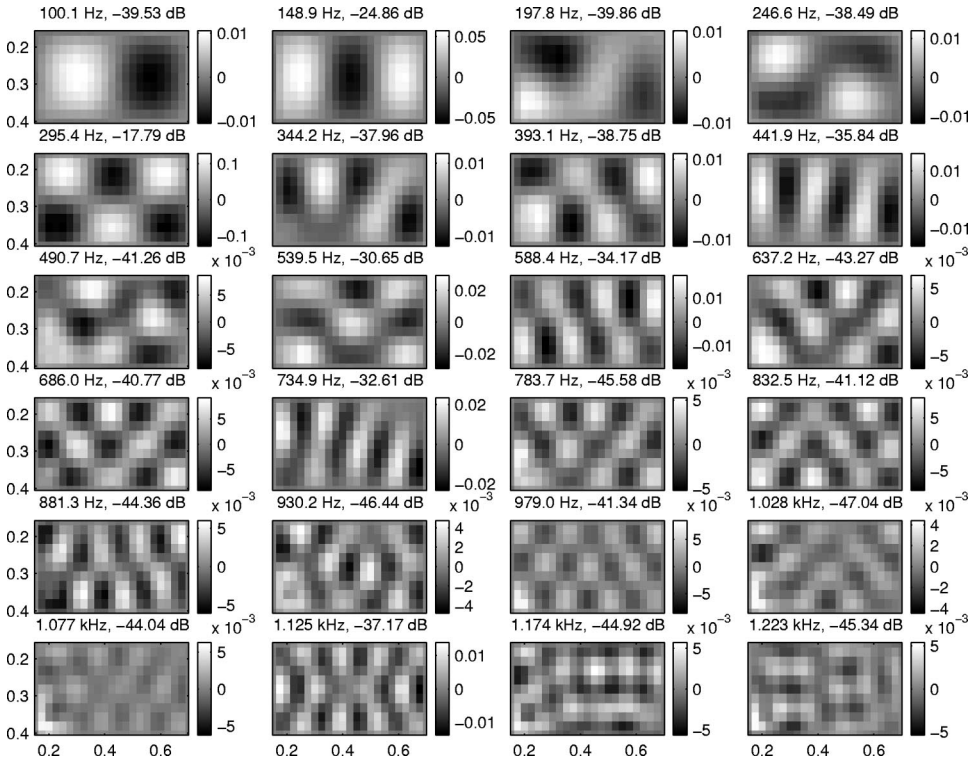


FIG. 11. Gray scale rendition of the real part of the normal velocity recovered using planar NAH for 24 equally distributed frequencies between 100 Hz and 1.22 kHz. At each frequency we show the maximum dB level over the plate.

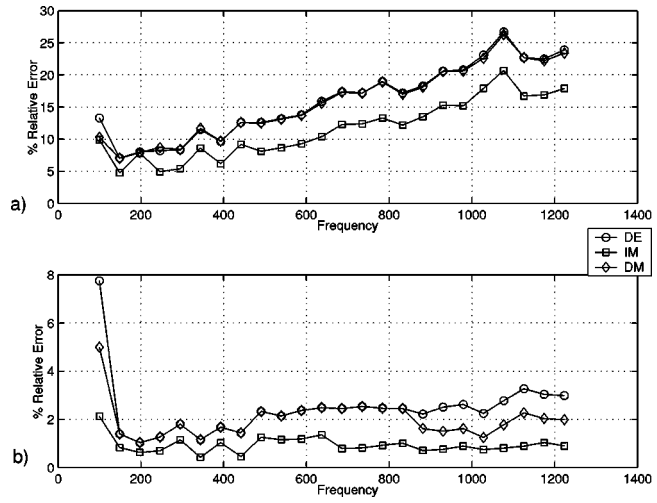


FIG. 12. Plot of % relative error for reconstructions of (a) normal velocity and (b) pressure. The vertical axis is the % error and the horizontal axis corresponds to the frequencies (Hz).

require special methods for the matrix inversion used to create the transfer functions $[\mathbf{G}_d]$, $[\mathbf{G}_n]$.

V. PHYSICAL EXPERIMENTS

The experimental configuration for the holographic measurement is similar to Ref. 46. A steel rectangular plate (of 55.9×26.7 cm and thickness of 1.6 mm) was driven by a Wilcoxon F3/F9 inertial shaker (0 to 3000 Hz) fixed in the underneath corner of the plate in order to generate a broad spectrum of spatial wave numbers. The resulting applied force was used to normalize the pressure measured with a ACO 7046 $\frac{1}{2}$-in. microphone above the plate. The scan is conducted in an automated point-by-point fashion. The microphone is initially positioned by the x-y scanner at a corner of the scan area, where the data is then acquired. Next, the microphone is moved to an adjacent point, and data is again acquired. This process is continued until the full scan area is measured. Synchronization between the source shaker, the data acquisition, and the scanner has proven to have high stability, yielding coherent data that can be thought of as having been acquired simultaneously. The pressure was measured on a grid of 69 points along the longer dimension and 46 points along the shorter dimension in a plane 0.4 cm above the plate surface. The step size in both the x and y directions was 1.25 cm, providing an overall scan dimension of 85×57.5 cm. The plate surface is located at the center of this area.

The points on the planar surface Γ_0 (35 points along the x direction and 23 points along the y direction) correspond to $\frac{1}{4}$ the original pressure measurements (and a spacing in both x and y directions of 2.5 cm). Γ will be a rectangular parallelepiped of 70×40 cm and thickness of 2.5 cm, located 0.4 cm below Γ_0. A uniform spacing of 2.5 cm is used for all points on Γ. The spacings over points on Γ_0 and Γ satisfy the minimal requirements to avoid aliasing when working with frequencies up to 2 kHz. Γ_0 and Γ are centered over the x-y coordinates at (42.5,28.125) cm (see Fig. 10). The BEM calculation are in MKS units.

In the results which follow, the ''exact'' normal velocity $\mathbf{v}^s$ and pressure $\mathbf{p}^s$ on the plate were determined by standard planar NAH reconstruction using the full 69×46 point array. It has been demonstrated that the reconstructions with the over-scanned hologram are extremely accurate.[1] The use of the BEM-NAH approaches discussed in Sec. III A will give the reconstructed normal velocity $\mathbf{v}_r^s$ or pressure $\mathbf{p}_r^s$ for 24 frequencies distributed equally from 100 Hz to 1.223 kHz over the parallelepiped Γ, but all the comparisons and calculation of errors will be taken over the x-y plane that corre-

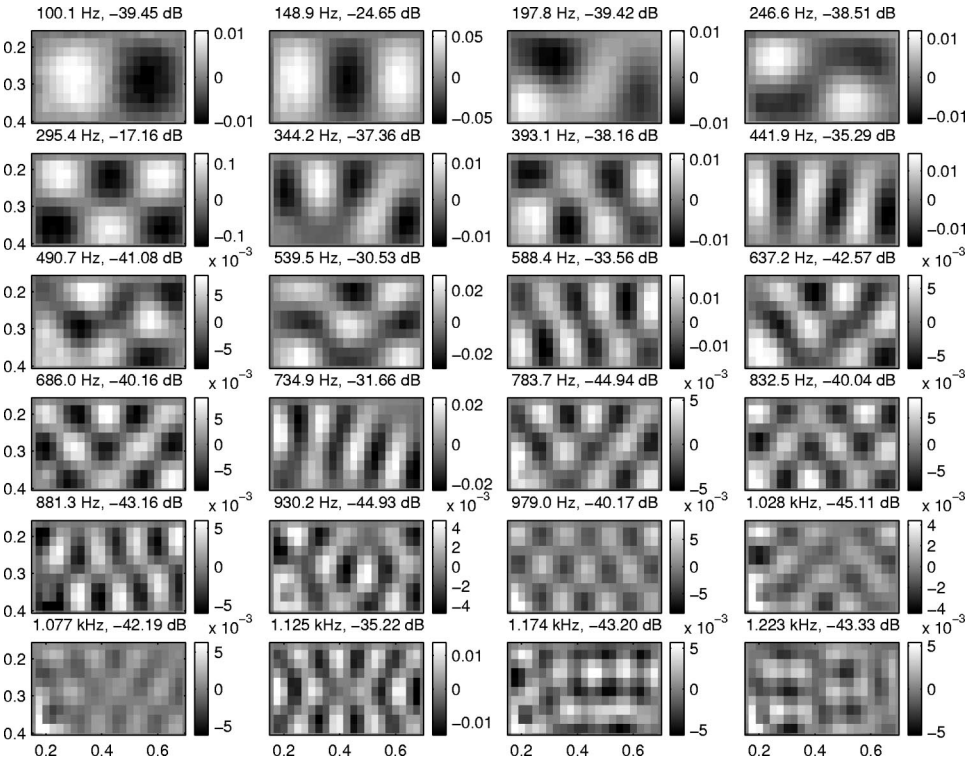


FIG. 13. Gray scale rendition of the real part of the normal velocity recovered using the direct-explicit (DE) approach for 24 equally distributed frequencies between 100 Hz and 1.22 kHz. At each frequency we show the maximum dB level over the plate.

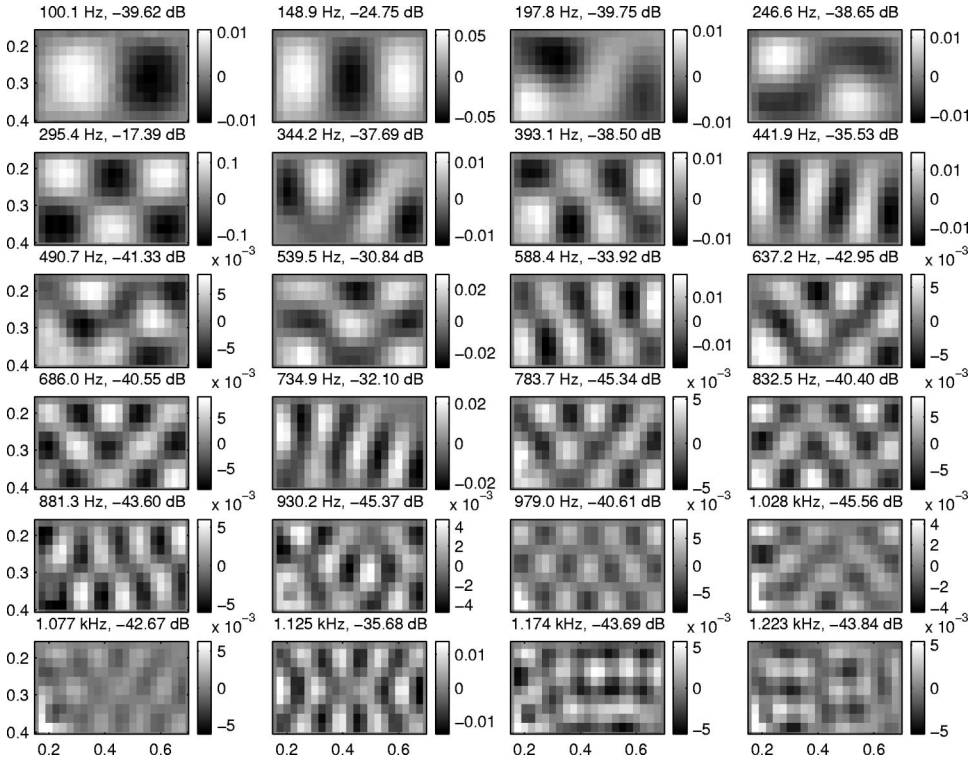


FIG. 14. Gray scale rendition of the real part of the normal velocity recovered using the indirect-implicit (IM) approach for 24 equally distributed frequencies between 100 Hz and 1.22 kHz. At each frequency we show the maximum dB level over the plate.

sponds to the plate [the upper x-y plane on Γ of dimensions 55.9×26.7 cm, center on (0.425,0.28125)m]. As mentioned earlier in Sec. III B, Tikhonov regularization using the Morozov discrepancy principle will be used to obtain the regularization parameter for all cases.

Figure 11 shows a gray-scale rendition on the real value of the "exact" normal velocity $\mathbf{v}^s$ (coming from planar NAH) for 24 frequencies that will be used to compute the relative error $\|\mathbf{v}^s-\mathbf{v}_r^s\|_2/\|\mathbf{v}^s\|_2\times100\%$ given in Fig. 12(a). The reconstructed normal velocity $\mathbf{v}_r^s$ will be computed using the DE, IM, and DM approaches, respectively. Similarly, we use the "exact" pressure $\mathbf{p}^s$ to compute the relative error $\|\mathbf{p}^s-\mathbf{p}_r^s\|/\|\mathbf{p}^s\|\times100\%$ on the 24 frequencies given in Fig. 12(b). An important property of the error on $\mathbf{v}_r^s$ reconstructions seen in Fig. 12(a) is that the DE approach and DM approach give similar errors. This is not surprising since both approaches use the direct formulation given in Eq. (2). The

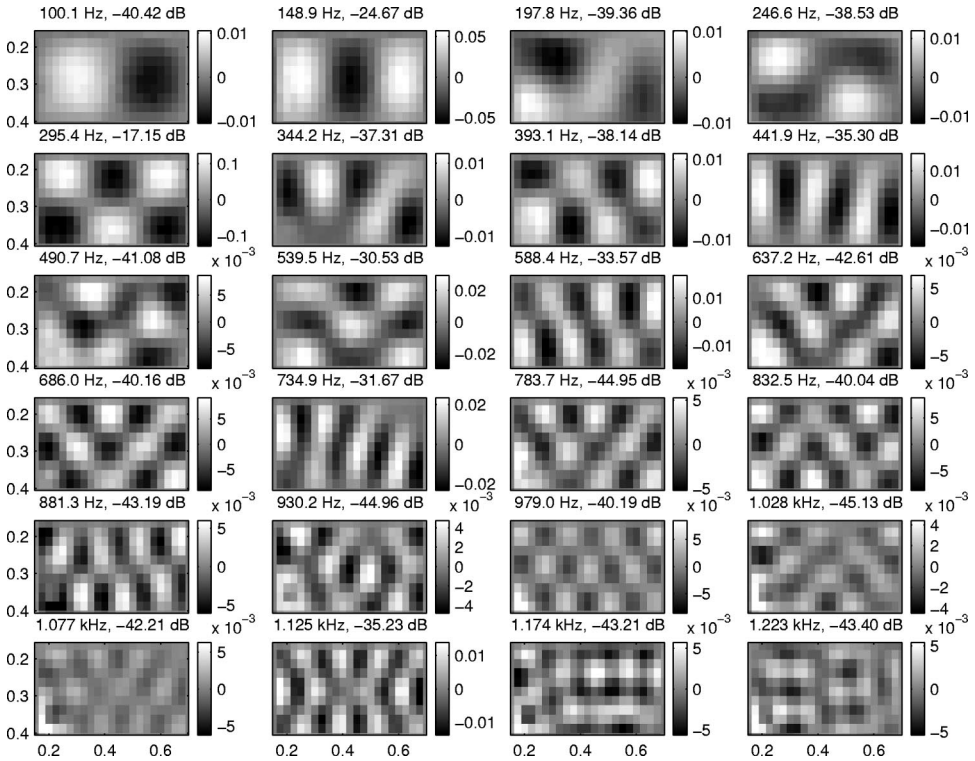


FIG. 15. Gray scale rendition of the real part of the normal velocity recovered using the direct-implicit (DM) approach for 24 equally distributed frequencies between 100 Hz and 1.22 kHz. At each frequency we show the maximum dB level over the plate.

reconstruction errors for the IM approach are slightly better for all frequencies (about 4% less). A similar behavior of the reconstruction error is found on the errors for $\mathbf{p}_r^s$ [see Fig. 12(b)]. In this case the relative error is much less than errors for the reconstruction of $\mathbf{v}_r^s$ for all frequencies (as expected).

Finally, Figs. 13–15 show a gray-scale rendition on the real value of the reconstructed normal velocity $\mathbf{v}_r^s$ coming from the DE, IM, and DM approaches, respectively. Each figure is given for 24 frequencies over the plate. Notice that little difference is seen between Figs. 13–15 and Fig. 11; all of the plots look similar.

VI. CONCLUSION

We have shown two implicit methods that are a good alternative for the reconstruction of the acoustic field using NAH. The numerical implementation of these methods is simple and generally requires less computational work than traditional methods.

In Sec. IV A we explained how the use of quadratic shape functions produced unbalanced CNW over the matrices created by BEM, which was shown to increase the reconstruction error of the acoustic field. This result applies to both direct and indirect formulations with implicit or explicit methods for the solution. Section IV B showed that implicit methods applied to the reconstruction of the acoustic field over problematic frequencies, without additional work, can still produce reconstructions with some moderated loss of accuracy. Methods that use explicit representations of Eq. (26) or (27) will require special techniques for the solution, which can be computationally expensive.

In Sec. V our experiment with the plate data showed that the use of implicit methods will produce similar errors as the traditional DE approach, for the reconstruction of the acoustic field. In particular for the IM approach, there is a small decrease in reconstruction errors. The reduction of the reconstruction error was not our principal motivation for the use of implicit methods, but instead we wanted to demonstrate the fact that these methods can be used to reconstruct the acoustic field with the same efficiency as the traditional DE approach.

Depending on the application, we can find problems that will involve the creation of very large matrices. These problems cannot be handled with either direct or indirect formulations that require the explicit inversion of matrices, since this will be prohibitively time and memory consuming. Instead implicit methods will be a good alternative (and maybe the only available alternative).

In this work the use of implicit methods was presented for exterior NAH, but these methods can also be used for interior NAH, which will only involve a slight change on the integral formulas given in Sec. II.

ACKNOWLEDGMENTS

The experimental measurement data from the planar hologram was provided by Brian Houston and Peter Herdic from the Naval Research Laboratory. The authors thank the reviewers for many helpful comments. This work was supported by ONR and NRC.

[1] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, UK, 1999).
[2] J. D. Maynard, ''Acoustic holography for wideband, odd-shaped noise sources,'' in *Proceedings Inter-noise '88*, Avignon, France, September 1988, Vol. I, pp. 223–231.
[3] W. A. Veronesi and J. D. Maynard, ''Digital holographic reconstruction of sources with arbitrarily shaped surfaces,'' J. Acoust. Soc. Am. **85**, 588–598 (1989).
[4] G.-T. Kim and B.-H. Lee, ''3-d sound source reconstruction and field reprediction using the helmholtz integral equation,'' J. Sound Vib. **136**, 245–261 (1990).
[5] M. R. Bai, ''Application of bem (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries,'' J. Acoust. Soc. Am. **92**, 533–549 (1992).
[6] G. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, ''Generalized nearfield acoustic holography for axisymmetric geometries,'' J. Acoust. Soc. Am. **88**, 199–209 (1990).
[7] A. Sarkissian, ''Near-field acoustical holography for an axisymmetric geometry: a new formulation,'' J. Acoust. Soc. Am. **88**, 961–966 (1990).
[8] A. Sarkissian, C. F. Gaumond, E. G. Williams, and B. H. Houston, ''Reconstruction of the acoustic field over a limited surface area on a vibrating cylinder,'' J. Acoust. Soc. Am. **93**, 48–54 (1993).
[9] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems* (Kluwer Academic, Boston, 1996).
[10] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia, 1998).
[11] B.-K. Kim and J.-G. Ih, ''On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method,'' J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[12] P. A. Nelson and S. H. Yoon, ''Estimation of acoustic source strength by inverse methods: Part i, conditioning of the inverse problem,'' J. Sound Vib. **233**, 643–668 (2000).
[13] S. H. Yoon and P. A. Nelson, ''Estimation of acoustic source strength by inverse methods: Part ii, experimental investigation of methods for choosing regularization parameters,'' J. Sound Vib. **233**, 669–705 (2000).
[14] B.-K. Kim and J.-G. Ih, ''Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by nah,'' J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[15] E. G. Williams, ''Regularization methods for near-field acoustical holography,'' J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[16] S. Sureshkumar and S. T. Raveendra, ''An analysis of regularization errors in generalized nearfield acoustical holography,'' in *2001 SAE noise and vibration conference*, no. 2001-01-1616, Grand Traverse, MI, April 2001.
[17] S.-C. Kang and J.-G. Ih, ''On the accuracy of nearfield pressure predicted by the acoustic boundary element method,'' J. Sound Vib. **233**, 353–358 (2000).
[18] S.-C. Kang and J.-G. Ih, ''The use of partially measured source data in near-field acoustical holography based on the bem,'' J. Acoust. Soc. Am. **107**, 2472–2479 (2000).
[19] S.-C. Kang and J.-G. Ih, ''Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography,'' J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[20] S. F. Wu and X. Zhao, ''Combined helmholtz equation-least squares method for reconstructing acoustic radiation from arbitrarilty shaped objects,'' J. Acoust. Soc. Am. **112**, 179–188 (2002).
[21] A. Tekatlian, P. Filippi, and D. Habault, ''Determination of vibration characteristics of noise sources solving an inverse radiation problem,'' Acustica **82**, 91–101 (1996).
[22] T. K. DeLillo, V. Isakov, N. Valdivia, and L. Wang, ''The detection of the source of acoustical noise in two dimensions,'' SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **61**(6), 2104–2121 (2001).
[23] T. K. Delillo, V. Isakov, N. Valdivia, and L. Wang, ''Computational methods for the detection of the source of acoustical noise,'' in *Proceedings of ASME2000*, Orlando, FL, November 2000.
[24] T. K. DeLillo, V. Isakov, N. Valdivia, and L. Wang, ''The detection of surface vibrations from interior acoustical pressure,'' Inverse Probl. **19**(3), 507–524 (2003).
[25] S. T. Raveendra, N. Vlahopoulos, and A. Glaves, ''An indirect boundary element formulation for multi-valued impedance simulation in structural acoustics,'' Appl. Math. Model. **22**, 379–393 (1998).
[26] N. Vlahopoulos and S. T. Raveerdra, ''Formulation, implementation and validation of multiple connection and free edge constraints in an indirect boundary element formulation,'' J. Sound Vib. **201**, 137–152 (1998).

[27] F. Augusztinovicz, ''Application and extension of acoustic holography techniques for tire noise investigations,'' J. Acoust. Soc. Am. **105**, 1373 (1999).
[28] E. G. Williams, B. H. Houston, P. C. Herdic, S. T. Raveendra, and B. Gardner, ''Interior nah in flight,'' J. Acoust. Soc. Am. **108**, 1451–1463 (2000).
[29] Z. Zhang, N. Vlahopoulos, S. T. Raveendra, T. Allen, and K. Y. Zhang, ''A computational acoustic field reconstruction process based on an indirect boundary element formulation,'' J. Acoust. Soc. Am. **108**, 2167–2178 (2000).
[30] Z. Zhang, N. Vlahopoulos, T. Allen, and K. Y. Zhang, ''A source reconstruction process based on an indirect variational boundary element formulation,'' Eng. Anal. Boundary Elem. **25**, 93–114 (2001).
[31] A. Schuhmacher, J. Hald, K. B. Rasmussen, and P. C. Hansen, ''Sound source reconstruction using inverse boundary element calculations,'' J. Acoust. Soc. Am. **113**, 114–126 (2003).
[32] D. Colton and R. Kress, *Inverse Acoustics and Electromagnetic Scattering Theory* (Springer-Verlag, New York, 1992).
[33] D. Colton and R. Kress, *Integral Equation Methods in Scattering Theory* (Wiley-Interscience Publication, New York, 1983).
[34] J. Pan and D. A. Bies, ''The effect of fluid-structural coupling on sound waves in an enclosure—theoretical part,'' J. Acoust. Soc. Am. **87**, 691–707 (1990).
[35] D. Ouellet, J. L. Guyader, and J. Nicolas, ''Sound field in a rectangular cavity in the presence of a thin, flexible obstacle by integral equation method,'' J. Acoust. Soc. Am. **89**, 2131–2139 (1991).
[36] E. G. Williams, ''On Green functions for a cylindrical cavity,'' J. Acoust. Soc. Am. **102**, 3300–3307 (1997).
[37] T. K. DeLillo, T. Hrycak, and N. Valdivia, ''Iterative regularization methods for inverse problems in acoustics,'' in *Proceedings of ASME2002*, New Orleans, LA, November 2002.
[38] C. Langrenne and A. Garcia, ''Integral formulations for the vibroacoustic characterization of a cello,'' J. Acoust. Soc. Am. **105**, 1088 (1999).
[39] A. F. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, ''An advanced computational method for radiation and scattering of acoustic waves in three dimensions,'' J. Acoust. Soc. Am. **77**, 362–368 (1985).
[40] D. Chien, ''Numerical evaluation of surface integrals in three dimensions,'' Math. Comput. **64**(210), 727–743 (1995).
[41] K. E. Atkinson, *The Numerical Solution of Integral Equations of the Second Kind* (Cambridge U.P., New York, 1997).
[42] A. H. Strout, *Approximate Calculation of Multiple Integrals* (Prentice–Hall, Englewood Cliffs, NJ, 1971).
[43] C. Schwab and W. L. Wendland, ''On numerical cubatures of singular surface integrals in boundary element methods,'' Numer. Math. **62**(3), 343–369 (1992).
[44] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).
[45] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 9th ed. (Dover, New York, 1972).
[46] E. G. Williams and B. Houston, ''Fast Fourier transform and singular value decomposition for patch nearfield acoustical holography,'' J. Acoust. Soc. Am. **114**(3), 1322–1333 (2003).

Subjective response of people to simulated sonic booms in their homes[a)]

David A. McCurdy[b)]
Structural Acoustics Branch, NASA Langley Research Center, MS 463, Hampton, Virginia 23681

Sherilyn A. Brown
Systems Analysis Branch, NASA Langley Research Center, MS 348, Hampton, Virginia 23681

R. David Hilliard
Wyle Laboratories, 3200 Magruder Boulevard, Hampton, Virginia 23666

(Received 2 August 2002; revised 5 May 2004; accepted 14 June 2004)

In order to determine the effect of the number of sonic boom occurrences on annoyance, a computer-based system was developed for studying the subjective response of people to the occurrence of simulated sonic booms in their homes. The system provided a degree of control over the noise exposure not found in community surveys and a degree of situational realism not available in the laboratory. A system was deployed for eight weeks in each of 33 homes. Each day from 4 to 63 sonic booms were played as the test subject went about his or her normal activities. At the end of the day, the test subjects rated their annoyance to the sonic booms heard during the day. The sonic booms consisted of different combinations of waveforms, levels, and occurrence rates. The experiment confirmed that the increase in annoyance resulting from multiple occurrences can be modeled by the addition of the term ''10 $*$ log(number of occurrences)'' to the sonic boom level. Of several noise metrics considered, perceived level was the best annoyance predictor. Comparisons of the subjective responses to the different sonic boom waveforms found no differences that were not accounted for by the noise metrics. [DOI: 10.1121/1.1781189]

PACS numbers: 43.50.Qp, 43.50.Ba, 43.50.Pn [DKW] Pages: 1573–1584

I. INTRODUCTION

Attempts to understand the impact of sonic boom exposure on the public began in the 1960s and continued in the following decade with efforts to determine sonic boom acceptance criteria for commercial overland supersonic flight. The advent of the first-generation American supersonic transport (SST) and European Concorde development programs in the 1970s resulted in the prohibition of overland supersonic flight in the United States and elsewhere. The National Aeronautics and Space Administration (NASA) revisited the topic as part of its second-generation commercial supersonic transport research program during the 1990s. Technology for shaping sonic boom pressure signatures by tailoring aircraft volume and lift distributions was developed in this program and validated in a flight test in 2003. The potential of this technology to reduce sonic boom impact, along with substantial current interest in developing a small supersonic civilian aircraft,[1] underscores the ongoing need for improved sonic boom acceptance criteria for overland supersonic flight.

Early attempts (e.g., Ref. 2) to develop acceptance criteria utilized subsonic and supersonic aircraft flyovers and juries of test subjects located both outdoors and indoors. These comparisons of single flyover events enabled sonic boom signatures to be equated with subsonic flyover noises. However, the use of actual supersonic aircraft flyovers precludes the systematic variation of sonic boom signature parameters and limits the available sonic boom signatures to those of existing aircraft. It is also expensive, difficult to arrange and execute, and can impact people not involved in the experiment. Therefore, there have been several efforts (e.g., Refs. 3–5) spanning the past four decades to utilize sonic boom simulators to determine subjective reaction to sonic booms. These simulator studies, almost exclusively laboratory-based, have identified key characteristics of sonic boom signatures and proposed metrics to enable the prediction of loudness and annoyance of single sonic booms. Studies such as these are useful for comparison of individual sonic boom signatures, but shed little light on criteria applicable to communities of people exposed to multiple sonic booms on a regular basis in their homes. Clearly, in addition to the characteristics of the individual sonic booms, the frequency or number of their occurrence would be an important variable in community reaction. A few studies (e.g., Refs. 6 and 7) have used surveys to determine community residents' reactions to sonic boom exposures resulting from actual aircraft overflights. Based on interviews with residents the important (negative) effects of sonic boom exposure have been identified, and include annoyance, startle, building vibration and associated rattles, and perceived building damage. In contrast to studies of residents' reactions to aircraft noise around airports, these sonic boom studies are extremely limited in terms of the numbers of communities and the ranges of exposures.

[a)]Portions of this work were presented in ''The effects of simulated sonic booms on people in their homes,'' 33rd AIAA Aerospace Sciences Meeting and Exhibit, 9–12 January 1995, AIAA 95-0834, Reno, NV.

[b)]Electronic mail: d.a.mccurdy@nasa.gov

Thus there is no well-accepted, standardized guidance on sonic boom acceptance criteria. Based on airport studies, integrated metrics such as equivalent continuous sound level (commonly known as L_{eq}) and day–night average sound level (commonly known as L_{dn} or DNL) have found widespread regulatory acceptance. Implicit in these metrics is a simple trade-off between the amplitude of individual flyover noises and their frequency of occurrence: 3 dB per halving or doubling of the number of events. This ''energy addition theory'' is modeled by the addition of the term ''10 $*$ log(number of occurrences)'' to the individual event level. Whether such a simple trade-off is applicable to impulsive sounds that occur relatively infrequently, as anticipated for sonic boom exposure, is the primary question addressed in this study.

A laboratory study designed to address this issue would be problematic and probably unsatisfactory due to the unrepresentative environment. A field study using real aircraft would be prohibitively complex and expensive and limited to existing sonic boom signatures. An alternative approach, adopted for this study, is the creation of a laboratory-style apparatus in people's homes. This allows long duration testing, situational realism not available in a laboratory, and a degree of control over the sonic boom exposure not found in community surveys. Such an approach has been previously demonstrated in a study[8] that was aimed at determining ''acceptable'' levels of sonic boom, but not the effect of number of events. The apparatus developed for the present study, the in-home noise generation/response system (IHONORS), is a computer-controlled audio system that generates simulated sonic booms, measures their levels, and records the test subject's ratings. Eight of the systems were rotated through 33 homes for 8 weeks in each home, providing a total of 1848 days of sonic boom exposures.

As indicated, the primary objective of this study is to determine the effect on annoyance of the number of sonic boom occurrences in a realistic environment. Analyses of the resulting data also yield information on a number of secondary questions, including (1) does the waveform shape affect annoyance; (2) what is the effect of startle on annoyance; and (3) which noise metrics best predict annoyance to sonic booms. Conclusions are based on statistical analyses of the test subjects' annoyance judgments and measured noise levels for each daily exposure to sonic booms.

II. NOISE METRICS, SYMBOLS, AND ABBREVIATIONS

DNL	day-night average sound level, dB
DNL(A)	day–night average sound level calculated using A-weighted sound pressure level, dB
DNL(C)	day–night average sound level calculated using C-weighted sound pressure level, dB
DNL(PL)	day–night average sound level calculated using perceived noise level, dB
IHONORS	in-home noise generation/response system
LLZd	Zwicker loudness level (diffuse field), dB
LLZf	Zwicker loudness level (frontal field), dB
n	number of occurrences of a sonic boom (i.e., the number of sonic booms actually heard by a test subject during the day)
PL	perceived level (Stevens Mark VII procedure), dB
PNL	perceived noise level, dB
SEL(A)	A-weighted sound exposure level, dB
SEL(C)	C-weighted sound exposure level, dB
SEL(U)	unweighted sound exposure level, dB

III. EXPERIMENTAL METHOD

A. Scope

As discussed earlier, the main effects of sonic boom exposure are annoyance, startle, building vibration and associated rattle, and perceived building damage. This experiment was designed to focus on annoyance response. Some information was obtained on startle, but the experiment was not designed to provide definitive results concerning startle. Building vibration and rattle, while recognized as important to the acceptability of sonic booms, were beyond the capability of the test apparatus to control or measure and, as a practical matter, beyond the reach of the study. The simulated sonic booms did cause varying degrees of building vibration and rattle in the homes. The extent of these effects varied by home, by room within the home, and by individual boom exposure. This variability, the large sample of homes, and their random selection controlled, at least to some extent, the impact of these effects on the results of the study. Perceived building damage was not addressed in the experiment and did not arise spontaneously as a test subject reaction.

Despite its crucial advantages, the test methodology chosen for this experiment does have limitations. One is the inability to reproduce the low frequency content of a sonic boom signature. Without actual supersonic overflights or the small spatial volume of a laboratory simulator, it is difficult to reproduce the frequencies below 20 Hz at realistic amplitudes. This is especially true if the sound generation system must be moved to, set up, and operated at existing homes without exposing the entire neighborhood to the resulting sonic boom stimuli. The IHONORS test apparatus, built in the early 1990s, has an essentially flat frequency response down to 20 Hz, the lower limit of hearing. Below 20 Hz its frequency response fell off rapidly. For determining the effect of number of sonic boom occurrences on annoyance, the advantages of this test methodology were judged to outweigh this shortcoming.

B. Test facility

A diagram of the IHONORS system is shown in Fig. 1. The system consisted of a computer and compact disc player that played the simulated sonic booms at randomly selected, preprogrammed times through a preamplifier and amplifier into three or four loudspeakers located in different rooms of the house. No warning was given as to when the booms would occur. The two indoor microphones and sound level meters measured the levels of the booms as they occurred and also continuously measured the ambient noise levels in the home. The measurements were then transferred to the computer and stored on its hard disk. At the end of the day, the test subject used the trackball to answer a series of questions about his or her activities during the day and his or her overall subjective response to the total sonic boom exposure

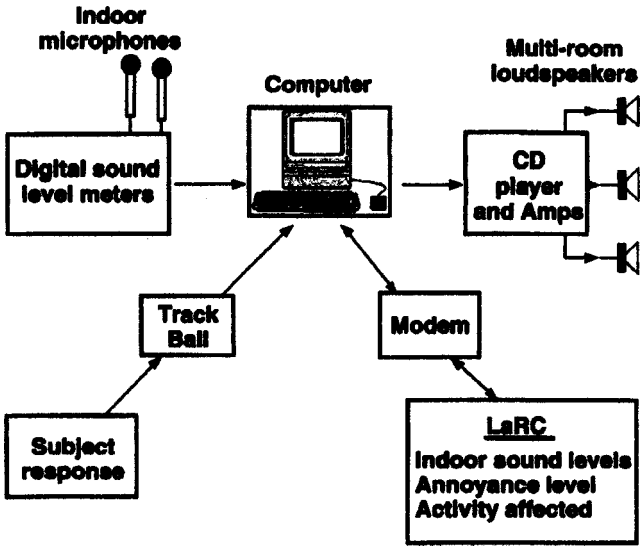


FIG. 1. In-home noise generation/response system diagram.

for the day (i.e., to all the sonic booms heard during the day). Once a week the noise measurements and test subject responses were downloaded via a modem to a central computer. The data were then checked to ensure that the system was operating correctly.

Figure 2 shows the actual components as deployed in a test subject's home. The homes selected were limited to single family detached dwellings so that the simulated sonic booms would not penetrate the walls into neighbors' homes, as might be the case in apartments or townhouses. The selected homes covered the economic range from lower middle class to upper middle class. The computer, compact disc player, trackball, and monitor were placed on the shelves of a microwave oven cart. The preamplifier, amplifiers, and sound level meters were placed inside the lower cabinet of the cart. The cart was placed in a position that was convenient for the test subject and simplified the installation of cables. A loudspeaker was placed in each of the three or four rooms that the test subject indicated he or she occupied most frequently during the 14-h boom period each day. The loudspeakers, which sat on the floor and were not attached to the building, were positioned to provide unobstructed coverage of the rooms. The two microphones were placed in two of the rooms with loudspeakers. To the extent possible, the system components were located so as to minimize their disruption of the family's routine and normality, and, in general, they blended in well with the existing home decor.

C. Test subjects

Thirty-three test subjects and their homes were randomly selected from a pool of local residents meeting certain participation requirements. First, to maximize the number of sonic booms heard by each test subject, the test subjects chosen were normally at home all day. This requirement eliminated people with jobs or other responsibilities that regularly took them outside the home for a significant portion of the day. Homes with children under eight years of age were eliminated because hearing testing equipment for small children was not available. Homes with pregnant women were eliminated as a safety precaution. Since everyone in a household was exposed to the sonic booms, all members of each household were required to complete a medical questionnaire for approval by a NASA physician. Only one occupant in each home was allowed to serve as a test subject and answer questions. Each test subject was paid $600 to participate in the experiment. All test subjects and home occupants were given audiograms prior to the experiment to verify normal hearing. The 33 test subjects, 25 female and 8 male, ranged in age from 32 to 71 years old. Table I gives the gender and age data for the subjects.

D. Noise stimuli

The simulated sonic booms presented to the test subjects represented combinations of three sonic boom pressure signatures or waveforms, three A-weighted sound exposure levels, and seven sonic boom occurrence rates.

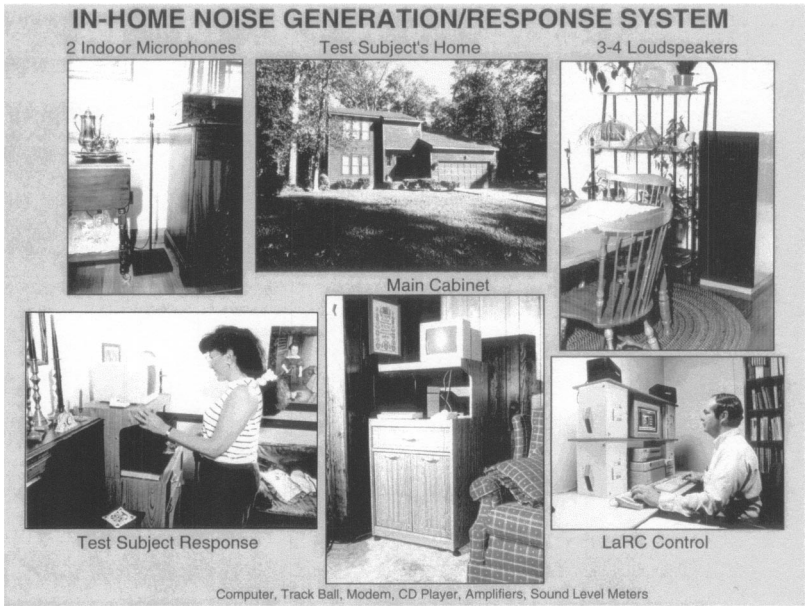


FIG. 2. In-home noise generation/response system in a test subject's home.

TABLE I. Test subject data.

Gender	No. of participants	Mean age	Median age	Age range
Male	8	52	54	32–71
Female	25	50	52	34–71
All subjects	33	51	53	32–71

Initially, the three sonic boom waveforms were to be a combination of conventional N-wave pressure signatures and projected ''shaped'' pressure signatures as heard indoors. To obtain indoor pressure signatures, digitally generated outdoor N-wave and ''shaped'' waveforms were passed through a digital frequency filter that modeled the transmission of sonic booms through a house with windows closed.[9] This house filter eliminated most of the higher frequency content that differentiated the outdoor N-wave and ''shaped'' sonic booms, resulting in the indoor signatures sounding very similar to one another. Since the original purpose of including more than one waveform in the experiment was to examine the effect of number of sonic boom occurrences over as wide a range of spectral content as possible (i.e., audibly different sonic booms), it was decided to use both indoor and outdoor waveforms in the experiment. Intuitively, it may seem contradictory to use outdoor sonic boom signatures inside a house, but doing so enabled the effect of number of occurrences, the primary test objective, to be studied over a more robust spectral content range.

The three pressure signatures eventually selected for the experiment represented an N-wave as heard indoors, an N-wave as heard outdoors, and a ''shaped'' wave as heard outdoors. These digitally generated waveforms provided one spectrum with relatively little high frequency content and two spectra with audibly different high frequency content. Figure 3 illustrates the difference between an outdoor N-wave and an outdoor waveform ''shaped'' to minimize the sonic boom.[9] (This shaping decreases the high frequency content of the sonic boom, which mainly affects loudness and annoyance. It has little effect on the low frequency content primarily responsible for building vibration and rattle.) Both waveforms were symmetrical with rise times of 4 ms and durations of 300 ms. The ''shaped'' waveform had a secondary rise time of 30 ms and the ratio of the front-shock overpressure to the peak overpressure was 0.3.

Given that the frequency response of the sound reproduction system rolled off rapidly below 20 Hz, the frequency content of the waveforms below 20 Hz was digitally filtered out and, therefore, not present in the final three simulated sonic boom signatures. This high-pass filtering enabled the sound system to produce higher sonic booms levels. The final three simulated sonic boom signatures are shown in Fig. 4.

In order to create the desired sound levels, the amplitudes of the three sonic boom signatures were adjusted to produce a constant SEL(A) across all three signatures. Each of the three signatures was then reproduced at three levels, 4 dB apart, for a range of 8 dB. These nine combinations of

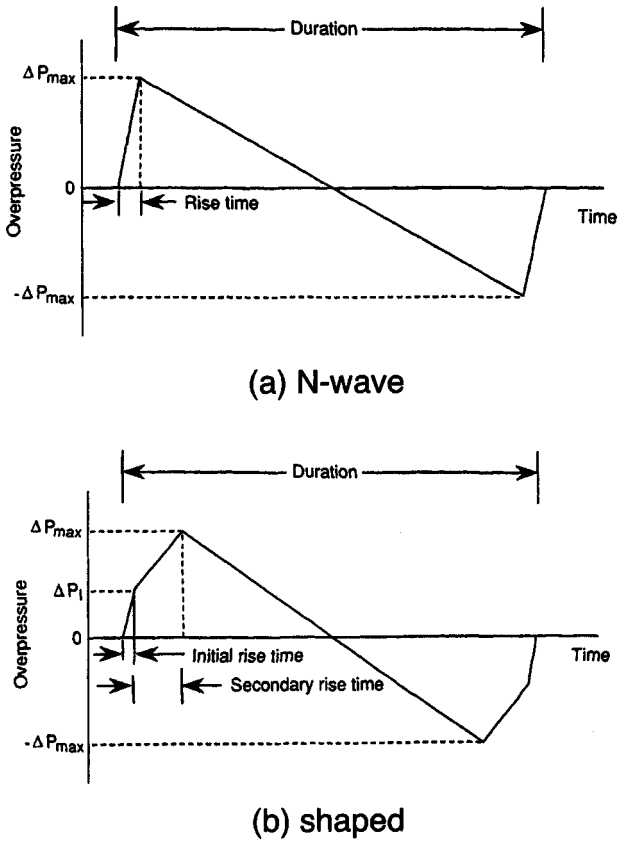


FIG. 3. Shape parameters for outdoor sonic boom pressure waveforms: (a) N-wave and (b) shaped.

waveform and level were then recorded on separate tracks of a compact disc.

The three SEL(A) levels at which the simulated sonic booms were presented in each home were nominally 66, 70, and 74 dB. These values covered the range of indoor sonic boom levels estimated for a variety of second-generation supersonic transport designs. (The use of outdoor waveforms in the experiment might, instinctively, lead one to consider presenting the sonic booms at outdoor levels. All other issues aside, the sound reproduction system was incapable of presenting the sonic booms at outdoor levels.) Actual levels achieved in each home for each waveform varied based on the size, shape, and layout of the house. Instead of the nominal 8 dB span, the typical range of levels within a home was 10 to 11 dB. In other words, although the sonic booms were recorded onto the compact disc at equal SEL(A) levels, differences in the acoustical characteristics of the rooms and homes resulted in changes in levels due to the spectral differences amongst the booms. (Home to home variation increased the range of levels in the data to 15 dB, 2 dB of which is directly attributable to just two homes.)

The occurrence rates were 4, 10, 13, 25, 33, 44, and 63 booms per daily 14-hour test period. Only one sonic boom waveform was presented each day. On most days the sonic boom was presented at only one SEL(A) level. On a few days the sonic boom was presented at two or three of the SEL(A) levels.

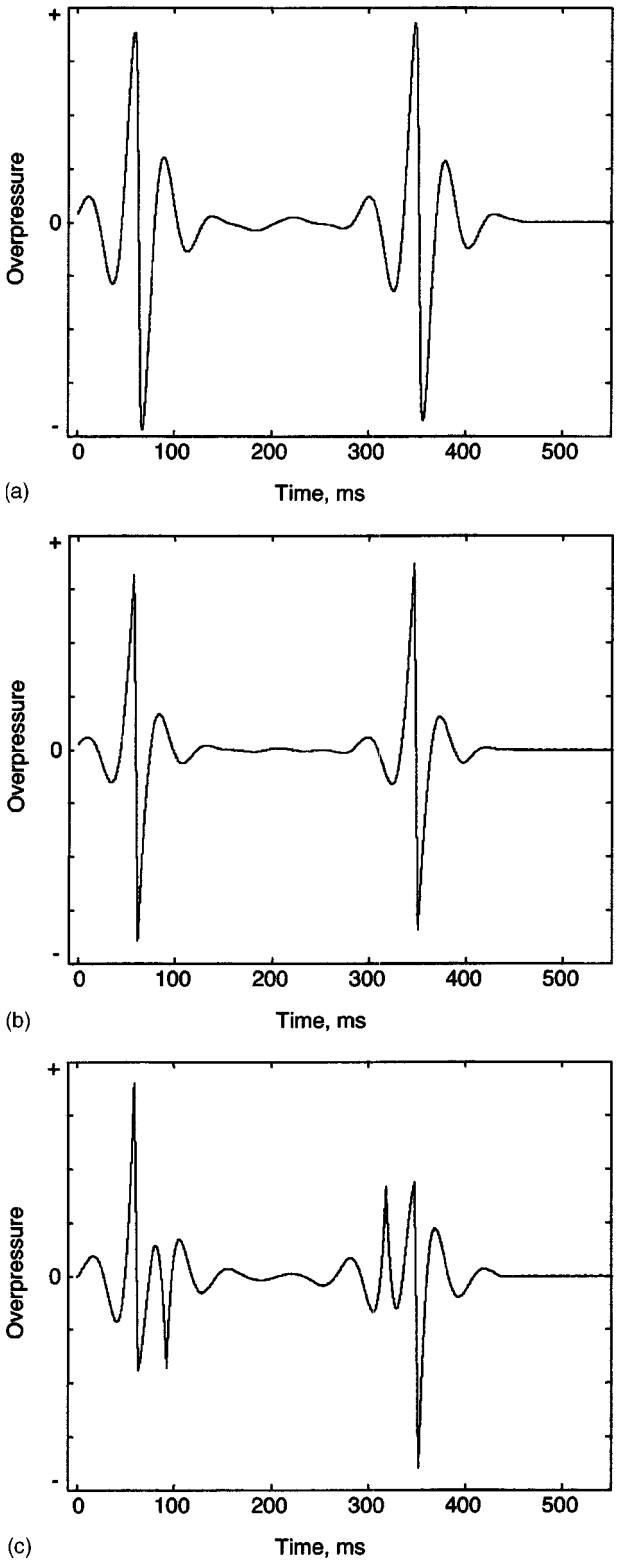


FIG. 4. Pressure signatures of sonic boom test stimuli after high-pass filtering at 20 Hz: (a) Indoor N-wave. (b) Outdoor N-wave. (c) Outdoor ''shaped'' wave.

E. Experiment design

The factorial combination of three levels, two waveforms (outdoor and indoor N-waves), and four occurrence rates (4, 10, 25, and 63 booms per day) yielded 24 daily exposures. The third waveform (outdoor, shaped) was presented at three levels and three occurrence rates (10, 25, and 63 booms per day) yielding nine additional daily exposures. Each waveform was then presented at more than one level within a day at occurrence rates of 13, 33, and 44 booms per day. This added another nine additional exposure days for a total of 42 unique exposure days. (The 13-occurrence, mixed-level days included 1 boom at the high level, 4 booms at the middle level, and 8 booms at the low level. The 33-occurrence, mixed-level days included 3 high level booms, 9 middle level booms, and 21 low level booms. The breakdown for the 44-occurrence, mixed-level days was 12 high level and 32 middle level booms.)

The constant-level, 63-occurrence exposure days were repeated for an additional 6 exposure days. Eight of the exposure days, including all three levels, all three waveforms, all seven occurrence rates and both constant- and mixed-level days, were selected to also be presented twice. The first presentations of these eight exposure days were assigned to the first eight days of each subject's testing period. The combination of the 42 unique exposure days, the 6 repeat exposure days, and the 8 ''introductory'' days resulted in a total of 56 exposure days (8 weeks).

The order of presentation of the 8 introductory days was varied between subjects based on a latin square design. The order of the remaining 48 exposure days was randomly assigned to days 9 through 56 of each subject's testing period. A new random order was determined for each test subject.

Eight IHONORS systems were used to conduct the in-home study of subjective response to simulated sonic booms. The testing began in February 1993 and was completed in December 1993. During the testing a system was deployed for eight weeks in each of 33 homes. On each day of a deployment the system played simulated sonic booms during a 14-h period as the test subject went about his or her normal activities. Although the booms were spread out evenly throughout the 14 h, the exact occurrence times were randomized. The combinations of waveforms, levels, and occurrence rates were also randomized across days for each test subject. The 14-h test period was selected so that no booms were played during the test subject's normal sleep period. At the end of the 14 h the test subject answered a series of questions and rated his or her overall annoyance to all the sonic booms heard during the day. The test subject could answer the questions anytime between the end of the test period and 4:00 a.m. the following morning.

The computer-generated questions were (1) when were you not inside your house, (2) what activities did you do while inside your house, (3) how annoying were all the sonic booms you heard today on a 0 to 10 scale, and (4) were you startled by any of the sonic booms today (yes or no)? Test subjects were instructed to indicate their individual reactions to the sonic booms and to facilitate this by not discussing the reactions of other household members.

Numerical category scaling was chosen as the psychophysical method for rating annoyance. The scale selected was a unipolar, 11-point scale from 0 to 10. The scale was positioned vertically like a thermometer. The lower end point was labeled ''NOT ANNOYING AT ALL'' and the upper end point was labeled ''EXTREMELY ANNOYING.'' The term

''ANNOYING'' was defined as ''UNWANTED, OBJECTIONABLE, DISTURBING, OR UNPLEASANT.''

F. Procedure

The process of conducting the experiment in each home consisted of six stages that included at least three visits to the home. The stages were (1) preliminary visit, (2) installation and calibration preparation, (3) installation and calibration visit, (4) testing period, (5) removal visit, and (6) posttest activities. The people conducting the experiment included the test conductor, the subject coordinator, the equipment manager, and the equipment team. Each stage of the process is discussed in the following paragraphs.

1. Preliminary visit

Once a potential test subject and home were identified by the subject coordinator, a preliminary visit to the home was conducted by the test conductor, the subject coordinator, and the equipment manager. At the start of the visit, which typically lasted about 3 h, the experiment and its requirements and procedures were explained to the subject. If the subject was still interested in participating, a test schedule for the home was determined. Locations for the system components and cables were selected and sketched. A system computer was setup and a communication check of the home's telephone line conducted via modem. Finally, the subjective response software was demonstrated for, and exercised by, the test subject.

2. Installation and calibration preparation

The time between the preliminary visit and system installation ranged from one to several weeks depending on the test schedule agreed to with the subject. During this time all the household members completed the necessary consent forms and medical questionnaire. They were also given hearing tests to confirm their hearing was normal. The medical questionnaire was reviewed and approved by a physician. The IHONORS system assigned to the home was checked for proper operation and its computer configured with the necessary information concerning the particular home involved.

3. Installation and calibration visit

The installation and calibration of a system in a home took from $1\frac{1}{2}$ to 2 days and was done by the test conductor, equipment manager, and equipment team. While the equipment team set up the system in the locations determined during the preliminary visit, the test conductor administered a pretest questionnaire to the subject. Once the system was installed, the amplifier output gains were adjusted for each loudspeaker so that the sonic boom sound levels in the center of each room were set as closely as possible to the nominal values. Noise measurements and digital recordings of the test stimuli were then made at each room's center position and final monitor microphone locations. Measurements and recordings were also made at other positions in the rooms (i.e., favorite chairs, stove) where the test subject might stand or sit for long periods of time. (After extensive comparisons of the noise measurements made using different equipment and at different locations within each room, the noise measurements used in the experiment analyses were based on the recordings made at the center position of the room occupied most often by the test subject.) In most cases the subject heard all the test stimuli during the calibration process. If not, the different stimuli were played for the subject. The subjective response software was again reviewed and exercised by the subject until the test conductor was confident the subject understood how to respond to the questions presented nightly by the computer. The system was then set into automatic mode and the experiment began the following day. Instructions for contacting the test conductor were left with the subject in case any problems or questions arose.

4. Testing period

During the testing period, the system would play each day's preprogrammed series of test stimuli and present the subjective response questions for the subject to answer each night. On the fourth day (nominally) of the period, the test conductor called the subject to ask if there were any problems or questions. This call was repeated midway through the testing period. Each week the equipment manager downloaded the noise measurement and subjective response data from the computer via modem. This data was then checked to confirm that the testing was proceeding as planned. If at any time a problem developed, arrangements were made with the test subject for the test conductor and/or the equipment manager to visit the home and correct the problem. The number of visits and their duration were kept to an absolute minimum in order to minimize any effect on the subject's responses.

5. Removal visit

The removal visit was normally conducted the day after the last day of testing. The system components were photographed in place before removal. The test conductor administered a posttest questionnaire to the subject while the equipment manager ran a check of the system. A series of noise measurements was made at the monitor microphone positions and at the centers of the rooms with loudspeakers to confirm proper system operation and as a backup to the installation calibration measurements. The equipment team then removed the equipment and the subject was thanked for his or her participation. The removal was done in one day.

6. Posttest activities

The members of the household were given hearing tests again as a safety precaution. Once the hearing tests were completed, the subject coordinator paid the test subject. The equipment manager downloaded all the data from the system computer, checked it for errors, and prepared data files for the analyses.

TABLE II. Summary of test exposure days.

Type of day	No. of days	Percentage of total days
Subject responded with annoyance judgment	1676	90.7%
Subject was outside all day	19	1.0%
Subject did not respond	47	2.6%
Day voided due to technical problem	106	5.7%
Total exposure days	1848	100.0%

IV. RESULTS AND DISCUSSION

A. Subjective data analyses

The eight IHONORS systems were deployed in 33 test subjects homes for a total of 1848 exposure days. Table II gives a breakdown of the test results during those 1848 days. Over 90% of the time, the subjects responded with an annoyance judgment. Subjects were outside of the home for the entire 14-h period only 1% of the time. Subjects were out of town or forgot to respond to the questions less than 3% of the time. Technical problems with the system caused the loss of subjective data on less than 6% of the exposure days. The systems were designed to, and did, recover automatically from power outages by rebooting. However, the most common technical problem was a momentary brownout or flicker in the power supply that did not trigger a reboot of the system but did cause the computer to lose its default memory settings.

B. Acoustic data analyses

The noise metrics considered were SEL(U), SEL(A), SEL(C), LLZd, LLZf, PNL, and PL. SEL(U) is the unweighted sound exposure level, which does not account for the frequency response characteristics of the ear. SEL(A) and SEL(C) are sound exposure levels based on the relatively simple A- and C-frequency weightings used to model the ear's response characteristics. SEL(A) is often used to assess airport community noise, while SEL(C), which does not attenuate low frequencies as much as SEL(A), is often used to evaluate impulse noise sources such as pile drivers and artillery, which produce noise similar in character to sonic booms. PL, LLZd, and LLZf are measures of loudness that are based on more complex level-dependent frequency weighting systems. PNL is a measure of noisiness that is also based on a more complex level-dependent frequency weighting system and is usually used in aircraft noise certification. The calculation method used to obtain the noise metrics from the sonic boom digital pressure signatures is described in Ref. 10. Detailed descriptions of all the different noise metrics can be found in Refs. 11 and 12.

Of the 1676 test exposure days for which a subjective annoyance response was obtained, noise measurements for 3 days were lost due to technical problems incurred during recording of the sonic booms. Therefore, the basic data set for the experiment consisted of 1673 exposure days.

C. Introductory exposure days

Initially, data analyses were based on the basic data set of 1673 test exposure days.[5,13] However, as mentioned in the noise stimuli section, the first 8 days of each testing period consisted of eight introductory exposures that were repeated on randomly selected days later in the testing period. Comparison of the introductory exposures with their repeats indicated that the test subjects adapted to the new sonic boom noise environment during the first few days of the testing period. Therefore, the introductory days were deleted from the data set used in the following analyses except where specifically stated otherwise. The resulting analysis data set consisted of 1441 exposure days. (The adaptation effect analysis is discussed later in this report. Eliminating the introductory days from the data set did not significantly affect the results and conclusions of the initial analyses previously reported.[5,13])

D. Effect of number of sonic boom occurrences

Although the number of sonic booms played in a home each day was set to one of seven occurrence rates (4, 10, 13, 25, 33, 44, and 63 booms per daily 14-h test period), the number of booms actually heard by the test subject depended on whether or not the subject was present in the home. Therefore, the number of sonic booms actually heard by a test subject during the day (i.e., the ''number of sonic boom occurrences'') is the number of booms presented by the system minus the number of booms that were presented during the time periods that the subject indicated that he or she was not inside the house. This number of sonic boom occurrences ranged from 1 to 63.

To reduce the number of plots required to illustrate the effect of number of sonic boom occurrences, the data is divided into five subsets, each having roughly the same number of data points. Subset (a) includes cases with number of occurrences between 1 and 5, subset (b) those between 6 and 11, (c) between 12 and 23, (d) between 24 and 41, and (e) between 42 and 63. Figure 5 displays the data for each subset in terms of subjective annoyance rating versus the SEL(A) of the individual sonic boom that was repeated during the day. The linear regression line for each subset is also plotted. As is seen in the Fig. 5 plots, the variance (scatter) in the data makes it difficult to discern effects by visual comparison of the data points only.

Figure 6 illustrates the effect of number of sonic boom occurrences by combining the five linear regression lines from Fig. 5 onto one plot of subjective annoyance rating as a function of the SEL(A) level of the individual sonic boom repeated during the day. This figure clearly illustrates that subjective annoyance response increases as the number of occurrences of a sonic boom increases.

This effect of number of occurrences is usually modeled by the addition of the term ''$k*\log(\text{number of occurrences})$'' to the measured sonic boom level. The value of ''k'' can then be determined from the coefficients of a multiple regression equation as follows:

$$A = a_0 + a_1 * [L + k * \log(n)], \tag{1}$$

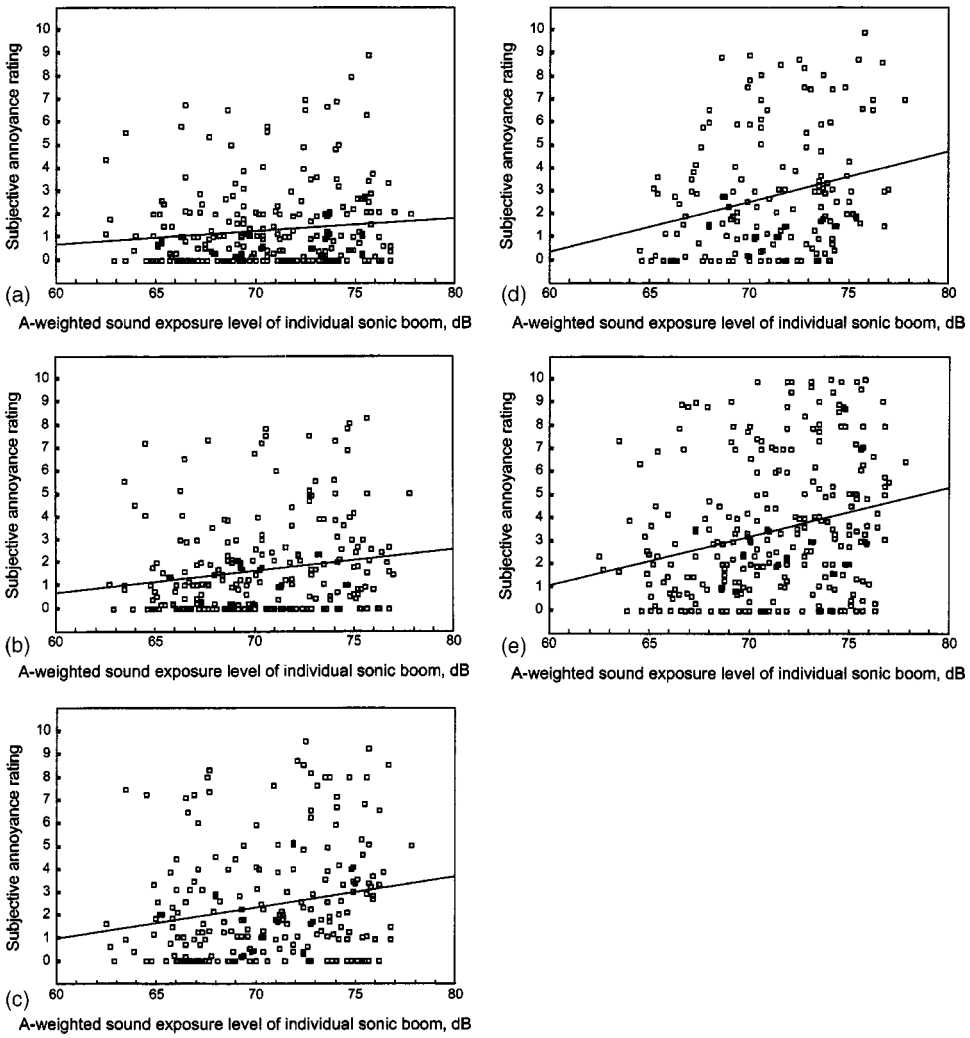


FIG. 5. Subjective annoyance rating versus individual sonic boom A-weighted sound exposure level for different numbers of sonic boom occurrences. (a) Number of occurrences equals 1 to 5. (b) Number of occurrences equals 6 to 11. (c) Number of occurrences equals 12 to 23. (d) Number of occurrences equals 24 to 41. (e) Number of occurrences equals 42 to 63.

where A = annoyance rating, L = individual sonic boom level, and n = number of sonic boom occurrences. Expanding the equation and replacing "$a_1 * k$" with the standard notation "a_2" gives the regression equation

$$A = a_0 + a_1 * L + a_2 * \log(n) \tag{2}$$

with "k" equal to a_2/a_1.

Figure 7 shows the values of "k" (i.e., a_2/a_1) and the corresponding 95% confidence intervals on "k" as determined from multiple regression analyses for each of the noise metrics. The metrics were SEL(A), SEL(C), SEL(U), LLZd, LLZf, PL, and PNL. As shown in Fig. 7, the calculated values of "k" ranged from 10 to 15 depending on the noise metric considered. Based on energy addition theory, the predicted value of "k" is 10, as indicated by the dashed line. For all but two of the metrics, the 95% confidence in-

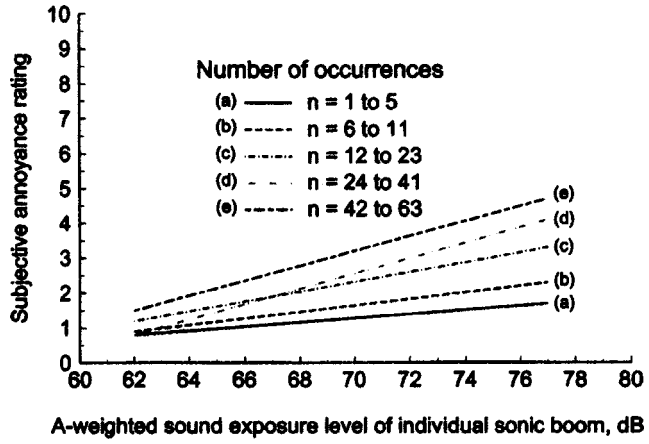


FIG. 6. Effect of number of sonic boom occurrences on annoyance versus A-weighted sound exposure level of individual sonic boom.

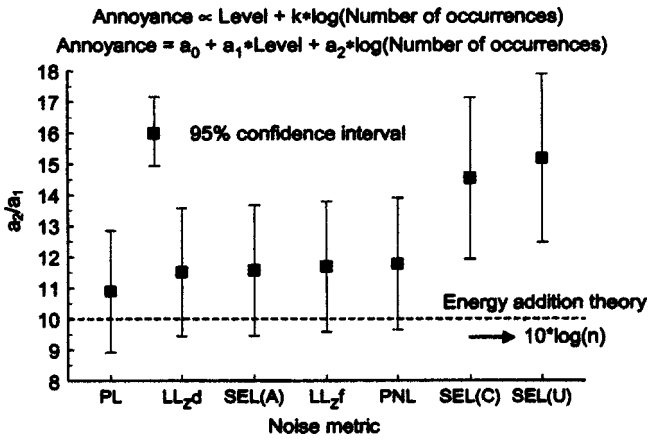


FIG. 7. Determination of "log(n)" coefficient in modeling the effect of number of sonic boom occurrences on annoyance.

TABLE III. Results of two-tailed Student's *t*-test for significance of difference between correlation coefficients of subjective annoyance response with noise metrics: * indicates significant at 0.05 level ($t_{1438}=1.960$), ** indicates significant at 0.01 level ($t_{1438}=2.576$), otherwise not significant.

Rank	Annoyance correlation coefficient	Noise metric procedure	*t*-statistic					
			PL	SEL(A)	PNL	LLZd	LLZf	SEL(C)
1	0.3495	PL	···					
2	0.3392	SEL(A)	3.166**	···				
3	0.3385	PNL	4.568**	0.157	···			
4	0.3288	LLZd	5.310**	1.930	3.411**	···		
5	0.3281	LLZf	5.303**	1.974*	3.604**	1.406	···	
6	0.3193	SEL(C)	4.626**	2.312*	3.378**	1.706	1.647	···
7	0.3150	SEL(U)	4.697**	2.580**	3.518**	2.157*	2.122*	2.874**

terval about the calculated value includes this value of 10. Therefore, in these cases, the value of "*k*" cannot, statistically, be said to be significantly different from 10.

The two metrics whose 95% confidence intervals on "*k*" did not include the value of 10 were examined further. The value of "*k*" was calculated from the regression coefficients a_1 and a_2 for each metric. Comparison of a_2, the regression coefficient for "log(*n*)," across all metrics found that the value of a_2 was constant across all metrics, indicating that the differences in "*k*" for SEL(U) and SEL(C) were due to the metric slope a_1 and not the effect of number of occurrences.

E. Comparison of noise metrics

Using the energy addition model for summing the effect of multiple sonic booms ($k=10$), the total daily sonic boom exposure was calculated in terms of each of the noise metrics for comparison with the test subjects' daily annoyance judgments. Daily exposure is commonly expressed in terms of day–night average sound level (DNL). DNL uses "10∗ log(number of occurrences)" to sum multiple events and then averages the noise energy across 24 h.[11,12] (Since no sonic booms were presented during the late night hours, no late night penalties were applied.) Although DNL is usually associated with A-weighted sound pressure level, for the purpose of this study, a DNL was calculated using each of the noise metrics.

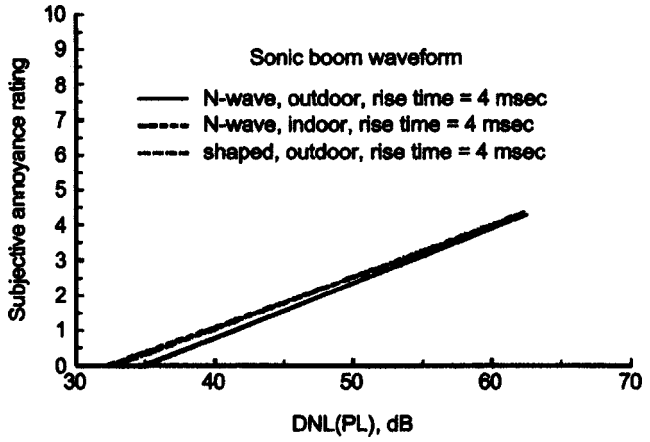


FIG. 8. Effect of sonic boom waveform shape on annoyance versus day–night average sound level based on the perceived level noise metric.

The predictive ability of the noise metrics was then compared based on the correlation coefficients between the subjective annoyance ratings and the different DNLs. The resulting rank order of metrics was (1) PL, (2) SEL(A), (3) PNL, (4) LLZd, (5) LLZf, (6) SEL(C), and (7) SEL(U). Perceived level was the best predictor of annoyance to the simulated sonic booms.

The differences in predictive ability were then assessed using a two-tailed Student's *t*-test for significance of difference (at the 0.05 level) between two correlation coefficients when samples are not independent.[14] The comparisons of the metrics, shown in Table III, yielded the following results of interest. PL was a significantly better predictor of annoyance than the other metrics. SEL(A) was significantly better than SEL(C). SEL(U) was significantly worse than the other metrics.

F. Effect of sonic boom waveform shape

Figure 8 shows the effect of sonic boom waveform on annoyance for day–night average sound level based on perceived level. The figure shows the regression line of subjective annoyance rating on DNL(PL) for each of the three sonic boom waveforms used in the study. Indicator (dummy) variable analyses[15] found no significant difference in either slope or intercept between the three different sonic boom waveforms. This result was the same for all the noise metrics considered.

G. Effect of startle

At the end of each day the test subject was asked if he or she was startled by any of the sonic booms during the day. Figure 9 shows the effect of startle on annoyance for day–night average sound level based on perceived level. The regression lines for both "startled" and "not startled" are shown for subjective annoyance rating plotted against DNL(PL). Indicator (dummy) variable analyses indicated a significant difference in both slope and intercept between annoyance ratings of the test subjects who were "startled" and those who were "not startled" by a sonic boom. Annoyance is greater when the test subject is "startled" and the magnitude of the increase in annoyance increases as sonic boom exposure increases. This effect was the same for all the noise metrics considered.

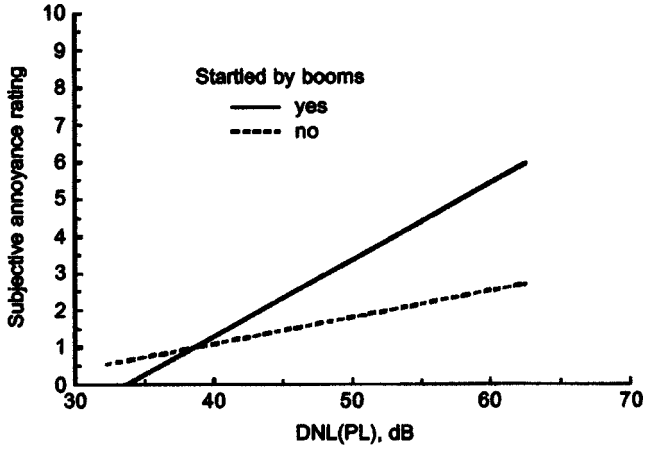


FIG. 9. Effect of startle on annoyance versus day-night average sound level based on the perceived level noise metric.

The daily exposures in which sonic booms were presented at just one noise level were used to determine how often startle occurred as a function of PL. At the end of each day, the test subjects were asked if they were startled by any of the sonic booms heard that day. Figure 10 shows the percentage of exposures causing startle as a function of the PL of the individual sonic boom repeated during the day. Each data point represents at least 20 daily exposures.

H. Effect of indoor ambient noise level on annoyance

Indoor ambient noise SEL(A) levels were collected 24 h a day from each of two microphones located in different rooms in each home. Ambient measurements were made between sonic boom presentations and at the end of each hour. This resulted in from 56 to 174 ambient measurements per day per home.

In order to reduce the complexity of the calculation of daily indoor ambient noise levels from this massive amount of data, only those ambient measurements representing at least 50 continuous minutes were included in calculating the daily ambient noise level. Furthermore, the ambient measurements were limited to the 14-h testing period during which the sonic booms occurred. This resulted in 623 exposure days for which a daily indoor ambient SEL(A) level was determined. The analyses of the effect of indoor ambient noise level were based on these 623 days, which did represent a wide range of the test conditions in the experiment design.

To determine if the daily indoor ambient SEL(A) level had a significant effect on annoyance, it was included in a multiple regression equation with DNL(PL). The resulting coefficient for the ambient level was not significantly different from zero. The multiple regression was repeated using the difference between the ambient level and the DNL(PL) level in place of the ambient level. Again the resulting coefficient for the ambient term was not significantly different from zero. Therefore, annoyance was not significantly affected by indoor ambient noise level.

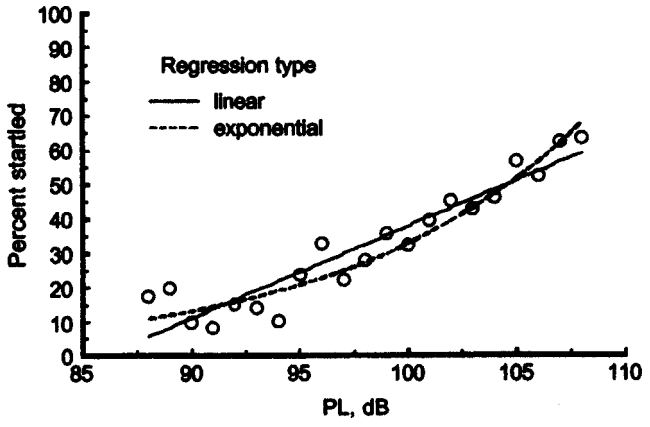


FIG. 10. Percentage of sonic boom exposures causing startle as a function of the perceived level of the individual sonic boom repeated during the day.

I. Effect of adaptation

As mentioned previously, the first 8 days of each testing period consisted of eight introductory exposures that were repeated on randomly selected days later in the testing period. These introductory exposures were included to determine if the test subjects adapted to the new noise environment during the first few days of the testing period.

In order to calculate an exposure level for a day in which more than one boom is heard, it is necessary to determine the effect of number of sonic booms occurrences. Based on energy addition theory and initial analyses of all the data collected (1673 test exposure days),[5,13] the effect of number was modeled by adding ''10∗log(number of occurrences)'' to the individual boom noise level. (The model is discussed and confirmed in a previous section.)

Since the number of booms actually heard each day by a test subject could vary, it was not possible to directly compare the response to an introductory day with the response to its repeated day. Therefore, indicator (dummy) variable regression analysis was used to compare the responses to the introductory exposures with the responses to their repeats. The exposures were defined in terms of SEL(A). Since the order of the 8 introductory days was varied across test subjects, each of the first 8 test days included each of the introductory exposures. For each test day, 1–8, the introductory exposures and subjective responses were compared with the repeated exposures and responses. The resulting regression equations for each day found no significant differences in slope between the regression models for the introductory exposures and the repeat exposures. Each day's difference in subjective response between the introductory and repeated exposures was determined from the regression equation (line intercepts) and converted into an equivalent SEL(A) difference using the slope (the SEL(A) coefficient) of the regression equation. These subjective differences are plotted against test day in Fig. 11. The regression analysis indicated that the introductory exposures were more annoying than the repeated exposures for each of the first 8 test days. The difference in annoyance was statistically significant for test days 1 and 3. Although the difference for the other 6 days was not significant, the trend in Fig. 11 is clear. The test subjects were more annoyed by a given sonic boom exposure occurring in the first few days of the test period than the same exposure occurring later in the test period.

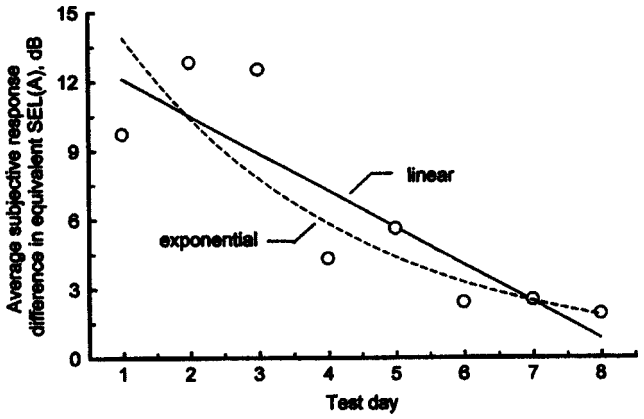


FIG. 11. Effect of adaptation on test subject annoyance.

J. Effect of activity

As part of the subjective response questions presented at the end of each day, the test subjects were asked to indicate, to the best of their memories, what type of activities they were engaged in and when during the 14-h test period. The six types of activities were defined as (1) ''sleeping, napping, resting,...;'' (2) ''talking with other people;'' (3) ''listening to TV, radio, music...;'' (4) ''playing...hobby, reading, etc.;'' (5) ''working...chores, job, etc.;'' and (6) ''other... .''

Subjects were instructed to choose the lowest numbered activity if engaged in two or more activities at the same time. Subjects were instructed to provide the information based on their recollections of the day and to not to keep a log during the day.

Since each daily annoyance response almost always covered more than one activity, the activities could not be compared directly. Instead the amount of time engaged in each activity each day was calculated and then included in a multiple regression analysis with DNL(PL). The regression coefficients for activities 1, 2, and 3 were significantly different from zero, indicating significant effects on the annoyance response (Fig. 12). Annoyance increased on average an equivalent of 1.5 dB per hour of activity related to sleeping, napping, or resting. Annoyance decreased on average an equivalent of 0.75 and 0.5 dB per hour of activity related to, respectively, talking with other people and listening to TV, radio, and music. The other activities had no significant effects on annoyance response.

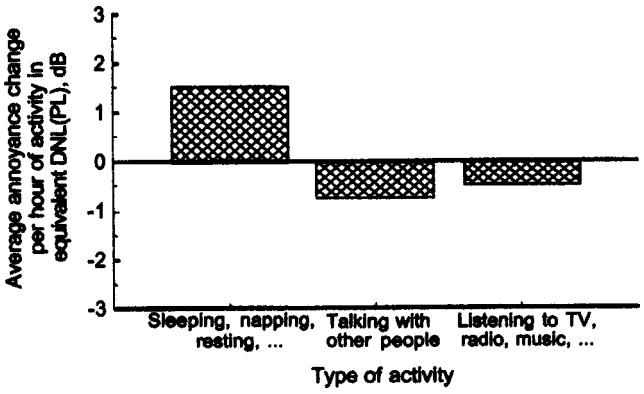


FIG. 12. Effect of activity on annoyance to sonic booms.

K. Effects of other test subject and household parameters

1. Gender

The 33 test subjects included 25 females and 8 males. No significant effect of gender on annoyance was found using indicator (dummy) variable analysis with DNL(PL) as the noise metric.

2. Age

The age of the test subjects at the beginning of the test period ranged from 32 to 71 with a mean of 50.6 and a median of 53. To determine if age had a significant effect on annoyance, it was included in a multiple regression analysis with DNL(PL). The resulting equation was

$$\text{Annoyance} = 0.141 * \text{DNL(PL)} + 0.034 * \text{Age} - 6.373 \quad (3)$$

Both coefficients and the constant were significantly different from zero. Comparing the two coefficients indicates that the increase in annoyance with one year of age was equivalent to an increase in DNL(PL) of 0.25 dB. (The addition of age to the regression equation raised its correlation coefficient from 0.350 to 0.374.)

3. Occupation

Of the 33 test subjects, 19 were homemakers, 9 were retired, and 5 either worked from their home or were temporarily unemployed. The effects on annoyance of this limited range of occupation categories were determined using indicator variable analysis with DNL(PL) as the noise metric. No significant effect on annoyance of occupation was found.

4. Spouse

Of the 33 test subjects, 26 had spouses living in the household. Indicator variable analysis using DNL(PL) as the noise metric found no significant effect on annoyance of the presence of a spouse.

5. Number of household members

Of the 33 households in the experiment, 3 had one member, 13 had two members, 5 had three members, 11 had four members, and 1 had five members. The number of people in the household was included in a multiple regression equation with DNL(PL). The resulting coefficient for the number of household members was not significantly different from zero. Therefore, annoyance was not significantly affected by the number of people living in a household.

6. Children

Eleven of the 33 households in the experiment included children under the age of 18. To determine if the presence of children affected annoyance response, it was included in an indicator variable analysis with DNL(PL). No interaction between DNL(PL) and the presence of children was found, but the presence of children did have a significant effect on the annoyance response. The resulting equation was

$$\text{Annoyance} = 0.149 * \text{DNL(PL)} - 0.817 * \text{Children} - 4.686, \quad (4)$$

where Children = 1 if children under 18 were members of the household and 0 otherwise. The effect of the presence of children on the test subjects' annoyance response was equivalent to a decrease in DNL(PL) of 5.5 dB. (The addition of children to the regression equation raised its correlation coefficient from 0.350 to 0.382.)

V. CONCLUDING REMARKS

Integrated sound exposure metrics such as equivalent continuous sound level and day–night average sound level are routinely applied to multiple occurrences of transportation noise sources such as road traffic and aircraft flyovers. The results of this experiment indicate that they are equally applicable to noise exposures consisting of infrequent, impulsive sounds such as sonic booms. In other words, multiple occurrences of a sonic boom can be modeled by the addition of "10 * log(number of occurrences)" to the level of the individual sonic boom. The implicit trade-off between the amplitude of individual events and their frequency of occurrence used in these integrated metrics, 3 dB per halving or doubling of number of events, is thus suitable for sonic boom acceptance criteria.

Of the several noise metric frequency-weighting procedures considered, perceived level was a significantly better predictor of annoyance. All of the metric procedures considered adequately accounted for the wide range of spectral content of the three sonic boom waveforms used in the experiment.

The experiment also provided information on a number of other questions, including startle, ambient level impact, activity effects, and demographic differences. Not surprisingly, startle was found to increase annoyance. The louder the sonic boom was, the greater the increase in annoyance caused by startle. Annoyance response did not appear to be affected by the indoor ambient noise level.

The results for different activities varied. Whether a subject was "working" or "playing" had no effect on their annoyance response to sonic booms. Annoyance increased if the test subject spent time sleeping or resting. Interestingly, talking with other people or listening to television, radio, or music appeared to distract the test subjects from, or mask the effects of, sonic boom exposure. Annoyance decreased if the test subjects were engaged in either communications pursuit.

Test subject gender did not affect annoyance to sonic booms. Neither did the presence of a spouse or the number of household members. Annoyance increased with test subject age at a rate equivalent to 0.25 dB per year. Annoyance decreased an equivalent of 5.5 dB if children were present in a household.

Two results of potential interest to future experiment designers are noted. First, the test subjects adapted to the new sonic boom noise environment during the first few days of testing, with annoyance decreasing. This adaptation effect should be taken into account in designing experiments of extended duration. Second, the experiment successfully validated the approach of creating a laboratory-style test apparatus in people's homes to conduct long term testing. This approach provided situational realism not available in the laboratory and a degree of control over the noise exposure not found in community surveys. Continuing improvements in computer, sound reproduction, and noise measurement technologies should facilitate future studies using this approach.

[1] P. Henne, "The case for small supersonic civil aircraft," AIAA Paper 2003–2555, presented at the American Institute of Aeronautics and Astronautics International Air and Space Symposium and Exposition: The Next 100 Years, Dayton, OH (July, 2003).

[2] K. D. Kryter, D. J. Johnson, and J. R. Young, "Psychological experiments on sonic booms conducted at Edwards Air Force Base," Contract AF 49 (638)-1758, Stanford Research Institute (August 1968).

[3] K. D. Kryter, "Laboratory tests of physiological-psychological reactions to sonic booms," J. Acoust. Soc. Am. **39**, Pt.2, S65–S72 (1966).

[4] A. Niedzwiecki and H. S. Ribner, "Subjective loudness of 'minimized' sonic boom waveforms," J. Acoust. Soc. Am. **64**, 1622–1626 (1978).

[5] J. D. Leatherwood, B. M. Sullivan, K. P. Shepherd, D. A. McCurdy, and S. A. Brown, "Summary of recent NASA studies of human response to sonic booms," J. Acoust. Soc. Am. **111**, 586–598 (2002).

[6] P. N. Borsky, "Community reactions to sonic booms in the Oklahoma City area," NORC Report No. 101, AMRL Report AMRL-TR 65-37, Wright Patterson Air Force Base, OH (1965).

[7] J. M. Fields, "Reactions of residents to long-term sonic boom noise environments," NASA Contractor Report 201704 (1997).

[8] J. E. Mabry and P. B. Oncley, "Establishing certification/design criteria for advanced supersonic aircraft utilizing acceptance, interference, and annoyance response to simulated sonic booms by persons in their homes," FAA-RD-75-44, Federal Aviation Administration (March 1975).

[9] J. D. Leatherwood and B. M. Sullivan, "Loudness and annoyance response to simulated outdoor and indoor sonic booms," NASA Tech. Memo. 107756, 1–36 (1993).

[10] K. P. Shepherd and B. M. Sullivan, "A loudness calculation procedure applied to shaped sonic booms," NASA Technical Paper 3134, 1–10 (1991).

[11] K. S. Pearsons and R. L. Bennett, "Handbook of noise ratings," NASA Contractor Report 2376 (1974).

[12] R. L. Bennett and K. S. Pearsons, "Handbook of aircraft noise metrics," NASA Contractor Report 3406 (1981).

[13] D. A. McCurdy, S. A. Brown, and R. D. Hilliard, "The effects of simulated sonic booms on people in their homes," AIAA Paper 95-0834, pp. 1–8, presented at the American Institute of Aeronautics and Astronautics 33rd Aerospace Sciences Meeting, Reno, NV January, 1995.

[14] H. E. Kluge, *Statistics: The Essentials for Research* (Erlbaum, Hillsdale, NJ, 1986), pp. 260–266.

[15] N. R. Draper and H. Smith, *Applied Regression Analysis* (Wiley, New York, 1966), pp. 241–257.

Sound level distribution and scatter in proportionate spaces[a)]

Stephen Chiles and Mike Barron[b)]
Department of Architecture & Civil Engineering, University of Bath, Bath, BA2 7AY, England

(Received 13 February 2004; revised 1 June 2004; accepted 1 June 2004)

Measurements have been conducted in two scale models of proportionate spaces each containing measures to promote a diffuse sound field. Contrary to classical statistical theory, it was found that reflected sound levels decrease with the distance from the source. Measured levels follow behavior predicted by a revised theory previously proposed for concert halls. The scatter of the measured reflected sound levels is shown consistently to exceed values predicted by Lubman/Schroeder's theory. However, it is demonstrated that this theory does accurately predict scatter of late sound levels when the early reflections are omitted. Investigation into the scatter of reverberation times shows that it is generally well predicted by Davy's theory. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1775279]

PACS numbers: 43.55.Br, 43.55.Cs, 43.55.Mc, 43.55.Ka [NX] Pages: 1585–1595

I. INTRODUCTION

Practitioners regularly use classical statistical theory that predicts the same reflected sound level at all points in a room. For concert halls, measurement data[1,2] show this to be an oversimplification as the reflected levels significantly decrease with increasing distance from the source. From simple reasoning that the onset of reflected sound cannot occur prior to the arrival of the direct sound, Barron and Lee[2] developed a revised theory that predicts the average behavior of sound levels in concert halls more accurately. So far this revised theory has only been applied to auditoria, where there is a particular acoustic environment with virtually all the acoustic absorption (the audience) on one surface. Variations of revised theory have been proposed for application in reverberation rooms[3] and specific types of churches.[4,5]

A diffuse sound field is taken as the reference condition for room acoustics and a model investigation is reported here that tests whether this revised theory is applicable to proportionate spaces designed to have diffuse sound fields. Two commonly accepted ways to promote a diffuse sound field are nonparallel geometry and scattering surfaces. The results given below are for measurements in two physical acoustic scale models: one with nonparallel geometry and the other with scattering surfaces.

II. SOUND LEVEL IN ROOMS

A. Classical theory

Sound received in a space from a single source can be split into a direct and reflected component. The prediction of the direct component is common in all approaches and accepted as giving an accurate result. The following discussion considers only the reflected component, for which the various prediction methods give significantly different results. To obtain the total sound level by any of these methods would require the addition of the direct component to the expressions given below for the reflected component.

Equation (1) is the classical statistical expression for reflected sound energy,

$$\frac{\overline{p^2}}{\rho c} = \frac{4W}{A}, \tag{1}$$

where p is the sound pressure, ρ is the density of air, c is the speed of sound in air, W is the sound power of the source and A is the total acoustic absorption in metric Sabines. This equation does not account for receiver (or source) locations and it therefore predicts the same reflected sound level throughout a space. Equation (1) is usually derived through consideration of the balance of acoustic energy introduced into and absorbed in a room. An alternative derivation of reflected sound energy can be based on an image model, which predicts exponentially decaying sound in a space.[6] In this model, the reflected level in Eq. (1) is obtained by summing the energy arriving at a point, from the moment of direct sound leaving the source to infinite time. Implicit in this alternative derivation is that the onset of the reflected sound occurs simultaneously at all points throughout the space at the moment the direct sound is emitted.

B. Revised theory

Revised theory[2] can be derived by reasoning that, as reflected sound cannot arrive earlier than direct sound, the arrival time of the direct sound can also be taken as the onset time of the reflected sound at that location. During a sound decay the instantaneous sound level at any moment is still taken to be the same throughout the space and the sound is still taken to decay exponentially. Therefore, the later the delayed onset the lower the reflected sound level will be at a given point. This is illustrated in Fig. 1, which shows three idealized decay curves for receiver positions (A, B, and C) located progressively farther away from the source. The direct sound is emitted from the source at time $t=0$ and reaches each of the receivers at times t_A, t_B, and t_C, respec-

[a)]Portions of this work were presented in "Distribution of reflected levels in rooms with diffuse sound fields," Proceedings of the 18th International Congress on Acoustics, Kyoto, April 2004, and "Sound level distribution in rooms: is it really the same everywhere ?," Proceedings of the Institute of Acoustics, Oxford, November 2003.

[b)]Electronic mail: M.Barron@bath.ac.uk

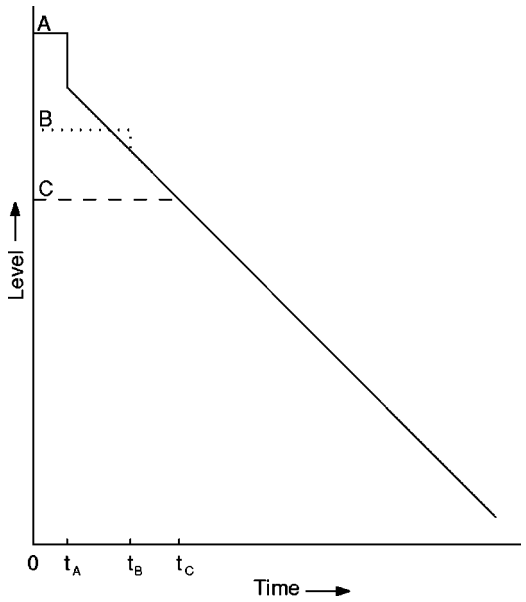


FIG. 1. Decay curves at three source–receiver distances according to revised theory. Here, $t=0$ is the time when the sound is emitted from the source and times t_A, t_B, and t_C are when the direct sound reaches receivers A, B, and C, respectively.

tively. At times later than the direct sound arrival at the farthest position, the decay curves all coincide as the instantaneous level is the same throughout the space. Moving away from the source, the direct sound becomes less prominent until at receiver C it is not visible in the decay trace. The image model using these revised integration times gives a revised value for reflected energy:

$$\frac{\overline{p^2}}{\rho c}=\left(\frac{4W}{A}\right)e^{-0.04r/T}, \tag{2}$$

where r is the source–receiver distance in meters and T is the reverberation time in seconds. Barron and Lee[2] examined the regression of measured sound levels against the source–receiver distance for 17 concert halls. From this measurement data it was apparent that in concert halls there is generally a linear correlation with distance and, extrapolating from the regression lines, Eq. (1) correctly predicts the reflected sound level close to the source. An analysis showed that the average behavior in concert halls was well described by Eq. (2).

For practical application, revised theory, Eq. (2), can be rewritten in terms of the reflected sound pressure level, L_{p_r}, and sound power level, L_W:

$$L_{p_r}=L_W+10\log(A)+6-0.174(r/T), \tag{3a}$$

or, using the Sabine equation,

$$L_{p_r}=L_W+10\log(T)-10\log(V)+14-0.174(r/T), \tag{3b}$$

where V is the room volume in m^3.

It can be seen that Eq. (2) differs from Eq. (1) by the factor $e^{-0.04r/T}$, which when expressed in decibels gives a decrease in the reflected sound level at a rate of 0.174/T dB/m as one moves away from the source. The maximum effect due to revised theory is thus a function of $r_{\max}/T$. If we take a cube of side z, the length of the diagonal is $\sqrt{3}z$, and $r_{\max}$ will be roughly 75% of this, namely $0.75\sqrt{3}z$. According to the Sabine equation $T=0.16V/S\overline{\alpha}=0.16z^3/6z^2\overline{\alpha}$, giving $0.174r_{\max}/T=\text{const.}\overline{\alpha}$, where S is the total area of the room surfaces in m^2 and $\overline{\alpha}$ is their average absorption coefficient. For spaces with the same proportion, the maximum revised theory effect is thus a function of $\overline{\alpha}$ alone, not size, for instance. It can be seen above that the value of the constant depends on the proportions of the space; for a cube it is 8.5, for a double cube 10.0, and for the models in this investigation it is approximately 9. This means that the maximum revised theory effect, $\Delta G_{r,\max}$, for these models is

$$\Delta G_{r,\max}\approx 9\overline{\alpha}\ \text{dB}. \tag{4}$$

For a maximum revised theory effect of, say, 3 dB, an average absorption coefficient of around 0.33 is required. With a small average absorption coefficient Eq. (2) gives very similar results to Eq. (1). For spaces such as reverberation chambers (typically $\overline{\alpha}=0.03$) there is little benefit in using Eq. (2) rather than Eq. (1), whereas for concert halls, which have a higher average absorption coefficient, there is a significant difference between the two results, by up to 3 dB at remote seats.

C. Alternative theories

Vorländer[3] proposed that, rather than starting the onset of reflected sound at the arrival of the direct sound, it should be started at the arrival of the first reflection. It can be shown that the average arrival time of the first reflection is the time for sound to travel the mean free path, 4 V/S. If this is substituted for r in Eq. (2) and the reverberation time, T, is expressed as 0.16 V/A, then the constants and the volume cancel out to leave

$$\frac{\overline{p^2}}{\rho c}=\left(\frac{4W}{A}\right)e^{-A/S}. \tag{5}$$

Vorländer went on to show that if the Eyring absorption exponent is used, ignoring air absorption so that $A=S\ln(1-\overline{\alpha})$, then Eq. (4) can be written as

$$\frac{\overline{p^2}}{\rho c}=\left(\frac{4W}{A}\right)(1-\overline{\alpha}). \tag{6}$$

This was proposed primarily for reverberation chambers where the measurement results are spatially averaged. Equation (6) is a well-established and commonly used modification of Eq. (1), notionally to account for a decrease in the reflected level due to sound absorbed at the first reflection, but there has been no robust derivation of this prior to Vorländer's work.

Several recent measurement exercises in religious buildings[4,5,7,8] also show reflected sound levels to generally decrease with an increasing source–receiver distance, contrary to classical theory, Eq. (1). However, the reflected levels in these religious buildings generally fall below those predicted by revised theory. In their study of Gothic-Mudejar churches Sendra *et al.*[4] introduce an adjustment parameter, ß, to Eq. (2), giving

$$\frac{\overline{p^2}}{\rho c} = \left(\frac{4W}{A}\right) e^{-\beta r/T}. \quad (7)$$

For each octave band, average values of ß are empirically derived from measurement data and Eq. (7) is proposed for general use in this specific type of church. During a study of Apulian Romanesque churches, Cirillo and Martellotta[5] theoretically derive an alternative adjustment to Eq. (2). Like Vorländer, they identify that the onset of reflected sound does not in reality start at the arrival time of the direct sound. Their solution is to assume that the delay is proportional to the source–receiver distance and derive a correction on this basis. They also derive a correction to account for the specific early reflection pattern. While these corrections do not rely on empirical data, their formulation makes many approximations and results in a complex set of expressions only tested for these Apulian Romanesque churches.

D. The present study

Revised theory, Eq. (2), does not account for the distribution of absorption or room geometry and defines receiver positions solely in terms of their distance from the source. These assumptions allow revised theory to retain most of the simplicity of the classical statistical formula, with no new parameters required (source–receiver distance is already needed for calculation of the direct sound level). As previously mentioned, it has been shown to be significantly more accurate than classical theory for concert halls. The question behind this investigation is whether revised theory can also give sufficiently robust predictions for other spaces and specifically in the reference condition of a proportionate space with a diffuse field. Alternative theories, illustrated by those presented above, tend to have a limited application and introduce such complexities that they represent a more detailed level of analysis that would perhaps be better handled by standard computer modeling packages.

III. MODELING

A. Measurement system

The use of physical scale modeling is an established technique in room acoustics.[9] For this investigation, key characteristics of the University of Bath's modeling system were reviewed: verification measurements were made of transducer directivity, spark source linearity, and air absorption corrections. In this investigation a greater accuracy is desired than is needed for say auditorium modeling. The microphone directivity and air absorption are discussed below. Measurements of the spark source showed it to have nonlinear propagation to a distance of approximately 100 mm. To ensure that measurements are only in the region of linear propagation, a minimum source–receiver distance of 200 mm (5 m full-size at a scale of 1:25) is used in this investigation. As shown previously by Barron,[10] the energy of the spark source over the frequency range used in this investigation increases at approximately $+9$ dB/octave. A first-order low pass filter (-6 dB/octave roll-off) is included in the analysis system to attenuate the high frequencies and effectively reduce the slope of the spark spectrum to approximately $+3$ dB/octave.

The modeling system uses two 1/8-in. microphones. To account for slight variations in the spark source level, the energy of the direct sound is measured with the second reference microphone. This also allows for situations where the direct sound is obscured to the first microphone. To reduce the effect of any variation in the spark source directivity, all results presented are average values from the individual analysis of three separate sparks. Variations of the spark source level/directivity are not significant in the results presented below. For both models, the spark source and two microphones are located at the bottom of adjustable height tubes held in randomly distributed sockets on the models' top surfaces. Spark discharges are directly recorded to obtain impulse responses, which are then stored and sequentially filtered in third-octave/octave bands. All normal room acoustics parameters are obtained using standard reverse integration, with a noise and tail correction to extend the dynamic range available. An iterative method[11] is used to determine the decay/noise cross-point and the decay is truncated at 5 dB above this point. The energy that would exist from a linear decay after this truncation point is added to the total energy of the truncated decay when calculating the reverse integrated response. The average noise energy is determined from the end section of the measurement and is subtracted from the pressure squared impulse response. Reverberation time T_{20} is obtained from the slope of the reverse integrated decay curve's regression line from 5 dB to 25 dB below the initial level.

The upper frequency used in the models is determined by the directivity of 1/8-in. microphones which, the manufacturer's data shows and the verification measurements confirmed, are omnidirectional to within ± 1 dB up to the 16 kHz third-octave band. The spark source was designed to have maximum energy at a frequency slightly above 16 kHz, so that its spectrum changes at a constant rate ($+9$ dB/octave) over the measurement range. The lower frequency limit (5 kHz third-octave band) is determined by the energy from a single spark that still maintains a signal-to-noise ratio of at least 25 dB. In the context of this investigation the absolute full-size equivalent frequencies are not critical. For convenience, a scale factor of 1:25 was chosen, since the models then represent realistic spaces and it gives results in a familiar frequency range: 250 Hz and 500 Hz octave bands. To avoid confusion, the later results and discussion are always for the full-size equivalent frequencies and distances.

To reduce air absorption at high frequencies, previous modeling at the University of Bath has been conducted in dry air or nitrogen. However for this investigation, the relatively large models make these options less practical. The measurements presented here were made in air with numerical corrections for absorption according to ISO 9613-1.[12] The validity of the correction procedure used is demonstrated in Fig. 2, which shows sample decays from measurements that were conducted in air and then repeated in nitrogen. The graph shows the same measured decay in air with

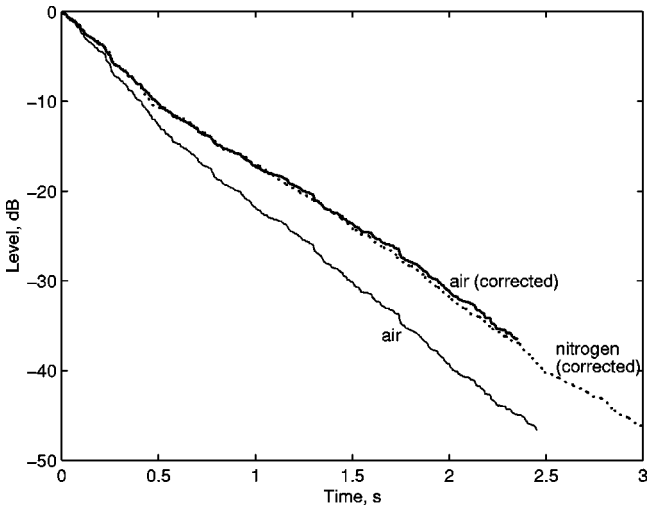


FIG. 2. Reverse integrated sound level decay curves for the same source–receiver combination when the model is filled with air (solid lines) and filled with nitrogen (dotted line). The decay in air is shown uncorrected for air absorption and corrected using the equations from ISO 9613-1 to standard atmospheric conditions: 20 °C, 50% R.H., 1013.25 mb. The decay in nitrogen has been corrected for air absorption but the magnitude of the correction is minimal.

and without correction for air absorption. When the correction has been applied the decay in air can be seen to be close to the decay measured in nitrogen. The decay in nitrogen has also been corrected for air absorption but the adjustment is very minor. These graphs illustrate that a penalty of measuring in air rather than nitrogen (or dry air) is that there is a significant loss of dynamic range in the corrected decay.

B. Model constructions

Figures 3 and 4 show the two models used in this investigation, which are both of similar size with the longest dimension just over 1 m actual size. Model 1 has flat surfaces and three of the walls are angled by 5° (one from each pair of opposite walls). Model 2 is rectangular and has highly irregular surfaces comprising scattering hemicylinders and deep absorptive devices. Certain ratios of dimensions for rectangular spaces encourage a good distribution of modes and hence promote a diffuse sound field.[13] Two of these ratios are the basis for the two models: 1/0.7/0.59 (Model 1), 1/0.83/0.65 (Model 2). In Model 1, the ratio is only used to obtain a reasonable starting point as the walls are angled to promote the uniform distribution of modes. In Model 2 there are not simple internal dimensions, due to the depth of the irregular surfaces, and there is some deviation from the chosen ratio. Model 1 is constructed from 9 mm thick, melamine faced, medium density fiberboard. The melamine facing gives a smooth, acoustically reflective finish. Model 2 is constructed from 12 mm thick plywood but the internal surfaces are completely covered with the absorbing and scattering finishes described below.

FIG. 3. Photograph of Model 1. The rear, left, and bottom surfaces are angled at 5° from parallel with their opposites. The surfaces are acoustically reflective other than the squares seen in the photograph, which are a velvet material providing acoustic absorption. The front surface is transparent acrylic. Transducers are suspended from the top surface.

FIG. 4. Photograph of Model 2 during construction. The black hemicylinders are 75 mm diameter rainwater guttering used to provide a scattering surface. In the completed model the wells between the guttering are covered with a velvet material.

From Eq. (4) it can be seen that revised theory only predicts a significant difference (>1 dB) from classical theory, at the maximum source–receiver distance, when the average absorption coefficient, $\overline{\alpha}$, in a proportionate space is greater than about 0.1 ($\overline{\alpha}$ is typically 0.3 in concert halls). For the purposes of testing revised theory a reasonable amount of absorption is therefore included in both models. This absorption is evenly distributed on all surfaces to further promote a diffuse sound field. Figure 3 shows Model 1

TABLE I. Dimensions and average measured reverberation times of Models 1 and 2. The dimensions are given as the full-size equivalent values for the scale of 1:25. Neither of the models are simple cuboids and the linear dimensions given are the maximum values in the case of Model 1 and the values inside the scattering/absorbing surfaces in Model 2. The volume for Model 1 is that calculated by the acoustic computer model and for Model 2 it is the volume inside the scattering/absorbing surfaces.

	Full-size equivalent dimensions				Rev. Time	
	Volume	Length	Width	Height	250 Hz	500 Hz
Model 1	9798 m^3	30.0 m	17.7 m	21.0 m	3.39 s	1.88 s
Model 2	9453 m^3	26.2 m	16.7 m	21.7 m	1.24 s	1.32 s

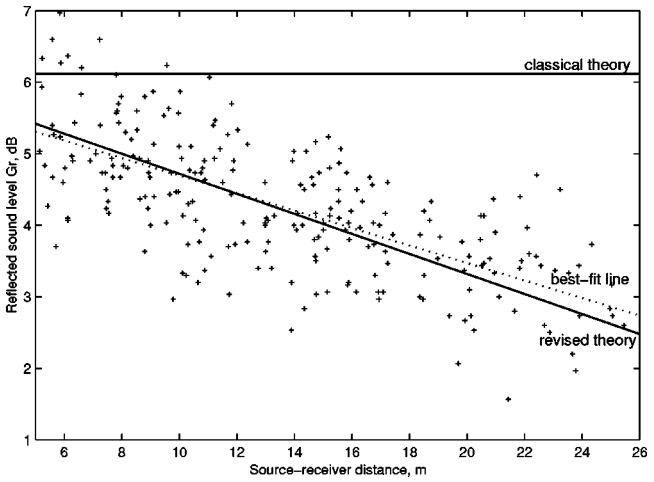


FIG. 5. Reflected sound level measurements plotted against the source–receiver distance for the 500 Hz octave band in Model 2. Lines are shown for the theoretical values of G_r according to classical and revised theories and also the best-fit line through the measurement data.

with 200 mm square patches of absorbing velvet material. The average absorption coefficient in Model 1 is 0.16 to 0.29 at the measurement frequencies. Figure 4 shows Model 2 during construction; when completed, the 40 mm deep wells, between the black hemicylinders, are covered by strips of velvet material as porous absorbers. The average absorption coefficient in Model 2 is 0.42 to 0.45. Previous measurements of hemicylinders[14] show the scattering coefficient, plotted against frequency, to have a small value at low frequencies and then rise steeply to a fluctuating plateau. The frequency at which the steep rise in scattering coefficient occurs is inversely proportional to the radius of the hemicylinders. For Model 2, the hemicylinders were selected on the basis of a practical material (rainwater guttering) with a radius large enough that the steep rise in scattering coefficients occurs below the measurement frequencies. Full-size equivalent dimensions of the models and the average measured reverberation times are given in Table I.

IV. SOUND LEVEL DISTRIBUTION

A large number of measurements were made in both models. In room acoustics the parameter G is defined as the total sound level relative to the direct sound at 10 m from the source.[15] This investigation is concerned only with reflected sound and not the total sound which includes direct sound. The parameter G is modified here by excluding the first 5 ms, which contains the direct sound, resulting in just the reflected sound level, denoted G_r,

$$G_r = 10 \log_{10} \frac{\int_5^\infty p^2(t)dt}{\int_0^5 p_{10}^2(t)dt}, \tag{8}$$

where p is the pressure at the receiver position and p_{10} is the pressure that would occur at a position 10 m from the source. Measurements were taken at 156 source–receiver pairs in Model 1 and 240 source–receiver pairs in Model 2. Figure 5 shows the measured G_r values plotted against distance for the 500 Hz octave band in Model 2. There are no measurements closer than 5 m to the spark source to avoid nonlinear propagation. Results in Model 1 and at 250 Hz in Model 2 have the same overall trends seen in Fig. 5. For both models and octave band frequencies, statistics of the regression lines compared to the revised theory predictions are given in Table II.

From these results it appears that revised theory is substantially more accurate than classical theory for predicting sound levels in proportionate spaces with diffuse sound fields. The scatter of sound levels about the revised theory prediction line, seen in Fig. 5, is discussed in Sec. VI. For each model and each frequency band there is clear evidence that the reflected sound level decreases with the distance from the source. Although the best-fit lines through the data often vary slightly from revised theory, no systematic difference was identified and the mean differences are small. Within 95% confidence limits, the differences between the slopes and intercepts of the best-fit lines and the revised theory predictions are not statistically significant except in the case of the slopes of the lines in Model 2. These slopes fall outside the 95% confidence limits but the difference is not of practical significance and would not be of statistical significance were it not for the large number of data points (240). The close agreement between measured levels and revised theory is an important result with major implications. For many practical spaces, where acousticians are currently using classical theory to predict sound levels, there is now a firm basis for the adoption of revised theory.

V. SOUND LEVEL DECAY CURVES

A further test of revised theory can be made by examining superimposed decay curves measured in the models. According to revised theory, the instantaneous level in a space

TABLE II. Statistics for sound level measurements and predictions in Models 1 and 2. For measurements, the statistics are taken from the linear regression line of reflected sound levels against the source–receiver distance and for predictions the mean level and slope are taken from the revised theory line against the source–receiver distance. The mean levels are the reflected sound level at the mean measurement distance, which happens to be the same in both models, 13.4 m.

		Mean level (dB)			Slope (dB/m)		
		Measured	Predicted	Difference	Measured	Predicted	Difference
Model 1	250 Hz	9.7	9.7	0.0	−0.05	−0.05	0.00
	500 Hz	6.9	6.6	−0.3	−0.08	−0.09	0.01
Model 2	250 Hz	4.2	4.6	0.4	−0.17	−0.13	−0.04
	500 Hz	4.3	4.2	0.0	−0.12	−0.14	0.02

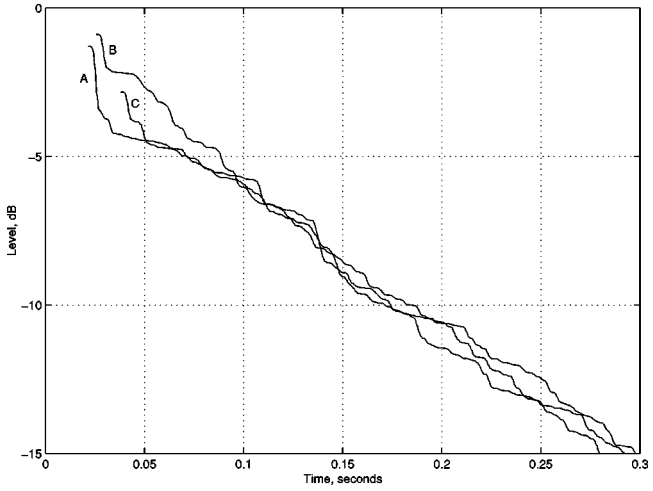


FIG. 6. Initial part of the reverse integrated (Schroeder) decays for three receiver positions. The moment when sound was emitted from the source is $t=0$ on the time axis.

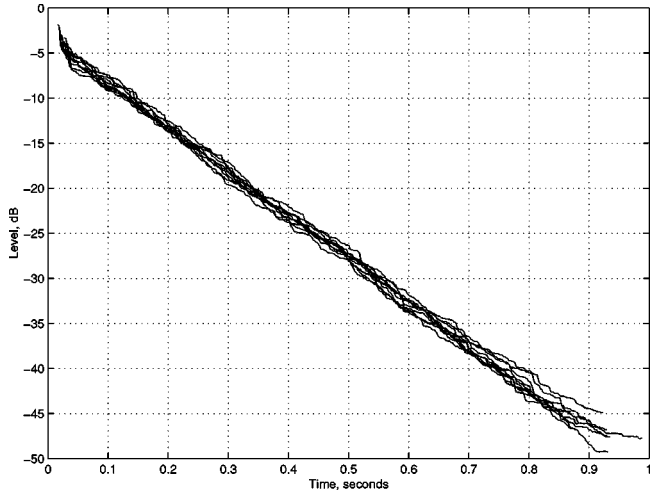


FIG. 7. Reverse integrated decays for ten receiver positions with the same source position in Model 2 at the 500 Hz octave band. The moment when sound was emitted from the source is $t=0$ on the time axis.

is the same at all positions during a sound decay. Therefore, as shown in Fig. 1, the measured sound level decay curves should coincide if they are each plotted with $t=0$ set as the time the source pulse is emitted. Normally individual decay curves are displayed starting at 0 dB. However, to compare different curves here, the magnitude of the start points are determined from the total measured sound level, G, at each position.

Figure 6 shows the initial part of three sound decay curves for the 500 Hz octave band in Model 2 at three receiver locations and the same source position. The three curves start at different times according to their source–receiver distances (7.5 m, 8.8 m and 13.0 m) and different levels in proportion to the measured total sound levels. Curve B has several strong early reflections visible as steps in the decay curve from the start point up to around 100 ms. Curve A has fewer early reflections which leaves the initial step of the direct sound more prominent. Due to these differences in early reflections, curve B has a slightly higher total level than curve A despite being marginally farther away from the source. However, although curve B is above curve A up to around 100 ms, they then converge, showing the later reflected sound to be uniform throughout the space. Position C is at a greater distance from the source than A or B and, as predicted by revised theory, has a correspondingly lower total level, although it converges with the other curves after the early reflections. The fact that the three curves converge is further evidence in support of revised theory. The differences in total levels show that generally the total levels decrease with an increasing source–receiver distance (curve C is lower than curves A and B) but there are minor fluctuations within this pattern (curve B is slightly higher than curve A).

Figure 7 shows the full decay curves measured for ten receiver positions with the same source position. These curves have been time and level aligned as described above. For most of the decay, the curves are reasonably close although towards the end the curves diverge. Investigations have shown this divergence at the end to be caused primarily by inaccuracies inherent in noise/tail correction to extend the measurement dynamic range (without any noise/tail correction the divergence would be greater). In the region where the curves are reasonably closely aligned, there is still some variation seen in Fig. 7. This has been quantified by the standard deviation of the decay curves' instantaneous level at each time step, for each model, source position and frequency band. In all cases, the standard deviation of the curves is greater at the start of the decay, due to direct sound and early reflections illustrated in Fig. 6, and at the end of the decay, due to the noise/tail correction. In between these regions, the average standard deviation in Model 1 is 0.65 dB for the 250 Hz octave band and 0.59 dB at 500 Hz. The values in Model 2 are 0.74 dB (250 Hz) and 0.57 dB (500 Hz). For Model 2, these values are very close to the theoretical standard deviations of reflected sound levels discussed below in Sec. VI. For Model 1, the values are higher than the theoretical standard deviations by 0.15–0.19 dB, which may be due to an insufficiently diffuse sound field, as discussed in Sec. VIII.

VI. SOUND LEVEL SCATTER

For the measured data, despite the close agreement of the average behavior with revised theory, it can be seen in Fig. 5 that there is significant scatter of the results around the best-fit line. The same is true at both frequency bands in both models and is partly explained by the theoretical fluctuation of sound levels within a space. This was investigated many years ago by Lubman[16] and Schroeder,[17] who both arrived at expressions that led to the following approximate equation for the standard deviation of sound levels measured in a room:

$$s_{G_r} = \frac{4.34}{\sqrt{\left(1+\left(\frac{BT}{6.9}\right)\right)}}\,dB, \tag{9}$$

where B is the bandwidth and T is the reverberation time. The derivation for this theoretical standard deviation starts

TABLE III. The scatter of reflected sound levels, measured as the standard error taken from the linear regression of the measured levels against the source–receiver distance, and predicted as the standard deviation using Eq. (9).

		Measured Standard Error	Predicted Standard Dev.
Model 1	250 Hz	0.58 dB	0.46 dB
	500 Hz	0.63 dB	0.44 dB
Model 2	250 Hz	0.99 dB	0.74 dB
	500 Hz	0.70 dB	0.54 dB

with Eq. (1) and the assumption that there are only *random* variations in level with the receiver position. The magnitude of the random variations is obtained by considering the steady state frequency response for a room to be a complex Gaussian process and the magnitude squared of the frequency response to be exponentially distributed.[17] The theory was originally only considered in the context of reverberation rooms. In such spaces it has been shown in Sec. II that the low average absorption coefficient means that revised theory would predict a negligible level variation with the source–receiver distance (the same result as classical theory). However, in the current investigation the standard deviation would also include the intrinsic variation of sound levels with the source–receiver distance, which was not accounted for in the derivation of Eq. (9). An alternative measure of scatter adopted here is to use the standard error taken from the linear regression of the measured reflected sound levels against the source–receiver distance. The use of standard error in place of standard deviation does not have a theoretical derivation, but in the absence of an equation for sound level scatter that accounts for the sound level decrease with distance, Eq. (9) used in this way gives useful values for comparison.

The data in Table III show that the standard errors for all measurements are consistently higher than the theoretical standard deviations. To explore the validity of Eq. (9) further, it is necessary to use data with a wider range of the two variables: reverberation time and bandwidth. This data was available from measurements made during the preparation of Model 1. The absorbing patches were originally added in three stages, with measurements made at each stage, giving results for a wide range of reverberation times. The last of these stages with 39 patches of absorption was the final condition for the main testing of Model 1. These impulse responses have been analyzed for the two octave bands at 250 Hz and 500 Hz and the six third-octave bands from 200 Hz to 630 Hz, giving a range of bandwidths. The standard errors for these measured sound levels are plotted in Fig. 8 against the predicted standard deviation. Figure 8 shows a general agreement between the measured standard error and predicted standard deviation, although it also highlights the consistent deviation at the two octave bands used in this study: the 250 Hz and 500 Hz octave bands (on the left of each set of points).

The theoretical model, Eq. (9), originated for reverberation rooms with low average absorption, whereas the spaces studied here have average absorption coefficients similar to

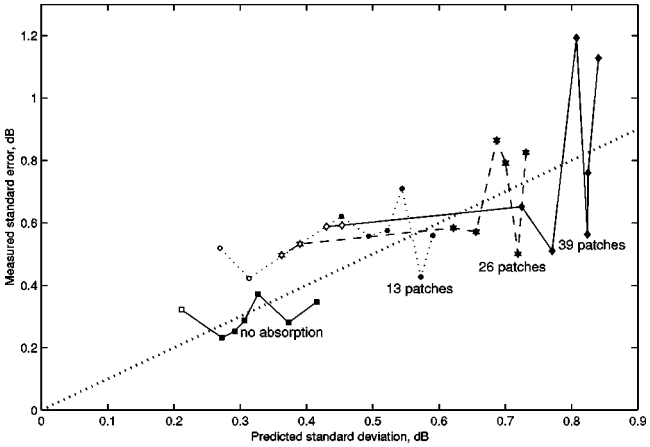


FIG. 8. Measured standard error plotted against the predicted standard deviation of reflected sound levels in Model 1 with four different amounts of absorption. In each set of data, the two left hand points with hollow symbols are for the 500 Hz and 250 Hz octave bands and the other six points to the right with solid symbols are for the third-octave bands from 630 Hz to 200 Hz. There is an exception for the set of data on the left, ''no absorption,'' where there is no 500 Hz octave band point. The straight dotted line is for equal predicted and measured values.

''real world'' rooms. However, even in a reverberation room, Lubman's measured standard deviations of sound level[18] consistently exceed the theoretical values by a similar percentage to that found in this investigation. The significant discrepancies found with standard deviations/standard errors of measured steady-state reflected sound levels indicate that the theory does not provide a reliable method for predicting this quantity. However, there is good agreement with average standard deviations of instantaneous late sound levels, presented above in Sec. V for Model 2.

The measured values for the late sound in Sec. V were obtained by calculating the average standard deviation of instantaneous sound levels for all receiver positions at each time step in the middle portion of the decays. As expected, very similar values are obtained in Model 2 by calculating the standard error against the source–receiver distance for the total late sound levels after 200 ms from the arrival of the direct sound. All these values are given in Table IV and show good agreement between the two different forms of analysis used and the theoretical predictions. For both analysis techniques, the choice of time window over which the statistics were obtained was determined experimentally. It can be seen in Fig. 6 that three of the decay curves converge after approximately 100 ms and it was found through an examination of statistics for 240 curves that they had all converged

TABLE IV. The scatter of late sound levels in Model 2. The theoretical standard deviation is from Eq. (9). The measured decays have been analyzed taking the average standard deviation, SD, of instantaneous sound levels during the middle portion of the decay and the standard error, SE, for correlation of the total late sound after 200 ms against the source–receiver distance.

	250 Hz	500 Hz
Theoretical Standard Deviation	0.74 dB	0.54 dB
Average SD of instantaneous levels	0.74 dB	0.57 dB
SE of late sound levels after 200 ms	0.79 dB	0.54 dB

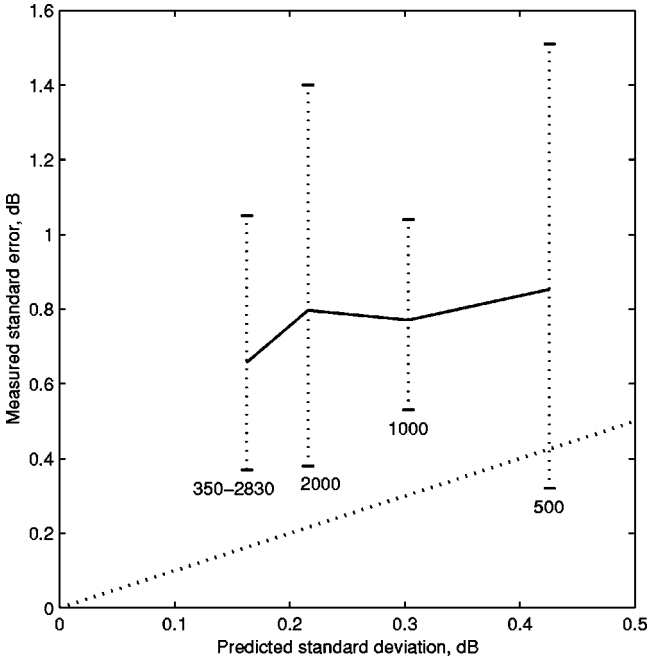


FIG. 9. Average measured standard error plotted against the predicted standard deviation of reflected sound levels in 16 concert halls. Error bars show the maximum and minimum values. The three right hand points are the average standard errors in the 2 kHz, 1 kHz, and 500 Hz octave bands and the left-hand point is the standard error for the average of these three bands. For simplicity, the reverberation time for the theoretical values at these three frequency bands has been taken as 2 seconds for all the halls. The straight dotted line is for equal predicted and measured values.

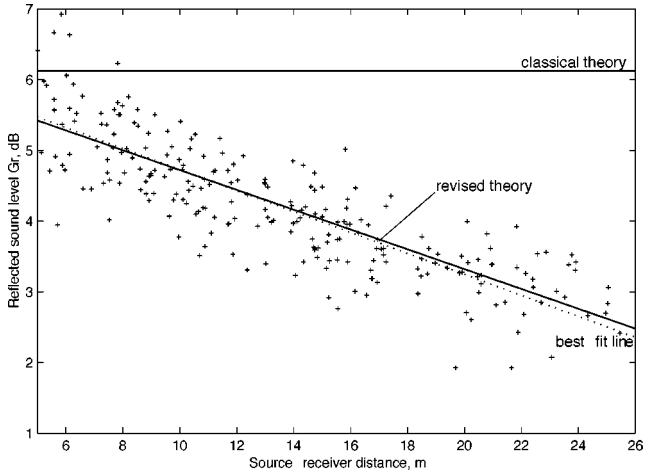


FIG. 10. Reflected sound level according to computer prediction plotted against the source–receiver distance for the 500 Hz octave band in Model 2. Lines are shown for the theoretical values of G_r according to classical and revised theories and also the best-fit line through the measurement data. The computer predictions in this graph are for the same scenario as Fig. 5 for the scale model measurements.

by around 200 ms for the 250 Hz octave band and 150 ms for the 500 Hz octave band. The statistics are not unduly sensitive to slight variations in these timings and, for the 500 Hz octave band, the standard error of the total late level after 150 ms is only 0.01 dB different from the value given for 200 ms. At the other end of the time windows, the average standard deviation of instantaneous levels did not include values after the decay curves started to diverge but the total late levels include the entire decays after 200 ms. This does not however, significantly affect the results as the end of the decays contribute little to the total late levels.

The good agreement between measurements and predictions seen in Table IV indicates that Eq. (9) does accurately predict the scatter of sound levels in the decaying sound field. It can be seen from examination of the decay curves that, on average, they do coincide for reflected sound, although there is always some variation of the instantaneous levels. It has been found that the magnitude of this variation is consistent over the length of the decays, apart from the initial section, which is controlled by the early reflections (divergence at the end of the curves is attributed to measurement/analysis issues and not the physical decays). This investigation has shown that the theory presented, Eq. (9), does accurately predict the variations during the main part of the decay but does not account for the variation of early energy. The total reflected sound level in most spaces is highly dependent on early reflections, thus making the theory limited in potential applications.

For reference, Fig. 9 shows the mean and range of standard errors of reflected sound levels measured at audience positions in 16 concert halls for the 500 Hz, 1000 Hz, and 2000 Hz octave bands. The left-hand bar in Fig. 9 is the standard error of the reflected sound levels averaged over these three octave-bands (350–2830 Hz). Concert halls, having a very uneven distribution of absorption (concentrated on the floor) and complicated geometries, could be expected to have a less uniform sound field than Models 1 and 2, and this is borne out by the high standard errors seen in Fig. 9 relative to predictions according to Eq. (9).

VII. COMPUTER MODELING

In parallel with the scale modeling, the two model spaces have been analyzed in a computer simulation (CATT-Acoustic). The simulation results for the 500 Hz octave band in Model 2 are shown in Fig. 10, which is equivalent to Fig. 5 for the scale model. For the predictions presented in Fig. 10 a scattering coefficient of 0.5 was used for all surfaces in the computer model. However, as the absorption in the model is evenly distributed the scattering coefficient has negligible effect on the results.

The overall trend of decreasing sound levels with increasing source–receiver distance in Fig. 10 is very similar to that seen in Fig. 5, although the slope of the decrease is higher for the computer simulation. Table V lists the scatter of sound levels measured in the scale model and predicted by the computer simulation for Model 1. It can be seen that there is significantly less scatter of reflected sound levels in

TABLE V. Scatter of reflected sound levels in Model 1 for 52 receiver positions and a single source position.

	250 Hz	500 Hz
Theoretical Standard Deviation	0.46 dB	0.44 dB
Standard Errors:		
Scale model measurements	0.56 dB	0.56 dB
Computer model predictions	0.19 dB	0.46 dB
Computer model with phase	0.71 dB	0.70 dB

TABLE VI. Relative standard deviations of reverberation times (T_{20}) measured in the two models and predicted using Eq. (10).

		Measured	Predicted
Model 1	250 Hz	6.4%	3.7%
	500 Hz	5.3%	3.5%
Model 2	250 Hz	5.7%	6.0%
	500 Hz	3.8%	4.3%

the computer simulation than for the scale model measurements, with a standard error even lower than the theoretical standard deviation at 250 Hz. The reason for the lesser scatter was assumed to be due to the exclusion of phase in the computer modeling. To test this hypothesis, data for one source position and 52 receivers in Model 1 were reanalyzed using the output from an auralization module, which includes an approximation for phase information. To enable realistic auralization, this module uses a synthesized phase and maintains the correct reflection density throughout the decay. To compute in a reasonable time, late energy is not processed as discrete reflections and for this late energy the application of the synthesized phase is approximate.

Table V shows that when the phase approximation is included it does significantly increase the reflected sound level scatter and demonstrates that the original lower scatter is due to the omission of phase in the standard calculations. However, the scatter with the phase approximation is slightly higher than the measured scatter due to the technique used to process the late sound discussed above. An examination of the correlation between computer predictions and scale model measurements has shown that the inclusion of phase approximations in this computer model does not appear to improve the accuracy of the individual sound level prediction at a particular receiver. Also, as seen in Table V, the omission of phase is only significant at the lower frequency, whereas most users of routine computer models are primarily interested in results at the midfrequencies.

VIII. REVERBERATION TIME SCATTER

Initially it was suspected that a reason for the measured sound level scatter being higher than predicted could be due to the sound field being insufficiently diffuse. A possible measure of sound field diffusion is the scatter of the reverberation time, expressed as the relative standard deviation. Davy *et al.*[19,20] derived a theoretical value of this quantity, which for impulse response measurements of a 20 dB decay is

$$\frac{\sigma(T_{20})}{\overline{T_{20}}} = \frac{0.96}{\sqrt{BT}}. \tag{10}$$

In this equation, the bandwidth, B should be the statistical bandwidth rather than the nominal or effective bandwidth.[21] The filters used in this investigation are 5th order Butterworth and the statistical bandwidths for them are obtained by increasing the nominal bandwidths by 11%.[22]

Table VI shows the measured and predicted relative standard deviations of reverberation times for the two models. It can be seen that for Model 2 there is reasonable agreement between measurements and predictions, as has been found previously in reverberation rooms by Bartel and Magrab,[23] Flenner *et al.*,[24] and Warnock.[25] The slight overestimation by Eq. (10) has been previously reported by Davy.[21] However, for Model 1 the measured values are significantly larger than predictions. Table VII lists standard deviations for Model 1 as the absorptive finishes were progressively introduced. This shows that the relative standard deviation of reverberation times is following the behavior predicted by Eq. (10) in that it increases with decreasing reverberation time but the magnitude is consistently greater than predictions. It is thought that Model 1 has an insufficiently diffuse sound field with excessive spatial variations, which have resulted in the discrepancies between measurements and predictions. This is discussed further later.

TABLE VII. Relative standard deviations of reverberation times (T_{20}) measured in and predicted for Model 1 as absorbing patches are progressively introduced.

	250 Hz		500 Hz	
	Measured	Predicted	Measured	Predicted
No absorption	2.9%	1.8%	–	–
13 patches	2.8%	2.6%	2.6%	2.3%
26 patches	3.5%	3.2%	3.4%	3.1%
39 patches	4.7%	3.7%	4.3%	3.6%

Theoretically, the values with all 39 absorptive patches in Table VII should be the same as the values given in Table VI, whereas it can be seen that there is a significant difference. Different source and receiver locations were used for the measurements summarized in the two tables and the data in Table VII comes from a comparatively small number of measurements (12 source–receiver pairs).

Bartel and Magrab[23] found that rotating diffusing panels significantly reduced the scatter of reverberation time, whereas fixed diffusing panels had little effect. However, Balachandran and Robinson[26] show that fixed diffusers can substantially improve sound field diffusion. Sixteen fixed plastic diffusers (80 mm by 120 mm) were suspended in Model 1 and Table VIII lists the measured scatter of reverberation time and sound levels. The data in Table VIII without the diffusers should be the same as Tables VI and VII but, again, different source–receiver pairs have been used and there are significant differences. The addition of the diffusers reduces the scatter of reverberation time to slightly below the predicted values.

An illustration of the results presented in Table VIII can

TABLE VIII. Measurements in Model 1 with and without diffusing panels. Results are for the relative standard deviation of the reverberation time and the standard error of the reflected sound level.

	250 Hz	500 Hz
Relative S.D. of T_{20}:		
Without diffusers	5.7%	4.3%
With diffusers	3.3%	2.9%
Standard error of G_r:		
Without diffusers	0.52 dB	0.55 dB
With diffusers	0.52 dB	0.54 dB

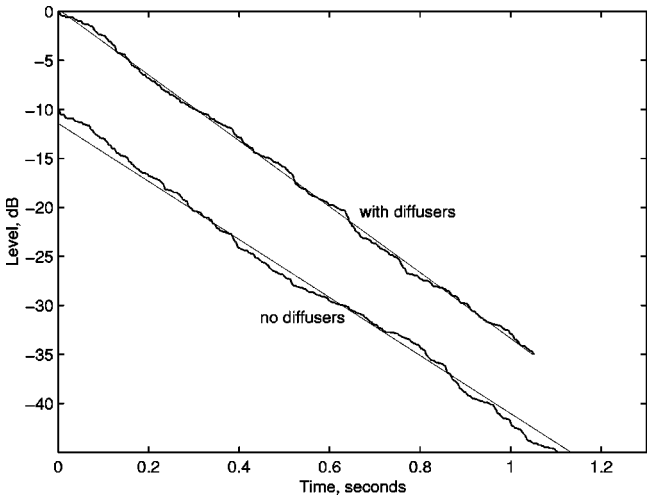


FIG. 11. A decay curve in Model 1 with and without diffusers. For ease of viewing, the curve for the model without diffusers is offset by −10 dB. The straight lines are the best-fit lines determined over the range −5 dB to −25 dB for each decay curve.

be made by plotting single decay curves. Figure 11 shows two decay curves for the same source–receiver pair, from the measurements summarized in Table VIII, with and without the diffusers present. For clarity, the curve without diffusers is offset by −10 dB. The figure also shows the best-fit lines determined over the range −5 dB to −25 dB, as used for determining the reverberation time. Without diffusers the decay curve is less linear with greater deviations from the best-fit line (the standard error from the best-fit line between −5 dB and −25 dB is 0.59 dB without diffusers and 0.37 dB with diffusers).

Therefore, for this particular source–receiver pair the automated calculation of reverberation time using these best-fit lines will be less accurate without the diffusers. This effect is considered to be the cause of the higher scatter of reverberation times without diffusers seen in Table VIII. This is also thought to be the reason for the scatter of reverberation time being above the predicted values for Model 1 in Tables VI and VII.

The consideration of reverberation time scatter is presented here as part of the investigation of sound level scatter. However, Table VIII shows that even when the reverberation time scatter significantly changes following the addition of diffusers, the scatter of sound levels remains largely unaffected. A similar result was found by Bartel and Magrab[21] who show that rotating diffusers substantially reduce the normalized standard deviation of reverberation time in the decaying sound field but have little effect on the scatter of steady-state sound levels.

IX. CONCLUSIONS

Revised theory has previously been shown to predict reflected sound levels in concert halls significantly more accurately than classical theory. Measurements of sound levels have been made in two scale models of proportionate spaces. To promote a diffuse sound field in these spaces, one model has nonparallel geometry and the other has acoustically scattering surfaces. Both models contain absorptive finishes to give average absorption coefficients ranging from 0.16 to 0.45. From these measurements it has been shown in all conditions tested that reflected sound levels decrease with distance from the source and revised theory is significantly more accurate than classical theory. The revised theory predictions are close to the best-fit lines through the data and it is concluded that it can be used for a calculation of the average behavior in proportionate spaces with diffuse sound fields. The magnitude of the revised theory effect is a function of the average absorption coefficient and the effect is therefore negligible in spaces such as a reverberation chambers which have very low average absorption coefficients, even when there is a highly absorbent sample on the floor.

The scatter of reflected sound levels measured in this investigation is significant. For spaces containing a moderate quantity of absorption, where the reflected sound level is correlated with the source–receiver distance, it is appropriate to use the standard error as a measure of scatter, rather than standard deviation. The scatter measured in the models is of a similar order to theoretical predictions but is consistently greater than them. However, it has been demonstrated that the theory does accurately predict the scatter of late sound levels for the part of the decays after the early reflections. Development of the theory is required to include the greater variation in reflected sound levels due to early reflections. A computer model analysis of the physical models gave a reduced scatter of predicted sound levels at lower frequencies. The use of a phase approximation in the computer model increased the scatter to a more realistic level but did not appear to improve the accuracy of individual values.

For these proportionate spaces, the measured scatter of reverberation time agrees well with predictions using theory proposed by Davy. An examination of this quantity in the two models suggested that the sound field in the model with nonparallel geometry was not as diffuse as in the model with scattering surfaces. The scatter of reverberation time in the model with nonparallel geometry was reduced by the addition of diffusing panels but this did not affect the scatter of sound levels.

ACKNOWLEDGMENTS

We are grateful for the assistance we have received from several people but in particular we would like to thank Michael Vorländer, Finn Jacobsen for drawing our attention to the previous work on sound level scatter, and Bengt-Inge Dalenbäck for his kind collaboration with the computer modeling and review of this paper. This research is funded by the UK Engineering and Physical Sciences Research Council.

[1] A.-C. Gade and J. H. Rindel, "Die Abstandsabhängigkeit vom Schallpegel in Konzertsälen," Fortschritte der Akustik-DAGA '85, 1985, pp. 434–438.
[2] M. Barron and L.-J. Lee, "Energy relations in concert auditoriums, 1," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[3] M. Vorländer, "Revised relation between the sound power and the average sound pressure level in rooms and consequences for acoustic measurements," Acustica **81**, 332–343 (1995).
[4] J. J. Sendra, T. Zamarreño, and J. Navarro, "An analytical model for evaluating the sound field in Gothic-Mudejar churches," in *Computational Acoustic and its Environmental Applications II*, edited by C. A. Brebbia *et al.* (Computational Mechanics Publications, Southampton, 1997), pp. 139–148.

[5] E. Cirillo and F. Martellotta, ''An improved model to predict energy-based acoustic parameters in Apulian-Romanesque churches,'' Appl. Acoust. **64**, 1–23 (2003).
[6] M. Barron, ''Growth and decay of sound intensity in rooms according to some formulae of geometric acoustics theory,'' J. Sound Vib. **27**, 183–196 (1973).
[7] A. Magrini and P. Ricciardi, ''Churches as auditoria: Analysis of acoustical parameters for a better understanding of sound quality,'' Building Acoust. **10**, 135–158 (2003).
[8] N. Prodi, M. Marsilio, and R. Pompoli, ''On the prediction of reverberation time and strength in mosques,'' in *Proceedings of the 17th ICA*, Rome, 2001.
[9] M. Barron, ''Acoustic scale model testing over 21 years,'' Acoust. Bull. **22**, 5–12 (1997).
[10] M. Barron, ''Impulse testing techniques for auditoria,'' Appl. Acoust. **17**, 165–181 (1984).
[11] A. Lundeby, T. E. Vigran, H. Bietz, and M. Vorländer, ''Uncertainties of measurements in room acoustics,'' Acustica **81**, 344–355 (1995).
[12] ISO 9613-1:1993, ''Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere,'' International Organization for Standardization, Geneva, Switzerland, 1993.
[13] L. W. Sepmeyer, ''Computed frequency and angular distribution of the normal modes of vibration in rectangular rooms,'' J. Acoust. Soc. Am. **37**, 413–423 (1965).
[14] M. Vorländer and E. Mommertz, ''Definition and measurement of random-incidence scattering coefficients,'' Appl. Acoust. **60**, 187–199 (2000).
[15] ISO 3382:1997. ''Acoustics-Measurement of reverberation time of rooms with reference to other acoustical parameters,'' International Organization for Standardization, Geneva, Switzerland, 1997.
[16] D. Lubman, ''Fluctuations of sound with position in a reverberant room,'' J. Acoust. Soc. Am. **44**, 1491–1502 (1968).
[17] M. R. Schroeder, ''Effect of frequency and space averaging on the transmission responses of multimode media,'' J. Acoust. Soc. Am. **46**, 277–283 (1969).
[18] D. Lubman, ''Precision of reverberant sound power measurements,'' J. Acoust. Soc. Am. **56**, 523–533 (1974).
[19] J. L. Davy, I. P. Dunn, and P. Dubout, ''The variance of decay rates in reverberation rooms,'' Acustica **43**, 12–25 (1979).
[20] J. L. Davy, ''The variance of impulse decays,'' Acustica **44**, 51–56 (1980).
[21] J. L. Davy, ''The variance of decay rates at low frequencies,'' Appl. Acoust. **23**, 63–79 (1988).
[22] J. L. Davy and I. P. Dunn, ''The statistical bandwidth of Butterworth filters,'' J. Sound Vib. **115**, 539–549 (1987).
[23] T. E. Bartel and E. B. Magrab, ''Studies on the spatial variation of decaying sound fields,'' J. Acoust. Soc. Am. **63**, 1841–1850 (1978).
[24] J. P. Flenner, J. P. Guilhot, and C. Legros, ''Dispersion des measures de la durée de réverbération d'un local,'' Acustica **50**, 201–208 (1982).
[25] A. C. C. Warnock, ''Some practical aspects of absorption measurements in reverberation rooms,'' J. Acoust. Soc. Am. **74**, 1422–1432 (1983).
[26] C. G. Balachandran and D. W. Robinson, ''Diffusion of the decaying sound field,'' Acustica **19**, 245–257 (1967/68).

Use of statistical energy analysis for the prediction of Sabine absorption coefficient of modally reactive panels

K. S. Sum[a)]
School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009, Australia

(Received 30 January 2004; revised 11 June 2004; accepted 19 June 2004)

Statistical energy analysis (SEA) is used to establish a relationship between the Sabine absorption coefficient of a modally reactive panel and the coupling of the panel with the sound field in a room. The acoustic-structural coupling is described by the coupling loss factors between the panel and the room. The application of the consistency relationship in SEA, which is based on the reciprocity principle, is shown to enable the coupling loss factors to be determined by directly exciting either the panel or the room. Each way of driving the acoustic-structural coupled system produces different values of the coupling loss factors and thus, of the absorption coefficient. The nature of discrepancy between the results of both ways of exciting the system is explained by using distributions of energy among structural modes of the panel and acoustic modes of the sound field, as well as distributions of modal coupling-loss factor among acoustic-structural coupled pairs of the modes. This paper provides an insight into the effect of the use of the consistency relationship on the prediction of the Sabine absorption coefficient of modally reactive panels by SEA. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1780571]

PACS numbers: 43.55.Br, 43.55.Dt, 43.55.Ev [NX] Pages: 1596–1606

I. INTRODUCTION

It is known that the sound absorption by modally reactive surfaces in a room depends on the coupling between the surfaces and the sound field in the room.[1] In the low-frequency range where acoustic modes of the uncoupled sound field are sparsely distributed, the absorption is described by the change in the resonance frequency and decay time of the modes due to the coupling.[2,3] Above this range, the practical size of rooms does not allow a modal description of the sound absorption to be applied because of a high modal density of the uncoupled sound field that causes an analysis of the coupling of individual acoustic modes to become impractical. Besides, a huge amount of details of such a large number of modes can also easily obscure the physical picture of behavior of the absorption. In this case, the absorption is amenable to an average modal description, which can be accomplished by statistical energy analysis (SEA).

In architectural acoustics, reverberation time is one of the main parameters that describes the sound quality in rooms. The average effect of an absorptive surface on the sound field in a room is usually quantified by the change in reverberation time of the sound field due to the surface absorption. For modally reactive surfaces, it is therefore important to relate the reverberation time to the coupling of the surfaces with the sound field. More than 20 years ago, SEA was used to analytically derive a relationship between the reverberation time and the coupling of two dynamic subsystems.[4,5] Instead of a modal description, these early works enable the effect of the coupling on the reverberation time to be described by using average modal properties of the uncoupled subsystems (i.e., internal loss factors and modal densities) and average modal-coupling quantities (i.e., coupling loss factors).

The literature on SEA is very extensive and the application of the method is not only broad in the areas of structural acoustics, dynamics, and vibration but also in room acoustics. SEA is known to have many implicit assumptions and these are summarized, for example, in Refs. 6 and 7. In the original development of the method, cross terms between modes of the subsystems in a general complex system were ignored and they contain information of the coherency between the modes.[8] The negligence of the cross terms allows a linear summation of power-flow terms in the energy-balance equations of the subsystems. Such negligence is justified when the driving force of the system is not a single-point excitation but a distributed and an uncorrelated excitation (i.e., implies statistical independence of the modes).[9] In the case of the former, the statistical independence can only be achieved from an ensemble-average result with respect to the point-excitation and response positions of the system. It was shown that consecutive point excitations at a minimum of three randomly chosen positions satisfy the requirement of the statistical independence.[10]

However, one of the main aspects of SEA is the introduction of the reciprocity principle into the method.[11] For a given pair of coupled subsystems, this principle provides a simple relationship between the modal densities of the subsystems in their uncoupled states and the two coupling loss factors of the pair (i.e., the consistency relationship). The relationship has often been considered as an integral part of SEA because if one of the coupling loss factors of the subsystems is known, then the other can be conveniently obtained once the modal densities are known (e.g., Refs. 12–15).

Nevertheless, two significant assumptions arise from the

[a)]Electronic mail: ksum@mech.uwa.edu.au

 © 2004 Acoustical Society of America

use of modal density in the consistency relationship. First, the power flow between a given pair of coupled subsystems is controlled only by their uncoupled modes whose resonance frequencies lie in the frequency band under consideration.[7,11] These will be called resonance modes. Second, all the resonance modes of a given subsystem are affected by the same amount of coupling and in the coupled state, there is also an equal distribution of energy among the modes.[7,11] As a result of these strict assumptions, the consistency relationship can be a major source of uncertainty in the application of SEA when one or both of the subsystems only have a few or no resonance modes in the band of interest. In other words, many resonance modes are required to minimize the adverse effect that results from the use of the relationship, because the lack of these modes can lead to three consequences. First, the total contribution of these modes to the power flow between the subsystems becomes weaker and thus, the influence of nonresonance modes (i.e., uncoupled modes whose resonance frequencies lie outside the band) becomes important. Second, the variation of energy between coupled resonance modes would be more pronounced and it is due to the fact that the coupling is very selective.[16] Finally, a large difference of modal density between the uncoupled and the coupled states may also occur especially with the presence of some low-order modes whose changes in resonance frequency due to coupling are significant.[17] So, a given mode does not necessarily lie in the same band for both states and thus, the modal density in the uncoupled state is prone to a large misprediction of that in the coupled state. Conversely, the misprediction is negligible when the band has many resonance modes. The criterion of a minimum of five resonance modes was often used[15,18] but four resonance modes have been suggested to be sufficient.[10]

Sabine absorption coefficient, α_{Sab}, is a parameter that describes the ability of an absorptive surface to dissipate sound and it is definable only within Sabine equation. Previous experimental data[19] indicated that α_{Sab} can be interpreted as the ratio of absorbed to incident energy of a diffuse sound field only if the surface absorption is small and the reverberation time of the sound field is long. Thus, this condition is required by various standards so that the definition of sound absorption coefficient holds for α_{Sab}. However, this interpretation is invalid for α_{Sab}, and α_{Sab} has different physical meanings if the surface absorption is not small and the reverberation time is short such that the sound field is not diffuse. There were many works which suggested that the reverberation time of a general sound field can be expressed in the form of Sabine equation. Early analytical results[20] showed that α_{Sab} in a general sound field can be interpreted as a nondimensional average half-power bandwidth of acoustic modes in the band of interest or a normalized modal-overlap factor of the sound field. Specifically, α_{Sab} of a locally reactive surface can be interpreted in terms of the acoustical impedance of the surface and the type of individual acoustic modes that dominate the sound-field decay.[21] On the other hand, α_{Sab} of a modally reactive surface can be interpreted in terms of the internal and the coupling loss factors of the sound field and the surface, where α_{Sab} not only accounts for the sound absorption but also the sound radiation of the vibrating surface.[15]

SEA has been employed for the estimation of reverberation time of the sound field and α_{Sab} of a modally reactive surface in an acoustic-structural coupled system.[15] But, unlike the uncoupled sound field whose modal density increases rapidly with frequency, the requirement of four or five resonance modes in a band is not readily met for the uncoupled surface. Below the high-frequency range, the measured and the SEA-predicted reverberation times were observed to agree well in some bands but differ considerably in other bands.[15] Although the observation was attributed to an insufficient number of modes of the surface and a nonuniform sound field, it was unclear how such conditions affected the irregularity of the prediction and how the consistency relationship contributed to the irregularity. Consequently, the use of SEA in the estimation of α_{Sab} of the modally reactive surface under different modal properties of the acoustic-structural coupled system was still not well understood. In order for SEA to be useful in the prediction of α_{Sab}, the uncertainty associated with the application of the consistency relationship must first be resolved. Therefore, it is necessary to investigate the extent of the influence of the relationship on the predictability of α_{Sab} by SEA. Good predictability is defined as the ability of the method to consistently provide accurate results when modal properties of the coupled system are altered.

In this paper, a weighted-average Sabine absorption coefficient of a modally reactive panel that couples with an enclosed sound field is defined in terms of the damping and the coupling loss factors of the acoustic-structural coupled system by using SEA. Mathematical descriptions are provided to show that these coupling loss factors can be determined in two different ways when the consistency relationship is applied, namely, by directly exciting either the panel or the sound field. The similarity and difference in the coupling loss factors obtained by both ways of driving the system, and the consequence on the predictability of the absorption coefficient, are illustrated by computational and experimental examples. The extent of the influence of the consistency relationship on the predictability is then investigated. Also demonstrated is the improvement in the estimation of the coefficient in terms of accuracy when the relationship is not used in SEA.

II. SEA FORMULATION FOR THE RESPONSE OF AN ACOUSTIC-STRUCTURAL COUPLED SYSTEM

A. Quasitransient solution to the sound-field response

Consider the coupling between a modally reactive panel at the boundary of an enclosure and a general sound field in the enclosure. When the panel and the sound field are simultaneously driven by time-dependent input excitations, $\Pi_p(t)$ and $\Pi_a(t)$, respectively, the SEA quasitransient energy-balance equations for the acoustic-structural coupled system can be written as

$$\frac{\partial \tilde{E}_p(t)}{\partial t} = \Pi_p(t) + \eta_{ap}\omega_0 \tilde{E}_a(t) - (\eta_p + \eta_{pa})\omega_0 \tilde{E}_p(t), \tag{1}$$

$$\frac{\partial \tilde{E}_a(t)}{\partial t} = \Pi_a(t) + \eta_{pa}\omega_0 \tilde{E}_p(t) - (\eta_a + \eta_{ap})\omega_0 \tilde{E}_a(t). \tag{2}$$

$\tilde{E}_p$ and $\tilde{E}_a$ are time-dependent energies of the panel and the sound field, ω_0 is the center frequency of the excitation band, and t is time. η_a denotes the damping loss factor of the uncoupled sound field, which includes losses at other parts of the boundary of the enclosure as well as sound absorption by air and any objects inside the enclosure. η_p denotes the damping loss factor of the uncoupled panel, and it represents the combined loss due to internal structural damping and other dissipations at the joints of the panel, and sound radiation from the panel surface opposite to the sound field. η_{pa} is the coupling loss factor from the panel to the sound field (i.e., also called the radiation loss factor of the panel) and η_{ap} is the coupling loss factor from the sound field to the panel.

By considering $\Pi_p(t)$ and $\Pi_a(t)$ to be steady-state excitations that operate for sufficiently long during $t<0$, the coupled system would reach a steady state before $t=0$. Then, the input excitations are turned off at $t=0$ [i.e., $\Pi_p(t) = \Pi_a(t) = 0$ for $t>0$] when the system is in the steady state. By using this initial condition and applying Laplace transformation to Eqs. (1) and (2), the decay of the sound field is obtained as[15]

$$\tilde{E}_a(t) = E_a[(d_1 - s_1)e^{-s_1 t} - (d_1 - s_2)e^{-s_2 t}]/(s_2 - s_1), \tag{3}$$

$$s_1 = 0.5\omega_0[\eta_a + \eta_{ap} + \eta_p + \eta_{pa} - \sqrt{(\eta_a + \eta_{ap} - \eta_p - \eta_{pa})^2 + 4\eta_{ap}\eta_{pa}}], \tag{4}$$

$$s_2 = 0.5\omega_0[\eta_a + \eta_{ap} + \eta_p + \eta_{pa} + \sqrt{(\eta_a + \eta_{ap} - \eta_p - \eta_{pa})^2 + 4\eta_{ap}\eta_{pa}}], \tag{5}$$

$$d_1 = \omega_0(\eta_p + \eta_{pa} + \eta_{pa}E_p/E_a). \tag{6}$$

In the above, E_p and E_a are initial energies of the panel and the sound field (i.e., at $t=0$).

Equation (3) shows that the coupled sound-field decay is controlled by two decay rates, s_1 and s_2. In general, as η_{pa} and η_{ap} are, respectively, small compared to η_p and η_a, while η_{pa} increases slowly with frequency[15,22,23] and η_{ap} decreases rapidly with frequency,[15] $4\eta_{ap}\eta_{pa}$ is small compared to $(\eta_a + \eta_{ap} - \eta_p - \eta_{pa})^2$ in Eqs. (4) and (5). So, s_1 and s_2 can be mathematically approximated such that they are, respectively, given by the lower and the higher of both $(\eta_a + \eta_{ap} - C)\omega_0$ and $(\eta_p + \eta_{pa} + C)\omega_0$ where $C = \eta_{ap}\eta_{pa}/(\eta_p + \eta_{pa} - \eta_a - \eta_{ap})$. This means that $s_1 < s_2$ and both exponential terms shown in Eq. (3) are significant in the early part of the coupled sound-field decay. As time increases, only the term with a lower decay rate (i.e., $e^{-s_1 t}$) is dominant because the other term (i.e., $e^{-s_2 t}$) becomes negligible. Thus, only the early part of the decay is nonexponential since both terms are dominant.

B. Sabine absorption coefficient of the panel in a general sound field

When the enclosure has M absorptive surfaces including the panel, the mean Sabine absorption coefficient associated with the total dissipation in the sound field is given by[20] $\bar{\alpha}_{\text{Sab}} = 4V_0\eta_T\omega_0/A_T c_0$, where $\bar{\alpha}_{\text{Sab}} = (A_p\alpha_{\text{Sab}} + \Sigma_{i=1}^{M-1} A_i\alpha_{\text{Sab},i})/A_T + \alpha_{\text{air}}$. η_T is the total loss factor that corresponds to the total dissipation, V_0 is the volume of the enclosure, A_T is the total area of the surfaces, c_0 is the speed of sound in air, A_p and α_{Sab} are the surface area and Sabine absorption coefficient of the panel, A_i and $\alpha_{\text{Sab},i}$ are the area and Sabine absorption coefficient of the ith surface, and α_{air} corresponds to the air absorption. From the above, $\alpha_{\text{Sab}} = (4V_0\eta_T\omega_0/c_0 - \Sigma_{i=1}^{M-1} A_i\alpha_{\text{Sab},i} - A_T\alpha_{\text{air}})/A_p$. Equation (2) show that if the sound field is not coupled to the panel, then its decay only depends on η_a and is not affected by η_p, η_{pa}, and η_{ap}. By setting $\alpha_{\text{Sab}} = 0$ and $\eta_T = \eta_a$, $\Sigma_{i=1}^{M-1} A_i\alpha_{\text{Sab},i} + A_T\alpha_{\text{air}} = 4V_0\eta_a\omega_0/c_0$ and thus,

$$\alpha_{\text{Sab}} = 4V_0\omega_0(\eta_T - \eta_a)/A_p c_0. \tag{7}$$

Since the reverberation time of the coupled sound field is not analytically definable in the early decay as there are two dominant exponential terms[5] (i.e., $e^{-s_1 t}$ and $e^{-s_2 t}$), approximate Sabine absorption coefficients, α_1 and α_2, which correspond to the individual terms are determined first. If the decay rate of the coupled sound field in Eq. (7), $\eta_T\omega_0$, is deduced only from s_1, then $\alpha_1 = 4V_0(s_1 - \eta_a\omega_0)/A_p c_0$. When the decay rate is deduced only from s_2, $\alpha_2 = 4V_0(s_2 - \eta_a\omega_0)/A_p c_0$. Then, α_{Sab} is derived from an average of α_1 and α_2, which are weighted with respect to the amplitude and decay rate of both exponential terms shown in Eq. (3):

$$\alpha_{\text{Sab}} = \frac{|d_1 - s_1|e^{-s_1}\alpha_1 + |d_1 - s_2|e^{-s_2}\alpha_2}{|d_1 - s_1|e^{-s_1} + |d_1 - s_2|e^{-s_2}}. \tag{8}$$

In this way, the predicted α_{Sab} takes into account the full quasitransient solution of the coupled sound-field decay. Once η_p, η_a, η_{pa}, and η_{ap} are known, α_{Sab} can be evaluated by Eq. (8).

C. Determination of coupling loss factors

It can be seen from the preceding section that given η_p and η_a, the value of α_{Sab} depends on the prediction of η_{pa} and η_{ap}. However, regardless of the frequency range of interest, exact analytical expressions of η_{pa} and η_{ap} for any modal properties of the panel and the sound field are not easily obtained due to the complicated nature of acoustic-structural coupling. So, the coupling loss factors are usually obtained through the SEA steady-state energy-balance equations of the acoustic-structural coupled system using the information of steady-state energies of the system. These energies can be either evaluated by any well-established numerical or analytical methods, or measured experimentally. If the sound field is directly driven by a steady-state excitation and the panel is driven only through the coupling, then the energy-balance equation of the panel gives

$$\eta_{ap} = (\eta_p + \eta_{pa})E_p^a/E_a^a. \tag{9}$$

Conversely, when the panel is directly driven by a steady-state excitation and the sound field is driven only through the coupling, the energy-balance equation of the sound field yields

$$\eta_{pa}=(\eta_a+\eta_{ap})E_a^p/E_p^p. \tag{10}$$

In Eqs. (9) and (10), E_p^a and E_p^p are steady-state energies of the coupled panel, and E_a^a and E_a^p are steady-state energies of the coupled sound field, where superscripts a and p refer to the direct excitation of the sound field and the panel, respectively. By solving both equations, η_{ap} and η_{pa} can be determined if the values of the energies, η_p, and η_a are available.

When modal densities of the uncoupled panel and the uncoupled sound field (i.e., n_p and n_a) are known, it is not necessary to simultaneously use Eqs. (9) and (10). In this case, either one of the equations can be used with the consistency relationship (i.e., $n_p\eta_{pa}=n_a\eta_{ap}$) but this results in two different ways of evaluating the coupling loss factors (i.e., by directly exciting either the sound field or the panel). For the respective direct excitations, η_{pa}^a and η_{ap}^a, and η_{pa}^p and η_{ap}^p, are employed in the following to denote the approximations of η_{pa} and η_{ap} attributed to the use of the consistency relationship. The corresponding approximations of α_{Sab} are, respectively, denoted as α_{Sab}^a and α_{Sab}^p. By using the consistency relationship in Eq. (9), it can be shown that

$$\eta_{pa}^a=\eta_p R_{pa}^a/(1-R_{pa}^a), \tag{11}$$

$$R_{pa}^a=\frac{E_p^a/n_p}{E_a^a/n_a}, \tag{12}$$

and thus,

$$\eta_{ap}^a=(\eta_p+\eta_{pa}^a)E_p^a/E_a^a. \tag{13}$$

On the other hand, when the consistency relationship is substituted into Eq. (10), the results are

$$\eta_{ap}^p=\eta_a R_{ap}^p/(1-R_{ap}^p), \tag{14}$$

$$R_{ap}^p=\frac{E_a^p/n_a}{E_p^p/n_p}, \tag{15}$$

$$\eta_{pa}^p=(\eta_a+\eta_{ap}^p)E_a^p/E_p^p. \tag{16}$$

R_{pa}^a is the modal-energy ratio of the panel to the sound field for the direct sound-field excitation and R_{ap}^p is the modal-energy ratio of the sound field to the panel for the direct panel excitation.

For the direct sound-field excitation, Eq. (11) indicates that η_{pa}^a is directly related to R_{pa}^a so it directly reflects the prediction error generated by the consistency relationship. However, Eq. (13) suggests that η_{ap}^a is affected less by the relationship because it is proportional to $\eta_p+\eta_{pa}^a$ (i.e., total dissipation by the panel), where the error in η_{pa}^a is masked by η_p. On the contrary, Eq. (14) shows that for the direct panel excitation, η_{ap}^p is directly related to R_{ap}^p so it directly reflects the prediction error associated with the consistency relationship. But η_{pa}^p is affected less by this relationship since it is proportional to $\eta_a+\eta_{ap}^p$ (i.e., total dissipation by the sound field), where the error in η_{ap}^p is masked by η_a [see Eq. (16)]. Thus, the deviation of α_{Sab}^a and α_{Sab}^p from α_{Sab} depends on the variation of α_{Sab}^a with η_{pa}^a and η_{ap}^a, and α_{Sab}^p with η_{pa}^p and η_{ap}^p. If α_{Sab}^a and α_{Sab}^p are, respectively, more sensitive to η_{pa}^a and η_{pa}^p than η_{ap}^a and η_{ap}^p, then α_{Sab}^a would deviate more than α_{Sab}^p when the masking of the error in η_{ap}^p by η_a is large enough such that η_{pa}^p has a smaller difference than η_{pa}^a from η_{pa}. If α_{Sab}^a and α_{Sab}^p are, respectively, more sensitive to η_{ap}^a and η_{ap}^p than η_{pa}^a and η_{pa}^p, then α_{Sab}^a would deviate less than α_{Sab}^p when the masking of the error in η_{pa}^a by η_p is large enough such that η_{ap}^a has a smaller difference than η_{ap}^p from η_{ap}.

D. Modal energies and modal coupling-loss factors

In Eqs. (12) and (15), $E_p^a/n_p\Delta f$, $E_p^p/n_p\Delta f$, $E_a^a/n_a\Delta f$, and $E_a^p/n_a\Delta f$ represent the mean energies of structural modes of the coupled panel and acoustic modes of the coupled sound field in a band of width, Δf (hertz). Therefore, η_{pa}^a, η_{ap}^a, η_{pa}^p, and η_{ap}^p represent the mean modal coupling-loss factors of acoustic-structural coupled pairs of the resonance modes. The valid use of these mean energies in the evaluation of the mean coupling-loss factors is based on two requirements. First, the resonance modes are in control of the energies of the panel and the sound field, and n_p and n_a are nonzero. Otherwise, the mean modal energies become infinite if the modal densities are zero, and the mean modal coupling-loss factors are undefined. Second, the mean modal energy of the subsystem that is directly excited is greater than that of the subsystem which is driven only by acoustic-structural coupling. Otherwise, R_{pa}^a and/or R_{ap}^p are greater than unity, and Eqs. (11) and/or (14) give negative values of η_{pa}^a and/or η_{ap}^p that are invalid. Subsequently, the values of η_{ap}^a and/or η_{pa}^p as obtained by Eqs. (13) and/or (16), can also become negative. This leads to a necessity of equipartition of energy among the structural modes and among the acoustic modes, where the modal energies are assumed to be equal to the associated mean modal energies. As a result, the coupled pairs of the modes have equal distributions of both coupling loss factors from the structural modes to the acoustic modes and from the acoustic modes to the structural modes. These modal coupling-loss factors are assumed to be equal to the corresponding mean modal coupling-loss factors.

However, equal distributions of modal energy in both subsystems are not always achievable because the selective nature of the coupling causes the modal energy to vary greatly from one mode to another. In this case, the deviation from the states of equipartition of modal energy can be shown by using the distribution of energies of resonance modes of the subsystems. The extent of the deviation is statistically determined by the scatter of the modal energies or spread of their distributions. If E_{pj}^a, E_{pj}^p, E_{ai}^a, and E_{ai}^p denote the energies of the jth structural mode and the ith acoustic mode and, respectively, replace $E_p^a/n_p\Delta f$, $E_p^p/n_p\Delta f$, $E_a^a/n_a\Delta f$, and $E_a^p/n_a\Delta f$, then Eqs. (11)–(16) become

$$\eta_{pa,ji}^a=\eta_p R_{pa,ji}^a/(1-R_{pa,ji}^a), \tag{17}$$

$$R_{pa,ji}^a=E_{pj}^a/E_{ai}^a, \tag{18}$$

$$\eta_{ap,ij}^a=(\eta_p+\eta_{pa,ji}^a)E_p^a/E_a^a, \tag{19}$$

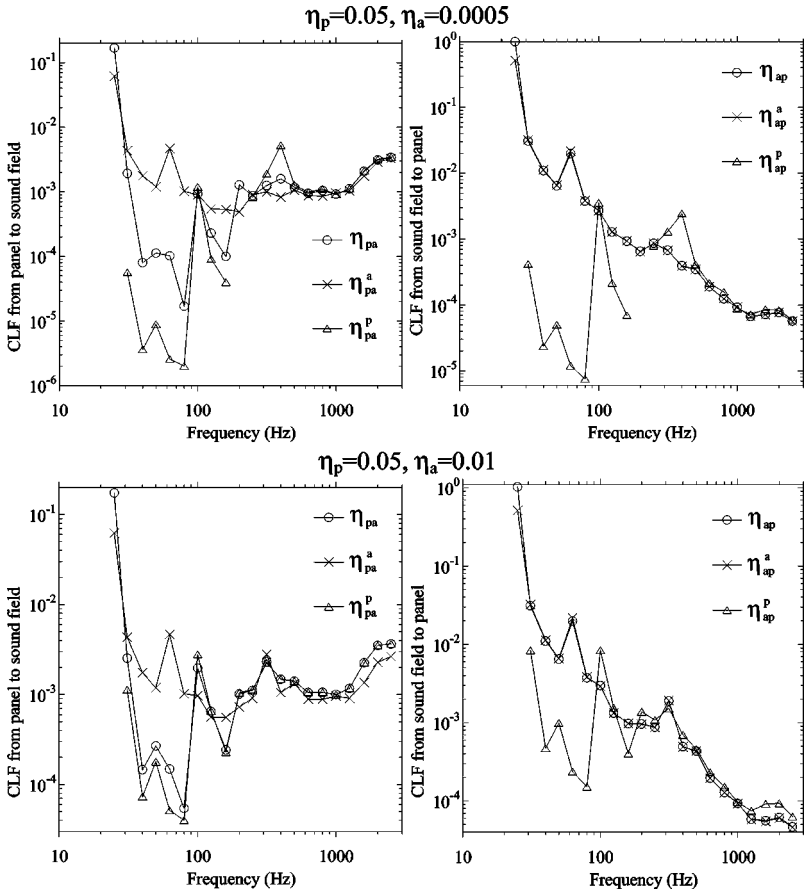


FIG. 1. Coupling loss factors for the coupling between the glass panel and the sound field in the enclosure.

$$\eta^p_{ap,ij} = \eta_a R^p_{ap,ij}/(1-R^p_{ap,ij}), \tag{20}$$

$$R^p_{ap,ij} = E^p_{ai}/E^p_{pj}, \tag{21}$$

$$\eta^p_{pa,ji} = (\eta_a + \eta^p_{ap,ij})E^p_a/E^p_p. \tag{22}$$

$\eta^a_{pa,ji}$ and $\eta^p_{pa,ji}$ are the coupling loss factors from the jth structural mode to the ith acoustic mode, and $\eta^a_{ap,ij}$ and $\eta^p_{ap,ij}$ are the coupling loss factors from the ith acoustic mode to the jth structural mode. The distribution of these modal coupling-loss factors can be obtained when the distribution of the modal energies is known. As the former determines η^a_{pa}, η^a_{ap}, η^p_{pa}, and η^p_{ap}, and the latter determines E^a_p/n_p, E^p_p/n_p, E^a_a/n_a, and E^p_a/n_a, both distributions illustrate the way that the mean modal energies affect the determination of the mean modal coupling-loss factors. Since the direct panel and direct sound-field excitations give different distributions of modal energy and modal coupling-loss factor, the influence of the consistency relationship on the estimation of η_{pa} and η_{ap} can be observed and compared for both ways of driving the system.

III. RESULTS AND DISCUSSION

A. Computational examples

Computational examples are first used to illustrate the effect of the use of the consistency relationship on the prediction of Sabine absorption coefficient of a modally reactive panel by SEA. In these examples, the acoustic-structural coupled system consists of a rectangular-parallelepiped enclosure and a simply supported rectangular glass panel. The enclosure has dimensions of $(L_x, L_y, L_z) = (0.880, 1.725, 1.540)$ m and the panel forms a boundary of the enclosure at $z = 1.540$ m. The properties of the panel are: $\eta_p = 0.05$, $A_p = 0.880 \times 1.725\ \mathrm{m}^2$, thickness, $h = 6$ mm, material density, $\rho_p = 2500\ \mathrm{kg\,m^{-3}}$, and longitudinal wave speed, $c_L = 5200\ \mathrm{m\,s^{-1}}$. E^a_p, E^a_a, E^p_p, and E^p_a are calculated by the well-established modal-coupling method.[3,24] For the direct sound-field excitation, a monopole sound source is used and it is located consecutively at twelve random positions in the enclosure. E^a_p and E^a_a are obtained by averaging the energies of the sound field and the panel over all driving locations. For the direct panel excitation, a mechanical point force is used and it is located consecutively at twelve random positions on the panel surface. E^p_p and E^p_a are calculated as the average energies for all driving locations. Values of 0.0005 and 0.01 are considered for η_a and they represent a lightly damped and a well-damped sound field in the uncoupled state, respectively. n_p and n_a are obtained by the asymptotic formulas of modal density for a flat plate in flexural vibration and a room.[12,15] By using the information of the energies, damping loss factors and modal densities, the values of η_{pa}, η_{ap}, η^a_{pa}, η^a_{ap}, η^p_{pa}, and η^p_{ap} are evaluated by Eqs. (9)–(16), and presented in Fig. 1 for some 1/3-octave bands. The results reveal a similarity and a difference in the characteristic of η^a_{pa} and η^a_{ap} from η^p_{pa} and η^p_{ap}, which are related to the two constraints imposed by the consistency relationship through the use of the modal densities (i.e., consideration of resonance modes only and the equipartition of energy among these modes).

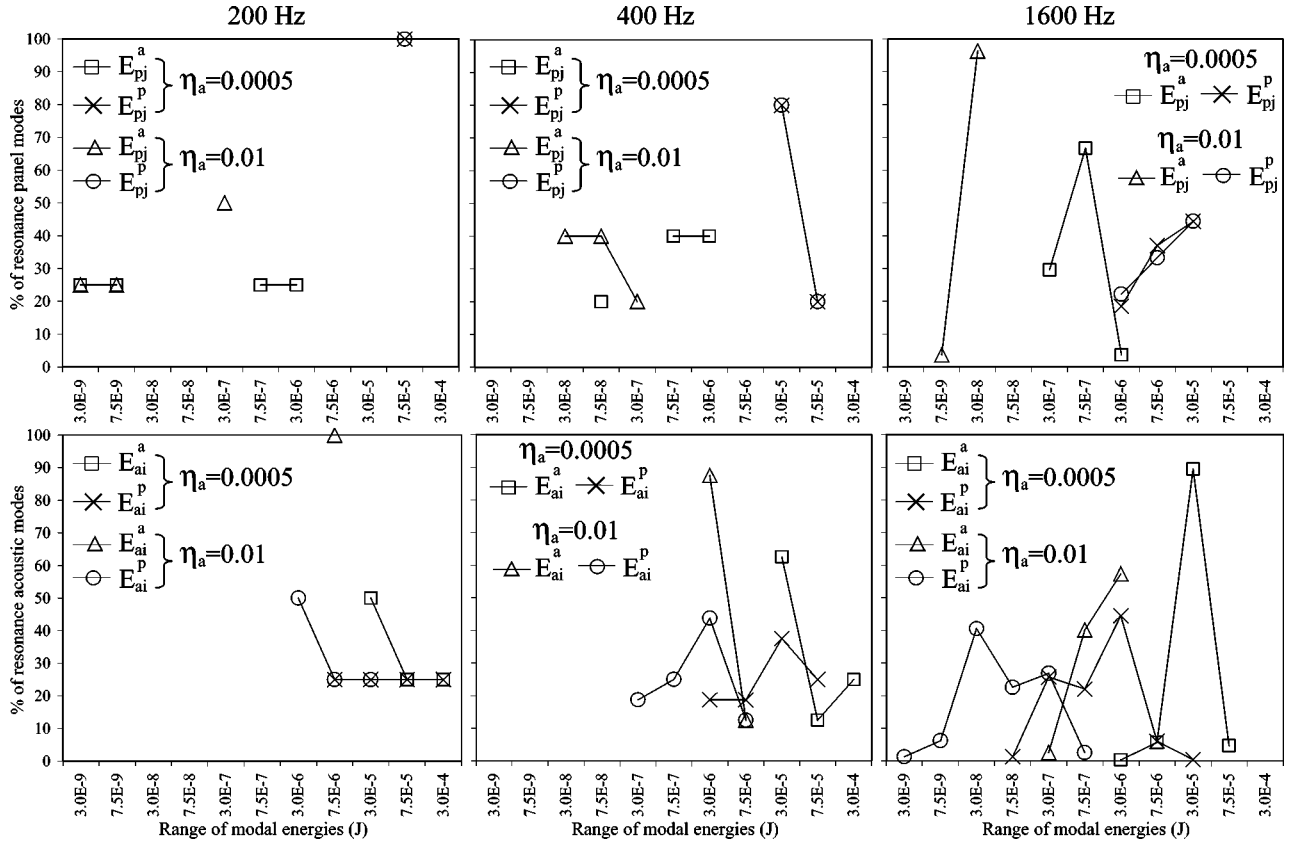


FIG. 2. Distributions of modal energy of the glass panel and the sound field in the enclosure in some selected 1/3-octave frequency bands.

In terms of the similarity, Fig. 1 shows that η^a_{pa} and η^p_{ap}, respectively, have poor predictability with respect to η_{pa} and η_{ap} in the 125 Hz band and below. The reason is because the consistency relationship directly affects η^a_{pa} and η^p_{ap} as explained in Sec. II C. Also, negative values of η^p_{pa} and η^p_{ap} are obtained for the 25 Hz band (not shown on the logarithmic scale of the figure). A comparison can be made between the number of resonance modes in each band obtained by the asymptotic formulas (i.e., n_p and n_a multiplied by Δf) and by the mode-count technique. In the latter, the number can be counted once the resonance frequencies of uncoupled modes of the panel and the sound field are known. It can be shown that the misprediction of n_p and n_a by the asymptotic formulas is very significant in the 125 Hz band and below, in particular, when the sound field has no resonance modes in most of the bands but the asymptotic formula gives non zero values of n_a. Therefore, η^a_{pa}, η^a_{ap}, η^p_{pa}, and η^p_{ap} in those bands do not have physical meanings since they involve the use of n_p and n_a but nonresonance modes are in control of the energy of one or both of the subsystems. This explains the reason for the poor predictability of η^a_{pa} and η^p_{ap} as well as the negative values of η^p_{pa} and η^p_{ap}.

Above the 125 Hz band, the number of resonance modes of both subsystems increases and it is reliably estimated by the asymptotic formulas. However, although both subsystems have a few resonance modes in the bands below 500 Hz, η^a_{pa} and η^p_{ap} also have a similar predictability that is not consistently good across the bands. In addition, η^p_{pa} and η^p_{ap} can still have negative values (e.g., in the 200 Hz band for $\eta_a=0.0005$, which is not shown on the logarithmic scale of Fig. 1). Above the 400 Hz band, the predictability of η^a_{pa} and η^p_{ap} becomes consistently good. These observations can be explained using distributions of modal energy and modal coupling-loss factor described in Sec. II D. The energy of each resonance mode is calculated by the modal-coupling method and consists of the total response that includes both acoustical- and structural-controlled peaks[2,3] of the coupled system. Modal coupling-loss factors of acoustic-structural coupled pairs of the modes are then obtained by Eqs. (17)–(22). As examples, histograms of modal energy and modal coupling-loss factor for the 200, 400, and 1600 Hz bands are presented in Figs. 2 and 3.

When the sound field is directly excited, the subsystem of a lower damping (i.e., the sound field) is driving the subsystem of a higher damping (i.e., the panel). So, values of E^a_{ai} are always greater than those of E^a_{pj} in all the three bands (see Fig. 2). In this case, $R^a_{pa,ji}$ given by Eq. (18) is less than unity and thus, values of $\eta^a_{pa,ji}$ obtained by Eq. (17) are definitely positive (see Fig. 3). Hence, as depicted in Fig. 1, η^a_{pa} has positive values. If the panel is directly excited, then the subsystem of a higher damping is driving the subsystem of a lower damping. When $\eta_a=0.01$, the sound field is well damped and Fig. 2 shows that all values of E^p_{ai} in the bands are smaller than those of E^p_{pj}. So, $R^p_{ap,ij}$ given by Eq. (21) is less than unity, which results in positive values of $\eta^p_{ap,ij}$ [see Eq. (20)] and thus, η^p_{ap}. When $\eta_a=0.0005$, the values of E^p_{ai} increase significantly and some of them become larger than those of E^p_{pj}, where values of $R^p_{ap,ij}$ above unity are pro-

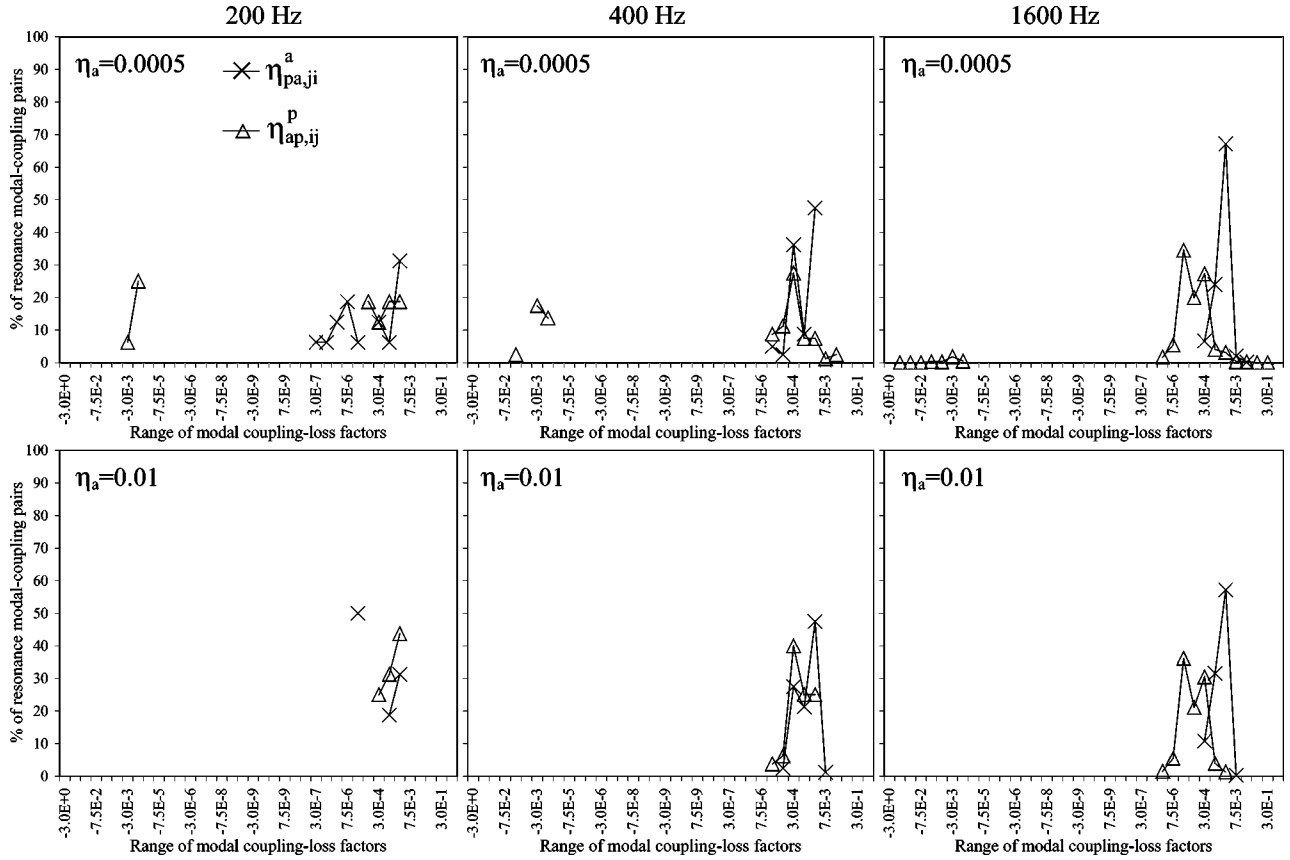


FIG. 3. Distributions of modal coupling-loss factor of acoustic-structural coupled pairs of resonance modes of the sound field in the enclosure and the glass panel in some selected 1/3-octave frequency bands.

duced. In this case, some values of $\eta^p_{ap,ij}$ are negative (see Fig. 3). However, whether or not η^p_{ap} also has a negative value would depend on the location of the mean value in the distribution of $\eta^p_{ap,ij}$. In the 200 Hz band, both subsystems have low modal-overlap factors (i.e., $M_p = n_p \eta_p f_0$ for the panel and $M_a = n_a \eta_a f_0$ for the sound field) and less than the suggested minimum of five resonance modes.[15,18] An enormous spread of values of E^p_{ai} results, which causes $\eta^p_{ap,ij}$ to have negative values from about 31% of coupled pairs of the resonance modes as well as a large scatter of positive values (see Figs. 2 and 3). So, it can be shown that $\eta^p_{ap,ij}$ of all the coupled pairs in this band has a negative mean value and correspondingly, η^p_{ap} in Fig. 1 has a negative value. However, M_p, M_a, and the number of resonance modes of the subsystems increase with frequency. Thus, given the same η_a of 0.0005 in the 400 and 1600 Hz bands, the scatter of values of E^p_{ai} as well as the population of negative values and spread of positive values of $\eta^p_{ap,ij}$ are smaller. Therefore, mean values of $\eta^p_{ap,ij}$ and hence, values of η^p_{ap}, are positive in these bands. The results suggest that the use of the consistency relationship can lead to invalid values of η^a_{pa} or η^p_{ap} if the subsystems have low M_p and M_a and less than five resonance modes, but the subsystem that is directly excited, has a much lower damping than the subsystem that is excited only by coupling.

As the bands above 200 Hz have five or more resonance modes, Fig. 1 shows that the suggested minimum number of modes[15,18] does not necessarily justify a good predictability of η^a_{pa} and η^p_{ap}. It is because the number would be too small for the modal energies to be close to equal distributions when M_p and M_a are low (e.g., the 400 Hz band). A good predictability is achieved in the bands above 400 Hz, which can be shown to have more than ten resonance modes. Such a large number of modes as well as the increase in M_p and M_a with frequency, cause the 1600 Hz band to have smaller scatters of modal energies and modal coupling-loss factors than the 400 Hz band (see Figs. 2 and 3). So, the predictability of η^a_{pa} and η^p_{ap} is better in the 1600 Hz band. The results suggest that the adverse effect from the use of the consistency relationship decreases with frequency regardless of whichever subsystem is directly excited. This is depicted in Fig. 1 as a convergence of η^a_{pa} to η_{pa} and η^p_{ap} to η_{ap}. In other words, as frequency increases, the subsystems are closer to the states of equal distribution of modal energy, which is a requirement of the reciprocity principle.

In terms of the difference in characteristic of η^a_{pa} and η^a_{ap} from η^p_{pa} and η^p_{ap}, Fig. 1 indicates that both direct excitations give different approximate values of η_{pa} and η_{ap}. When the sound field is directly excited, the discrepancy between η^a_{pa} and η_{pa} is larger than that between η^a_{ap} and η_{ap}. The reason is because the consistency relationship has a direct influence on η^a_{pa}, but for η^a_{ap}, this influence is masked by a large value of η_p [see Eqs. (11) and (13)]. The masking is shown in Fig. 4 where the influence of the relationship via η^a_{pa} is negligibly small such that $\eta_p + \eta^a_{pa}$ is indistinguishable from $\eta_p + \eta_{pa}$ in almost all the bands. If the panel is directly excited, an opposite behavior is observed where the discrepancy between η^p_{ap} and η_{ap} is larger

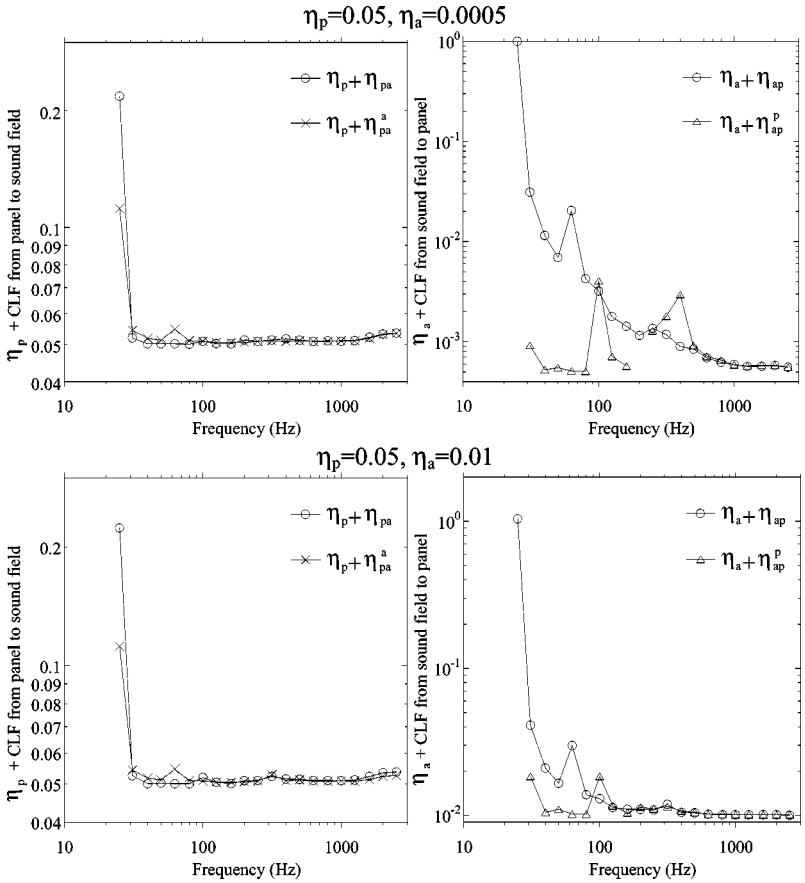


FIG. 4. The total of damping and coupling loss factors for the glass panel and the sound field in the enclosure.

than that between η_{pa}^{p} and η_{pa}. In this case, the consistency relationship has a direct influence on η_{ap}^{p}, but for η_{pa}^{p}, such influence is masked by η_a [see Eqs. (14) and (16)]. However, the extent of the masking depends on how large η_{ap}^{p} is, compared to η_a. When $\eta_a=0.0005$, Fig. 4 shows that $\eta_a+\eta_{ap}^{p}$ has considerable discrepancies from $\eta_a+\eta_{ap}$ in some bands since this value of η_a is much lower than those of η_{ap}^{p}. So, the difference of η_{pa}^{p} from η_{pa} in Fig. 1 is large in these bands. When $\eta_a=0.01$, the masking increases and $\eta_a+\eta_{ap}^{p}$ is closer to $\eta_a+\eta_{ap}$ (see Fig. 4). Hence, the difference of η_{pa}^{p} from η_{pa} is significantly reduced as depicted in Fig. 1. This explanation on the extent of the masking is also applicable to the above-mentioned $\eta_p+\eta_{pa}^{a}$ although only one value of η_p is considered. Figure 4 also indicates that for most of the bands below 125 Hz which have no resonance modes, a large masking can still lead to a negligible effect of the use of the consistency relationship. This means that from the point of view of prediction only, even η_{pa}^{p} and η_{ap}^{a} do not have physical meanings in these bands, they can still be used as good approximations to η_{pa} and η_{ap} (e.g., η_{ap}^{a} in Fig. 1) when the masking by η_a and η_p is large. Figure 5 presents some examples of the distributions of $\eta_{ap,ij}^{a}$ and $\eta_{pa,ji}^{p}$ that, respectively, have smaller spreads than those of $\eta_{pa,ji}^{a}$ and $\eta_{ap,ij}^{p}$ in Fig. 3 because of the masking by η_p and η_a. The results suggest that, in general, the consistency relationship generates a larger prediction error in the approximation of η_{pa} than in the approximation of η_{ap} when the sound field is directly excited, and vice-versa when the panel is directly excited.

C^a and C^p are evaluated for the approximate coupling loss factors by using the expression for C in Sec. II A. From the data in Fig. 1, it can be shown that $\eta_a+\eta_{ap}^{a}-C^a \ll \eta_p+\eta_{pa}^{a}+C^a$ and $\eta_a+\eta_{ap}^{p}-C^p \ll \eta_p+\eta_{pa}^{p}+C^p$. Therefore, following the discussion in Sec. II A, these terms, respectively, control s_1^a, s_2^a, s_1^p, and s_2^p, where $s_2^a \gg s_1^a$ and $s_2^p \gg s_1^p$. The symbols denote the sound-field decay rates that are calculated by Eqs. (4) and (5) for the approximate coupling loss factors. Hence, s_1^a and s_1^p, respectively, control $\alpha_{\mathrm{Sab}}^{a}$ and $\alpha_{\mathrm{Sab}}^{p}$ as from Eq. (8), the exponential terms have a more significant weighting than the absolute value terms. Furthermore, for s_1^a and s_1^p, it can be shown that $C^a \ll \eta_a+\eta_{ap}^{a}$ and $C^p \ll \eta_a+\eta_{ap}^{p}$. Thus, $\alpha_{\mathrm{Sab}}^{a}$ varies more with η_{ap}^{a} than η_{pa}^{a}, and $\alpha_{\mathrm{Sab}}^{p}$ varies more with η_{ap}^{p} than η_{pa}^{p}. Due to the good agreement between η_{ap}^{a} and η_{ap}, $\alpha_{\mathrm{Sab}}^{a}$ and α_{Sab} agree well (see Fig. 6). On the other hand, although η_{pa}^{p} agrees well with η_{pa}, $\alpha_{\mathrm{Sab}}^{p}$ is insensitive to the influence of η_{pa}^{p}. The poor agreement of η_{ap}^{p} with η_{ap} in some bands results in poor agreements between $\alpha_{\mathrm{Sab}}^{p}$ and α_{Sab} in these bands. Since the discrepancy between η_{ap}^{p} and η_{ap} is larger than that between η_{ap}^{a} and η_{ap}, Fig. 6 shows that $\alpha_{\mathrm{Sab}}^{p}$ has a larger difference than $\alpha_{\mathrm{Sab}}^{a}$ from α_{Sab}.

B. Experimental examples

In the following experimental examples, the acoustic-structural coupled system consists of a standard rectangular-parallelepiped reverberation room and six rectangular particle-board panels. The room has dimensions of $(L_x, L_y, L_z)=(6.840, 5.565, 4.720)$ m, and the panels are identical and located on the floor of the room. Each panel has

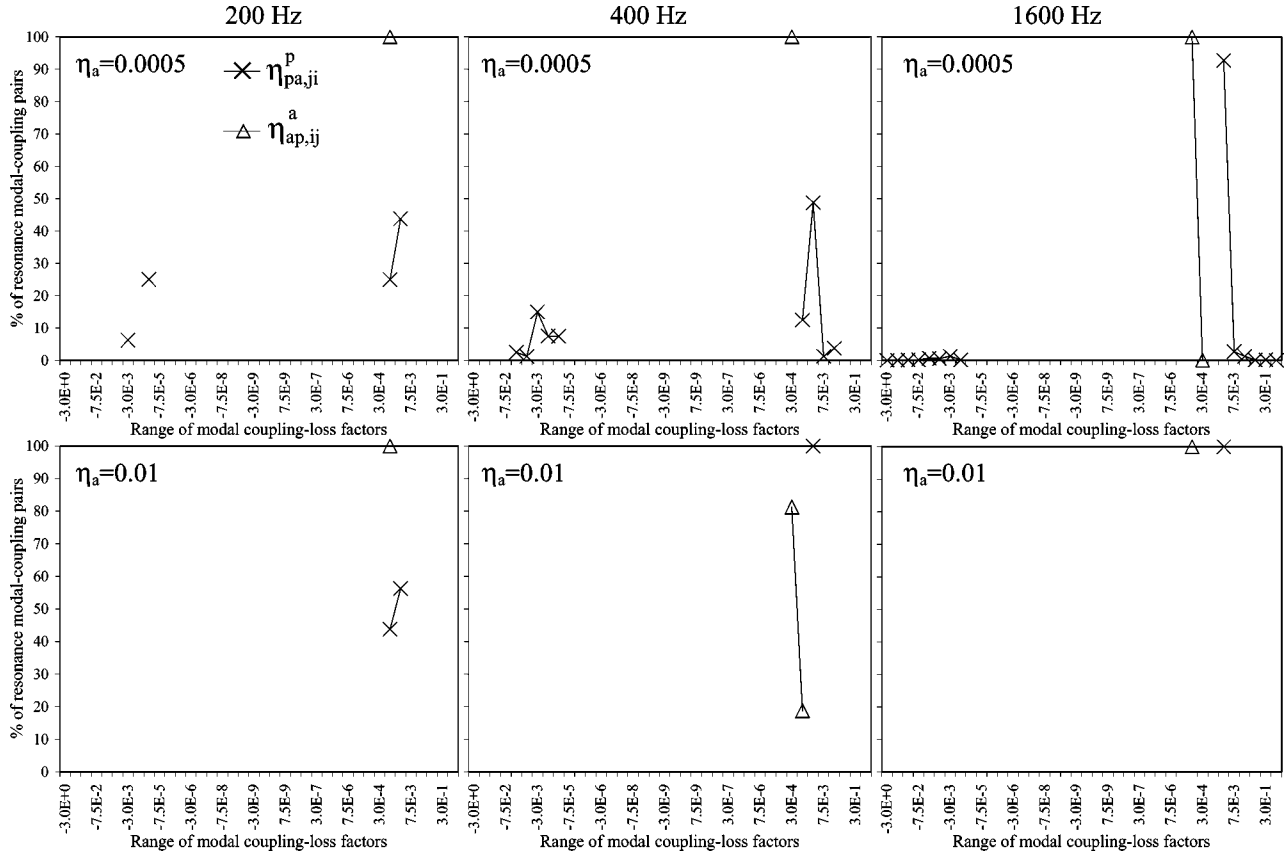


FIG. 5. Distributions of modal coupling-loss factor of acoustic-structural coupled pairs of resonance modes of the sound field in the enclosure and the glass panel in some selected 1/3-octave frequency bands.

these properties: $A_p = 1.5 \times 1.2\,\mathrm{m}^2$, $h = 10\,\mathrm{mm}$, $\rho_p = 691\,\mathrm{kg\,m}^{-3}$, and $c_L = 2540\,\mathrm{m\,s}^{-1}$. In order to avoid the contact between the panel surface and the floor, the panels are attached to wooden frames by screws along their edges on the underneath. The frames have a thickness of 20 mm and a height of 90 mm. As a result, there is an air gap of 90 mm in depth between each panel and the floor. 1/3-octave values of η_a for the sound field in the room in the absence of the panels are provided, and those of E_p^a/E_a^a and E_a^p/E_p^p for the coupling between the sound field and one panel are also available.[15] For the direct excitation of the sound field, a small loudspeaker was used. For the direct excitation of the panel, a noncontacting electromagnetic driver was used. The total loss factor of the panel in an anechoic chamber was also measured,[15] and it consists of the damping and the radiation loss factors of the panel (i.e., $\eta_p + \eta_{pa}$).[25] By substituting the values of η_a, $\eta_p + \eta_{pa}$ and the energy ratios into Eqs. (9) and (10), η_{ap}, η_{pa}, and η_p can be determined. The data for η_a and η_p are shown in Fig. 7. n_p and n_a are obtained by the asymptotic formulas.[12,15] Then, η_{pa}^a, η_{ap}^a, η_{pa}^p, and η_{ap}^p are evaluated and presented in Fig. 8.

Figure 8 indicates that the four approximate coupling loss factors exhibit a similar characteristic as that in the foregoing computational examples. It can be shown that the asymptotic formulas provide an accurate estimation of the number of resonance modes of both subsystems in most of the bands below 630 Hz. However, the panel only has around ten or less resonance modes in these bands and the number is

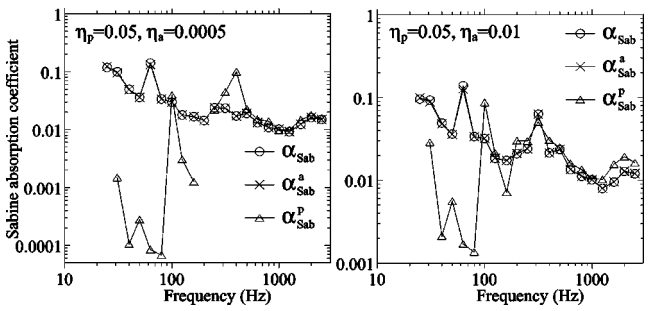


FIG. 6. Sabine absorption coefficient of the glass panel in the enclosure.

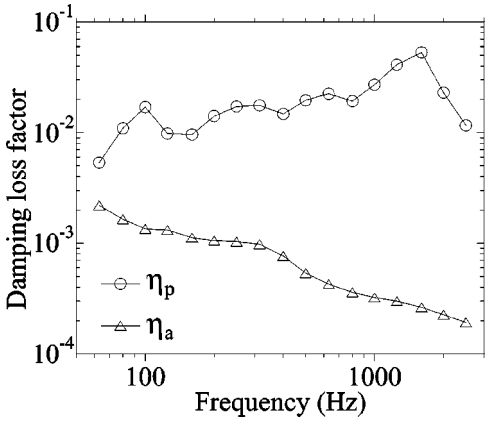


FIG. 7. Damping loss factors of the sound field in the reverberation room and the particle-board panels.

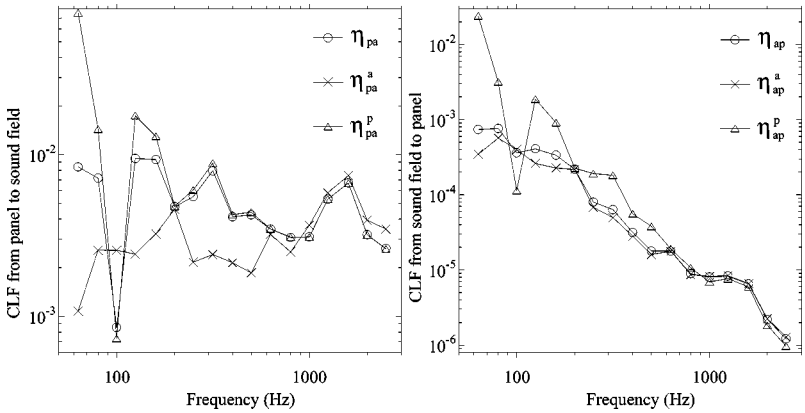


FIG. 8. Coupling loss factors for the coupling between one particle-board panel and the sound field in the reverberation room.

much less than that of the sound field. Also, as M_p and M_a in the bands are low, the modal energy of one or both of the subsystems is not necessarily close to the state of an equal distribution. Hence, the predictability of η_{pa}^a and η_{ap}^p with respect to η_{pa} and η_{ap} is poor. Negative values of η_{pa}^a and η_{ap}^p are not obtained because values of η_a and η_p in the bands are only different by an order of magnitude or less (see Fig. 7). In the 630 Hz band and above, M_p, M_a, and the number of resonance modes of both subsystems are large and thus, the subsystems are closer to the states of equipartition of modal energy. So, the adverse effect from the use of the consistency relationship decreases with frequency, and Fig. 8 illustrates this phenomenon where η_{pa}^a and η_{ap}^p, respectively, converge fairly well to η_{pa} and η_{ap}. On the other hand, as $\eta_{ap}^a \propto \eta_p + \eta_{pa}^a$ where the error in η_{pa}^a due to the use of the consistency relationship is masked by η_p, the discrepancy between η_{pa}^a and η_{pa} is larger than that between η_{ap}^a and η_{ap}. Also, the discrepancy between η_{ap}^p and η_{ap} is larger than that between η_{pa}^p and η_{pa} because $\eta_{pa}^p \propto \eta_a + \eta_{ap}^p$ where the error in η_{ap}^p is masked by η_a (see Fig. 9).

When the sound field is coupled to six panels, values of η_{ap}^a, η_{ap}^p, η_{ap}, E_p^a, E_p^p, and A_p are increased by six times in the calculation of the sound-field decay rates, the amplitudes of the exponential terms and their corresponding Sabine absorption coefficients. Also, $C^{a\#}$ and $C^{p\#}$ are evaluated where # denotes a multiple of six to η_{ap}^a and η_{ap}^p. It can be shown from the data in Fig. 8 that $\eta_a + 6\eta_{ap}^a - C^{a\#} \ll \eta_p + \eta_{pa}^a + C^{a\#}$ and $\eta_a + 6\eta_{ap}^p - C^{p\#} \ll \eta_p + \eta_{pa}^p + C^{p\#}$. So, following the discussion in Sec. II A, these terms, respectively, control s_1^a, s_2^a, s_1^p, and s_2^p, where $s_2^a \gg s_1^a$ and $s_2^p \gg s_1^p$. Therefore, α_{Sab}^a and α_{Sab}^p, respectively, depend mainly on s_1^a and s_1^p due to the enormous weighting of the exponential terms. However, for s_1^a and s_1^p, it can be shown that $C^{a\#} \ll \eta_a + 6\eta_{ap}^a$ and $C^{p\#} \ll \eta_a + 6\eta_{ap}^p$ in most of the bands. This means that α_{Sab}^a varies more with η_{ap}^a than η_{pa}^a, and α_{Sab}^p varies more with η_{ap}^p than η_{pa}^p. Since η_{ap}^a and η_{ap} agree well in most of the bands as shown in Fig. 8, the agreement between α_{Sab}^a and the SEA-predicted α_{Sab} is good in those bands (see Fig. 10). On the other hand, although η_{pa}^p and η_{pa} also agree well in most of the bands, α_{Sab}^p does not vary much with η_{pa}^p. As η_{ap}^p and η_{ap} have poor agreements in some bands, α_{Sab}^p also agrees poorly with α_{Sab} in these bands (see Fig. 10). Due to a larger discrepancy between η_{ap}^p and η_{ap} than between η_{ap}^a and η_{ap}, α_{Sab}^p has a larger difference than α_{Sab}^a from the SEA-predicted α_{Sab}. The measured α_{Sab} is also included in Fig. 10 for comparisons. It is deduced from the reverberation times of the sound field with and without the panels.[15] As obvious, the prediction accuracy of SEA can be improved when the consistency relationship is not used. In the bands below 100 Hz, the panels have two or less resonance modes so a requirement of SEA of incoherency between the modes is not satisfied.[6–8] Hence, SEA still gives a poor prediction of α_{Sab} even though the consistency relationship is not used.

IV. CONCLUSIONS

In this paper, the effect of the use of the consistency relationship in SEA on the prediction of Sabine absorption coefficient of a modally reactive panel in a room is investi-

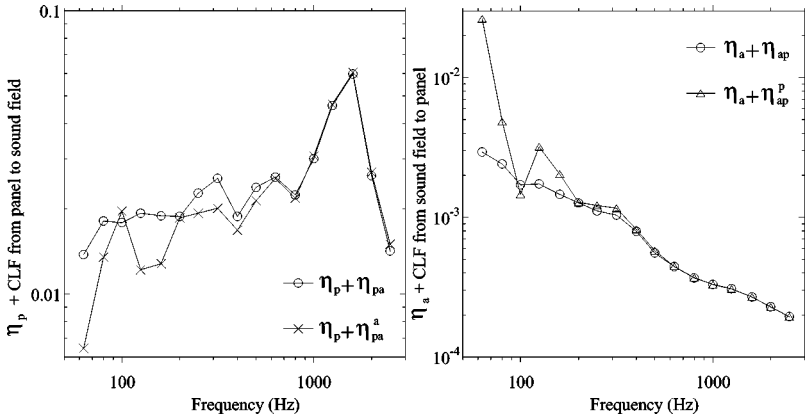


FIG. 9. The total of damping and coupling loss factors for one particle-board panel and the sound field in the reverberation room.

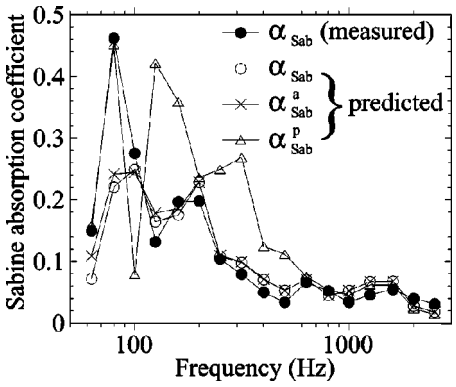


FIG. 10. Sabine absorption coefficient of the particle-board panels in the reverberation room.

gated. The use of the relationship enables SEA to predict the absorption coefficient in two different ways namely, through the direct excitation of either the panel or the room. Both ways of driving the acoustic-structural coupled system produce different prediction errors in the coupling loss factors between the panel and the room and thus, in the absorption coefficient. It is shown that for both direct excitations, η_{pa}^{a} and η_{ap}^{a} have a similar as well as a different characteristic from η_{pa}^{p} and η_{ap}^{p}.

In terms of the similarity, the consistency relationship has a direct influence on η_{pa}^{a} and η_{ap}^{p}, and the predictability of both approximate coupling loss factors is similar. In frequency bands where one or both of the subsystems have less than five resonance modes, the predictability is poor. Negative values of either of the coupling loss factors can also result when the subsystems have low modal-overlap factors (i.e., M_p and M_a) and the subsystem that is directly excited has a much lower damping than the subsystem that is excited only by coupling. Although this number of modes is the suggested minimum for a reliable use of SEA, it would be too small for η_{pa}^{a} and η_{ap}^{p} to have a good predictability. The reason is because the modal energies of the subsystems are not necessarily close to equal distributions if M_p and M_a are low. However, as the number of resonance modes in the bands of interest increases above ten when frequency increases, and M_p and M_a also increase with frequency, the subsystems are closer to the states of equipartition of modal energy. In this case, the predictability of η_{pa}^{a} and η_{ap}^{p} is consistently good across these bands. Therefore, as frequency increases, the adverse effect from the use of the consistency relationship decreases regardless of whichever subsystem is directly excited.

In terms of the difference, η_{pa}^{a} has a larger prediction error than η_{ap}^{a}, and η_{ap}^{p} has a larger prediction error than η_{pa}^{p} because the direct influence of the consistency relationship is on η_{pa}^{a} and η_{ap}^{p} but $\eta_{ap}^{a} \propto \eta_p + \eta_{pa}^{a}$ and $\eta_{pa}^{p} \propto \eta_a + \eta_{ap}^{p}$. Smaller errors in η_{ap}^{a} and η_{pa}^{p} result when the errors in η_{pa}^{a} and η_{ap}^{p} are, respectively, masked by η_p and η_a. In other words, the coupling loss factor from the panel to the room is predicted more poorly than the coupling loss factor from the room to the panel when the room is directly excited, and vice-versa when the panel is directly excited. When the masking is sufficiently large, η_{pa}^{a} would have a larger error than η_{pa}^{p}, and η_{ap}^{p} would have a larger error than η_{ap}^{a}. Thus, if $\alpha_{\mathrm{Sab}}^{a}$ and $\alpha_{\mathrm{Sab}}^{p}$ are, respectively, more sensitive to the change in η_{pa}^{a} and η_{ap}^{p}, then $\alpha_{\mathrm{Sab}}^{a}$ provides a poorer prediction of the absorption coefficient than $\alpha_{\mathrm{Sab}}^{p}$. On the contrary, $\alpha_{\mathrm{Sab}}^{p}$ gives a poorer prediction when $\alpha_{\mathrm{Sab}}^{a}$ and $\alpha_{\mathrm{Sab}}^{p}$ are, respectively, more sensitive to the change in η_{ap}^{a} and η_{ap}^{p}. It is also shown that the prediction accuracy of SEA for the absorption coefficient can be improved if the consistency relationship is not used because the uncertainty in its application is not present.

[1] K. S. Sum and J. Pan, "Some aspects of coupling-induced sound absorption in enclosures," J. Acoust. Soc. Am. **114**, 666–676 (2003).
[2] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure—Theoretical part," J. Acoust. Soc. Am. **87**, 691–707 (1990).
[3] J. Pan, "The forced response of an acoustic-structural coupled system," J. Acoust. Soc. Am. **91**, 949–956 (1992).
[4] G. Maidanik, "Some elements in statistical energy analysis," J. Sound Vib. **52**, 171–191 (1977).
[5] J. E. Brooks and G. Maidanik, "Loss and coupling loss factors of two coupled dynamic systems," J. Sound Vib. **55**, 315–325 (1977).
[6] F. J. Fahy, "Statistical energy analysis," in *Noise and Vibration*, edited by R. G. White and J. G. Walker (Horwood, West Sussex, 1982), Chap. 7.
[7] C. B. Burroughs, R. W. Fischer, and F. R. Kern, "An introduction to statistical energy analysis," J. Acoust. Soc. Am. **101**, 1779–1789 (1997).
[8] R. H. Lyon and G. Maidanik, "Power flow between linearly coupled oscillators," J. Acoust. Soc. Am. **34**, 623–639 (1962).
[9] F. J. Fahy, "Energy flow between oscillators: Special case of point excitation," J. Sound Vib. **11**, 481–483 (1970).
[10] D. A. Bies and S. Hamid, "In situ determination of loss and coupling loss factors by the power injection method," J. Sound Vib. **70**, 187–204 (1980).
[11] R. H. Lyon, "Energy sharing by coupled systems," in *Statistical Energy Analysis of Dynamical Systems: Theory and Applications* (MIT, Cambridge, 1975), Chap. 3.
[12] M. J. Crocker and A. J. Price, "Sound transmission using statistical energy analysis," J. Sound Vib. **9**, 469–486 (1969).
[13] J. P. Wilby and T. D. Scharton, "Acoustic transmission through a fuselage side wall," BBN Report No. 2742, BBN, 1974.
[14] B. L. Clarkson and K. T. Brown, "Acoustic radiation damping," J. Vib. Acoust. Stress Rel. Design **107**, 357–360 (1985).
[15] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on acoustical decays in a reverberation room in the high frequency range," J. Acoust. Soc. Am. **87**, 718–727 (1990).
[16] T. Khilman, "Sound radiation into a rectangular room. Applications to airborne sound transmission in buildings," Acustica **18**, 11–20 (1967).
[17] A. J. Pretlove, "Free vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **2**, 197–209 (1965).
[18] F. J. Fahy and A. D. Mohammed, "A study of uncertainty in applications of SEA to coupled beam and plate systems. I. Computational experiments," J. Sound Vib. **158**, 45–67 (1992).
[19] M. Hodgson, "Experimental evaluation of the accuracy of the Sabine and Eyring theories in the case of non-low surface absorption," J. Acoust. Soc. Am. **94**, 835–840 (1993).
[20] R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 545–565 (1969).
[21] P. M. Morse, "Standing waves of sound," in *Vibration and Sound* (American Institute of Physics, New York, 1976), Chap. 8.
[22] C. E. Wallace, "Radiation resistance of a rectangular panel," J. Acoust. Soc. Am. **51**, 946–952 (1972).
[23] F. G. Leppington, E. G. Broadbent, and K. H. Heron, "The acoustic radiation efficiency of rectangular panels," Proc. R. Soc. London, Ser. A **382**, 245–271 (1982).
[24] E. H. Dowell, G. F. Gorman, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542 (1977).
[25] M. Heckl, "Measurement of absorption coefficients on plates," J. Acoust. Soc. Am. **34**, 803–808 (1962).

Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces

Masayuki Morimoto[a)]
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501 Japan

Hiroshi Sato
Graduate School of Engineering, Tohoku University, Aoba, Sendai 980-8579 Japan

Masaaki Kobayashi
Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Nada, Kobe 657-8501 Japan

(Received 7 January 2004; revised 24 May 2004; accepted 31 May 2004)

Listening difficulty ratings, using words with high word familiarity, are proposed as a new subjective measure for the evaluation of speech transmission in public spaces to provide realistic and objective results. Two listening tests were performed to examine their validity, compared with intelligibility scores. The tests included a reverberant signal and noise as detrimental sounds. The subject was asked to repeat each word and simultaneously to rate the listening difficulty into one of four categories: (1) not difficult, (2) a little difficult, (3) fairly difficult, and (4) extremely difficult. After the tests, the four categories were reclassified into, not difficult [response (1)] and some level of difficulty, (the other 3 responses). Listening difficulty is defined as the percentage of the total number of responses indicating some level of difficulty [i.e. not (1)]. The results of two listening tests demonstrated that listening difficulty ratings can evaluate speech transmission performance more accurately and sensitively than intelligibility scores for sound fields with higher speech transmission performance. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1775276]

PACS numbers: 43.55.Hy, 43.71.Gv [MK] [NX] Pages: 1607–1613

I. INTRODUCTION

The acoustical design of rooms often involves optimizing speech transmission performance since speech communication is a basic function of most public spaces such as classrooms, meeting rooms, auditoriums, railway stations, airports, etc. In the past, many studies have tried to evaluate the performance of these spaces using both physical and subjective measures.

Some physical measures are based on the useful-to-detrimental ratio, which was developed by Lochner and Burger[1] and used in a simplified form as U50.[2] Clarity[3] and Deutlicheit[4] are closely related to useful-detrimental ratios but without the influence of ambient noise. Bradley[5] showed that measured values of these indices and the more recent STI (Speech Transmission Index)[6] are closely related and can be estimated from each other. Meanwhile, several studies have developed subjective measures, such as the Rhyme test,[7] the SPIN (Speech Perception in Noise) test,[8] the SRT[9] (Speech Recognition Threshold) and so on. Among them, SRT is defined as the presentation level of speech necessary for listeners to recognize the speech 50% correctly. Because we are interested in the near perfect scores required for good speech transmission and minimal communication difficulty, SRT values are not relevant to this study.

Conventional subjective measures can be roughly classified into two groups. One is the syllable articulation test using mono-syllables, or tri-syllables, etc. A defect of syllable articulation tests is that it is difficult to imagine speech transmission performance because test syllables used in listening tests lack the reality of real speech. The other choice is intelligibility tests using words and sentences which assess both phonetic and linguistic components of the process of speech perception. Listeners indicate the percentage of words or sentences that they recognize. In these tests it is easy to imagine the actual speech transmission performance. However, the intelligibility score depends on the choice of words and sentences used in the listening tests.[8] Thus, each of the syllable articulation tests and intelligibility tests using words and sentences has both merits and disadvantages. Furthermore, even if the intelligibility is perfect, the speech transmission performance is not always satisfactory. Although a 0 dB useful-to-detrimental sound ratio corresponds to above 90% intelligibility,[5] this is not always adequate for ideal speech communication in spaces. In response to these problems, subjective ratings of the ''easiness'' of speech recognition were presented by Sato *et al.*[10] as an alternative approach. However, ''easiness'' ratings result in a relative scale and one cannot say how easy is good enough or when a low rating corresponds to unacceptably bad acoustical conditions. In addition, ''easiness'' lacks objectivity and would be influenced by personal preferences. It should probably be considered as related to speech quality.

From an engineering point of view, it is important to obtain a physical measure, which can evaluate the speech transmission performance. To achieve this purpose, it is necessary to compare the subjective measures with objective measures of realistic situations that could be a reference for

[a)]Electronic mail: mrmt@kobe-u.ac.jp

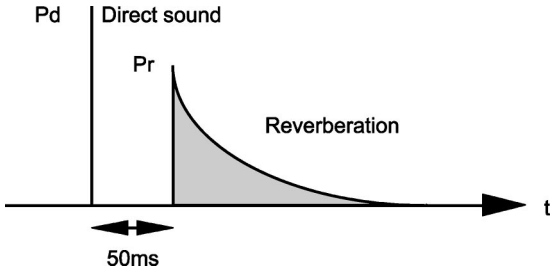


FIG. 1. Impulse response of the test sound field. Pd and Pr is a sound pressure level of a direct sound and a reverberation signal, respectively.

judging whether or not the physical measure is useful for evaluating speech transmission performance.

To solve the problems mentioned above, the present authors propose ''listening difficulty'' using words with high ''word familiarity,'' as a subjective measure, which is adequate to imagine the speech transmission performance and does not depend on the test words. ''Word familiarity'' is a subjective variable which indicates how familiar native speakers are with the word. The familiarity of about eighty thousand Japanese words has been rated from 1.0 (minimum) to 7.0 (maximum) by Amano *et al.*[11] Another reason why high familiarity words are used is to minimize the effect of the cognitive process on the difficulty rating since the main object of interest in this study is only the speech transmission performance of spaces and including these with public address systems. Moreover, the reason why ''difficulty'' is introduced, instead of ''easiness'' is that people more frequently experience listening difficulty in daily life. In other words, ''difficulty'' has reality and objectivity and matches our normal feelings more than ''easiness.''

Our purpose in this study is to examine the validity of ''listening difficulty,'' and to compare it with intelligibility scores.

II. LISTENING TEST I: LISTENING DIFFICULTY AND INTELLIGIBILITY FOR SOUND FIELDS WITH REVERBERATION

First, the validity of listening difficulty is examined by using a simple sound field that consisted of a direct sound and a reverberant signal as detrimental sound.

A. Listening test

1. Test sound fields and test words

Figure 1 shows schematically the impulse response of the test sound field. The sound field was composed of a direct sound followed by a reverberant decay. The reverberant decay started at 50 ms after the direct sound. As Bradley *et al.*[12] reported, increased early reflection energy arriving at the listener within the first 50 ms after the direct sound has the same effect on speech intelligibility scores as an equal increase in the direct sound energy. In this study, simulated sound fields are used including only the direct sound component as useful sound energy and reverberant signals as detrimental sound energy. The parameters were reverberation time and the sound pressure ratio of the initial amplitude of the reverberant decay to the direct sound amplitude described as Pr/Pd. The reverberation time was set at 0.5, 0.9, 2.0 and 6.0 s. The Pr/Pd was set at 0.5, 0.25, and 0.1. Table I shows the eight kinds of sound field used in the listening test.

TABLE I. Eight kinds of sound field used in the first listening test.

	Sound field							
Parameter	A	B	C	D	E	F	G	H
Reverberation time (s)	0.5	0.9	0.9	0.9	2.0	2.0	2.0	6.0
Pr/Pd	0.5	0.5	0.25	0.1	0.5	0.25	0.1	0.5

Two word lists with two different levels of word familiarity in Japanese[11,13] were used in this test. The word familiarity, categorized from 1.0 to 7.0, as mentioned above was divided into four levels. One was the word list with the level 4 familiarity corresponding to the categories from 7.0 to 5.5, which was used for the subjective measure proposed here. The other was the word list with the level 3 familiarity corresponding to the categories from 5.5 to 4.0. This word list was added to the test in order to determine the effect of word familiarity on the difficulty ratings. In this paper, the former and the latter are called ''high familiarity'' and ''low familiarity,'' respectively. Each word list included 50 phonetically balanced words. Each word has four syllables.

2. Subjects

A total of 122 subjects were used in the test. They varied in age from 19 to 31 years old. They did not report any known hearing impairment. Thus the subjects were not carefully selected listeners, but were more representative of general young listeners with normal hearing.

3. Procedure

In the test, five words of the 50 words with each familiarity were presented to each subject for each of eight different sound fields. Thus, each subject listened to a set of 80 test signals in total (5 words$\times$2 familiarities$\times$8 sound fields) in a random order. Moreover, each word was presented to a subject only once. A different set of 80 test signals was presented to 122 subjects so that all 800 test signals (50 words $\times$2 familiarities$\times$8 sound fields) might be presented in the whole of the listening test. At the beginning of the test, 12 test signals, every other of six words with each familiarity for each of three test sound fields A, E, and H, respectively, were added as a practice trial. The interval for presentation of each test signal was 9 s. The total length of the test was about 19 min for each subject.

The test signals were presented to two subjects at a time from a loudspeaker in an anechoic room. The listening positions were located at 3 m distance from the loudspeaker and at $\pm$30 degrees from the central axis of the loudspeaker. The frequency response of the loudspeaker was flat to within $\pm$5 dB from 100 to 10 kHz at both listening positions. The sound pressure levels of the test signals were set at 65.0 dB A, fast, peak on the average of all test words at listening positions when only the direct sound was presented. The range of the levels of the test words was within $\pm$2.0 dB A.

Each subject was asked to repeat by writing down each word as they listened and simultaneously to rate the listening

TABLE II. Four categories for listening difficulty.

(1) Not difficult
(2) A little difficult
(3) Fairy difficult
(4) Extremely difficult

difficulty into one of the four categories shown in Table II. The object of measuring listening difficulty is classifying conditions as either to ''no-difficulty'' or ''difficulty.'' However, it was difficult for listeners to identify the boundary between ''no-difficulty'' and ''difficulty'' and to answer in terms of one of the two categories. The solution employed in the listening test was to ask subjects to answer in terms of one of four categories described in Table II and then to calculate the difficulty rating with the boundary between ''no-difficulty'' and ''difficulty'' set in analysis which is going to be explained later.

B. Results and discussion

1. Definition of listening difficulty

Figure 2 indicates the results of the listening test for sound fields with reverberation. Panel (a) is the result for the high familiarity words and the panel (b) is that for the low familiarity words. The abscissa indicates the sound field and the ordinate indicates the score as a percentage. The number of responses was 610 (5 words per sound field $\times 122$ subjects). Open circles are the intelligibility scores. Open triangles are the listening difficulty ratings ''1.'' Closed circles are the sum of the difficulty ratings ''1'' and ''2.'' Crosses are the sum of the difficulty ratings ''1,'' ''2,'' and ''3.'' Both of these results indicate that open circles do not coincide with open triangles, but coincide well with closed circles. This means that even if the subjects can repeat words correctly, they feel them difficult to listen to in some cases. In other words, even if the intelligibility is 100%, the speech transmission performance in these cases cannot always be regarded as satisfactory. The speech transmission performance could be regarded as perfect when the percentage of responses for listening difficulty is zero.

Accordingly, listening difficulty is defined as a subjective measure of speech transmission performance that indicates the percentage of the sum of the difficulty responses ''2,'' ''3,'' and ''4'' (i.e., except ''1''—not difficult). From this definition and the results mentioned above, it can be considered that the listening difficulty can evaluate the speech transmission performance more accurately than intelligibility scores.

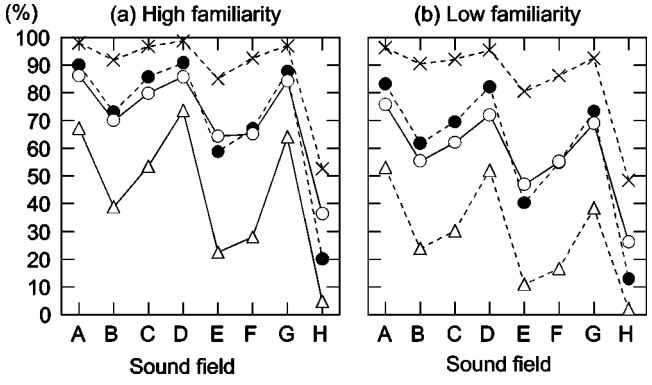


FIG. 2. Results of the first listening test. Left panel: high familiarity. Right panel: low familiarity. Open circles indicate the intelligibility scores. Open triangles, closed circles and crosses indicate the responses of difficulty (1), (1)+(2), and (1)+(2)+(3), respectively.

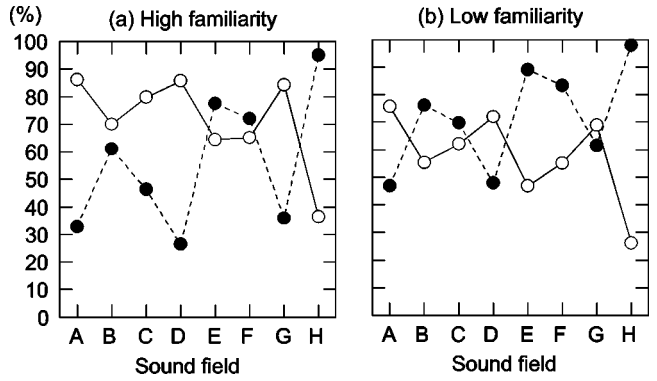


FIG. 3. Listening difficulty ratings and the intelligibility scores. Left panel: high familiarity. Right panel: low familiarity. Open circles indicate listening difficulty ratings and closed circles indicate the intelligibility scores.

2. Relation between listening difficulty and intelligibility

Figure 3 shows listening difficulty ratings and the intelligibility scores for each test sound field. Closed circles indicate the listening difficulty ratings and open circles indicate the intelligibility scores. For both word familiarities, it can be seen that there is a negative correlation between listening difficulty and intelligibility. Figure 4 shows the correlation between listening difficulty and intelligibility. Each point represents a different sound field. As indicated in Fig. 4, correlation coefficients indicate a strong negative correlation between them for both familiarities. The regression equations were obtained such that ''x'' is the listening difficulty and ''y'' is the intelligibility. For the high familiarity, the coefficient of x is 0.653 that is half of the coefficient of y. This means that the listening difficulty can evaluate the speech transmission performance more sensitively than the intelligibility. Also the same tendency can be seen for the low familiarity words, although the coefficient is a little larger than for the high familiarity words. In summary, it is

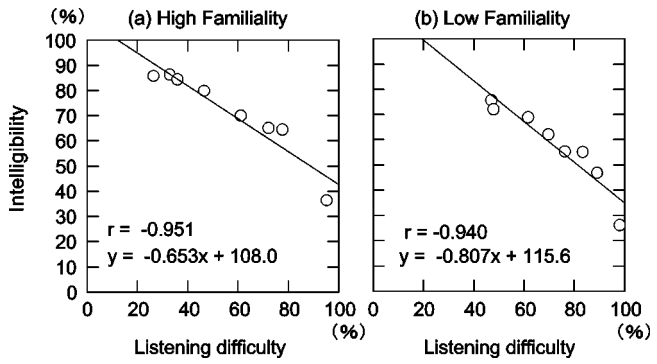


FIG. 4. The relation between listening difficulty and intelligibility. Left panel: high familiarity. Right panel: low familiarity. In the resulting regression equations, ''x'' is the listening difficulty and ''y'' is the intelligibility.

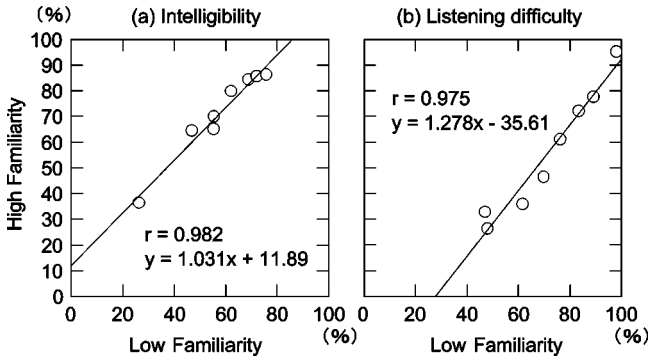


FIG. 5. Relation between the high familiarity and the low familiarity. Left panel: intelligibility. Right panel: listening difficulty. In the resulting regression equations, "*x*" is the low familiarity and "*y*" is the high familiarity.

concluded that listening difficulty ratings can evaluate speech transmission performance more accurately and sensitively than intelligibility scores in this test region.

3. Effects of word familiarity on listening difficulty and intelligibility

Figure 5 shows the correlation between results for words with the high and the low familiarities for intelligibility [panel (a)] and listening difficulty [panel (b)]. Each point represents a different sound field. It can be seen that when the word familiarity becomes lower, intelligibility becomes lower and the listening difficulty becomes higher. The correlation coefficients and the regression equations between results for the high familiarity and the low familiarity were obtained. As shown in Fig. 5, there is a strong correlation between both the intelligibility and listening difficulty for low and high familiarity words. Furthermore, the coefficients of x in the regression equations are almost equal. This means that the effects of word familiarity on intelligibility and listening difficulty do not depend on the test sound fields.

III. LISTENING TEST II: LISTENING DIFFICULTY AND INTELLIGIBILITY FOR SOUND FIELDS WITH REVERBERATION AND NOISE

In this second experiment, the validity of listening difficulty is examined by using sound fields consisting of a direct sound, plus a reverberation signal and noise as detrimental sound. Bradley[5] demonstrated that an optimum speech transmission performance can be obtained when the signal-to-noise (SN) ratio is at least 15 dB A for reverberation times less than 0.5 s. In this section, the effects of noise level and reverberation time on the listening difficulty are investigated and compared with the effects on intelligibility scores to determine acceptable SN ratios for good speech transmission.

A. Listening test

1. Test sound fields and test words

The impulse response of the test sound field is the same as shown in Fig. 1. The Pr/Pd is kept constant at 0.1. The sound pressure levels of the test signals were kept constant at 55.0 dB A, slow, peak for the average of all test words at listening positions when only the direct sound was presented. The range of the levels of test words was within ± 2.0 dB A. The parameters were reverberation time and background noise level. The reverberation time was set at 0.5 and 2.0 s. The background noise level was set at 0 (no noise was presented from the loudspeaker), 10, 25, 40, 55, and 60 dB A. The ambient noise level in the experimental room was 17.8 dB A. The level corresponds to the speech signal (direct sound) to noise ratio (SN ratio) of infinity (no additional noise), 45, 30, 15, 0, and -5 dB A, respectively. Hence the total number of sound fields used in the listening test was twelve (2 reverberation times$\times$6 background noise levels). The power spectrum of noise used as a background noise is the same as that found by Hoth[14] for each 1/1 octave band from 125 Hz to 8 kHz.

Six word lists (300 words) with the high familiarity in Japanese were used in the test. One of the lists was also used in the first test.

2. Subjects

Thirteen university students with normal hearing acted as the subjects in the test. They did not serve as subjects for the first test.

3. Procedure

In the test, 25 words of 300 words were presented to each subject for each of 12 different sound fields. Thus each subject listened to a set of 300 test signals in total (25 words$\times$12 sound fields). Moreover, each word was presented to a subject only once. Three hundred test signals were arranged in a random order and divided into six units of 50 words. Six units were presented separately. Different sets consisting of 300 test signals were presented to 13 subjects so that all 3600 test signals (300 words$\times$12 sound fields) might be presented during the whole of the listening test. The duration of each word was around 700 ms. The duration of each test signal depended on reverberation time. The duration of a background noise for each test signal was around 7 s depending on the duration of the test signal. The background noise was presented in advance of the test signal by 135 ms and was ended after each test signal ended. The interval for presentation of background noise was 6 s. During the interval no noise was presented. The total length of each unit was about 10 min.

The test signals were presented to two subjects at a time from a loudspeaker in an anechoic room as in the first test. The listening positions and the loudspeaker used for the test were the same as those in the first test.

Each subject was asked to repeat by writing down each test word as they listened and to rate the listening difficulty into one of the four categories shown in Table II.

B. Results and discussion

Figure 6 indicates the results of the listening tests for sound fields with reverberation and noise. The abscissa indicates the noise level and SN ratio and the ordinate indicates the score as a percentage. The number of responses was 325 (25 words per sound field$\times$13 subjects). Circles indicate the

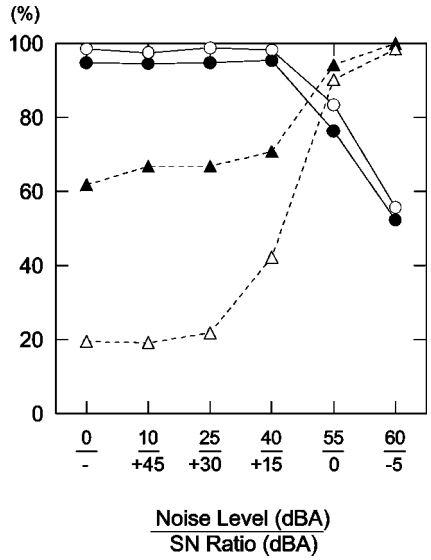


FIG. 6. Listening difficulty ratings and the intelligibility scores as a function of noise level and speech signal to noise ratio. Open and filled symbols indicate scores for the reverberation time of 0.5 and 2.0 s, respectively. Circles represent the intelligibility and triangles indicate the listening difficulty. A 0 dB A noise level means no additional noise was presented from the loudspeaker. The ambient noise level in the experimental room was 17.8 dB A.

intelligibility and triangles indicate the listening difficulty. Open and filled symbols represent scores for the reverberation time of 0.5 and 2.0 s, respectively.

First, let us discuss both types of scores for the sound fields with noise of less than 40 dB A (SN ratio$>$15 dB A). The intelligibility is close to 100%, regardless of the reverberation time. On the other hand, the listening difficulty for the reverberation time of 0.5 s is better than that for 2.0 s. The difference is about 20% for noise of 40 dB A (SN ratio $=$15 dB A) and 40% for less than 25 dB A (SN ratio $>$30 dB A). Furthermore, the listening difficulty approaches a fixed value for noise of less than 25 dB A (SN ratio $>$30 dB A). The value is about 20% for the reverberation time of 0.5 s and about 60% for 2.0 s. The listening difficulty never reaches 0% with reverberation present, but the intelligibility increases to 100% as the noise level decreases.

Next, let us discuss both scores for the sound fields with noise of more than 55 dB A (SN ratio$<$0 dB A). Listening difficulty is affected by neither the noise level nor the reverberation time. Listening difficulty reaches 100% for noise of more than 55 dB A (SN ratio$<$0 dB A). On the other hand, intelligibility is clearly affected by the noise level, but not by the reverberation time. The intelligibility is about 80% and 50% for noise levels of 55 dB A (SN ratio$=$0 dB A) and 60 dB A (SN ratio$=-5$ dB A), respectively.

These results suggest that listening difficulty is not always high when background noise is present. Listening difficulty ratings can evaluate speech transmission performance more sensitively and accurately than intelligibility scores when the background noise is lower than 40 dB A (SN ratio$>$15 dB A). Conversely, intelligibility scores can evaluate speech transmission performance more sensitively than listening difficulty ratings when background noise is higher than 40 dB A (SN ratio$<$15 dB A). Namely, listening difficulty ratings can evaluate speech transmission performance more sensitively than the intelligibility when the performance is high, and *vice versa*.

Finally let us discuss a necessary SN ratio for good speech transmission when the speech level is 55 dB A. Bradley[2] reported that the desired ratio is at least 15 dB A for a reverberation time of 0.5 s. The results of this listening test also demonstrated that the intelligibility is almost 100% under the same condition as that mentioned above. However, there is a difference of 20% in the listening difficulty between SN ratios of 15 and 30 dB A for the same reverberation time. In other words, listening difficulty evaluates the speech performance more strictly than does intelligibility. As a result, for the minimum listening difficulty, the necessary SN ratio is higher than 15 dB A, when the speech level is 55 dB A.

Additionally, there is probably an inverse correlation between useful-to-detrimental energy and listening difficulty but further study in various S/N and acoustical conditions is required to find a clear relation between them.

IV. FURTHER DISCUSSION ON SUBJECTIVE MEASURES OF SPEECH TRANSMISSION

The relation between subjective measures of speech transmission and the performance of speech transmission of sound fields is discussed based on the results obtained in the present paper.

Figure 7 illustrates different relations between them. Panel (a) presents the ideal relation between them. The subjective value monotonically increases as the performance is advanced. The subjective value reaches the lowest when the performance is minimum and it reaches the highest when the performance is maximum.

At present, however, no subjective measure demonstrates such a relationship as shown in panel (a).

At present, the most common listening test material for the subjective evaluation of speech transmission is an intelligibility test such as the rhyme test. Most previous[2,13,15] papers have demonstrated relations as illustrated in panel (b). Namely, that intelligibility scores can differentiate between the performance of sound fields where the performance is relatively low. On the other hand, it cannot clearly identify differences in the performance of sound fields where the performance is relatively high, because the intelligibility reaches a plateau near to 100%.

Meanwhile, in the present paper we propose listening difficulty ratings as a subjective measure for an evaluation of the speech transmission performance and demonstrate that (1) Listening difficulty and intelligibility scores show a highly negative correlation and listening difficulty ratings can evaluate speech transmission performance more sensitively than intelligibility scores when the performance is high, and intelligibility scores can evaluate speech transmission performance more sensitively than difficulty ratings when the performance is low. (2) Intelligibility is high even when the listening difficulty is close to 100%. (3) Listening difficulty changes significantly and does not reach 0% even if the intelligibility is 100%. These results yield the relation

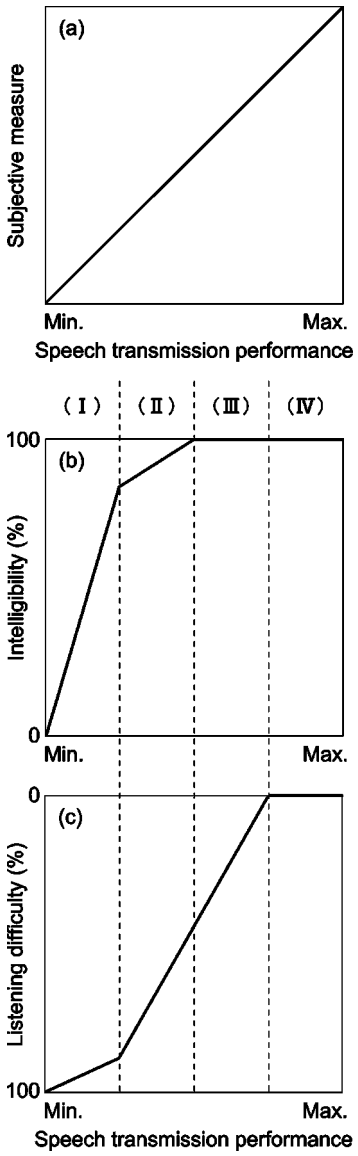


FIG. 7. Schematic relations between subjective measures of speech transmission and the performance of speech transmission. Panel (a) presents the ideal relation. Panel (b) illustrates the relation between the performance and the intelligibility. Panel (c) shows the relation between the performance and the listening difficulty.

between listening difficulty and speech transmission performance as shown in panel (c).

In summary, the relation between intelligibility and listening difficulty can be explained reasonably by dividing the speech transmission performance into four ranges from (I) to (IV) as shown in Fig. 7. In the range (I) where the speech transmission performance is the worst, the performance can be evaluated sensitively by using the intelligibility. On the other hand, the listening difficulty is around 100% in this range and it cannot describe clearly the differences in performance. This condition corresponds to the test sound fields with noise levels of 55 and 60 dB A used in the second test. In the range (II), the performance is a little higher, and listening difficulty and intelligibility show a highly negative correlation. Furthermore, listening difficulty can evaluate the speech transmission performance more accurately and sensitively than does intelligibility. The eight test sound fields used in the first test are in this range. In the range (III), where the performance is higher, the intelligibility reaches 100% and cannot differentiate among the conditions. On the other hand, listening difficulty can evaluate the performance sensitively. This condition corresponds to the test sound fields with noise levels of less than 40 dBA used in the second test.

Actually, it is usually not required to evaluate the performance of sound fields belonging to the range (I), since actual sound fields to be evaluated, such as classrooms, meeting rooms, railway stations, auditoriums, airports etc., have higher speech transmission performance than this range (I). Therefore, it can be concluded that listening difficulty is more suitable to evaluate speech transmission performance in public spaces than intelligibility scores, because difficulty can evaluate the performance more sensitively and accurately than intelligibility scores in the ranges (II) and (III).

Sound fields in the range (IV), in which nobody feels listening difficulty (0% difficulty) and of course in which the intelligibility is 100%, must be evaluated by using other measures such as the speech quality which may have higher sensitivity than listening difficulty ratings in this range.

V. CONCLUSIONS

In the present paper we have proposed listening difficulty ratings using words with the maximum familiarity as a subjective measure that can evaluate speech transmission performance realistically and objectively. Two kinds of listening tests were carried out to validate listening difficulty ratings compared with intelligibility scores. Sound fields with reverberation and others with reverberation and noise were used as test sound fields in the first and the second listening tests, respectively. The results of the two listening tests clarified that (1) Listening difficulty ratings can evaluate speech transmission performance more accurately and sensitively than intelligibility scores for sound fields with a higher speech transmission performance. (2) Conversely, intelligibility scores can evaluate the performance more sensitively than listening difficulty ratings for the sound fields with lower performance. (3) It is possible to estimate the listening difficulty for words of one word familiarity from those of another word familiarity in these test regions. (4) A necessary speech signal to noise ratio for good speech transmission is at least 15 dB A in terms of intelligibility, while it is at higher than 15 dB A in terms of listening difficulty. In conclusion, listening difficulty ratings are more suitable to evaluate actual sound fields, such as classrooms, meeting rooms, railway stations, auditoriums, airports etc., than intelligibility scores.

ACKNOWLEDGMENTS

The authors acknowledge NTT Basic Research Laboratories and Research Institute of Electrical Communication of Tohoku University for providing the phonetically balanced word lists based on word familiarity. The authors are grateful to Dr. John S. Bradley for his fruitful discussions. A part of this study was carried out as a project of the Sub Working Group for Developing Evaluation and Design Standards of Speech Transmission Quality at the Architectural Institute of

Japan. The research project was partially supported by the research grant from Kajima Foundation and by the Twenty-first Century Center of Excellence (COE) Program ''Design Strategy towards Safety and Symbiosis of Urban Space'' awarded to the Graduate School of Science and Technology, Kobe University. The Ministry of Education, Culture, Sports, Science and Technology of Japan sponsored the Program.

[1] J. P. A. Lochner and J. F. Burger, ''The intelligibility of speech under reverberant conditions,'' Acustica **11**, 195–200 (1961).
[2] J. S. Bradley, ''Predictors of speech intelligibility in rooms,'' J. Acoust. Soc. Am. **80**, 837–845 (1986).
[3] W. Reichardt and U. Lehmann, ''Optimierung raumeindruck und durchsichtigkeit von musikdarbietungen durch auswertung von impulsschalltests,'' Acustica **48**, 174–185 (1981).
[4] R. Thiele, ''Richtungsverteilung und zeitfolge der schallruckwurfe in raumen,'' Acustica **3**, 291–302 (1953).
[5] J. S. Bradley, ''On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility,'' J. Acoust. Soc. Am. **106**, 1820–1828 (1999).
[6] T. Houtgast and H. J. M. Steeneken, ''The modulation transfer function in room acoustics as a predictor of speech intelligibility,'' Acustica **28**, 66–73 (1973).
[7] G. Fairbanks, ''Test of phonetic differentiation: The Rhyme Test,'' J. Acoust. Soc. Am. **30**, 596–600 (1958).
[8] D. N. Kalikow, K. N. Stevens, and L. L. Elliott, ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351 (1977).
[9] R. Plomp and M. Mimpen, ''Speech-reception threshold for sentences as a function of age and noise level,'' J. Acoust. Soc. Am. **66**, 1333–1342 (1979).
[10] H. Sato, H. Yoshino, and M. Nagatomo, ''Relationship between speech transmission index and easiness of speech perception in reverberatory fields,'' J. Acoust. Soc. Am. **103**, 2999 (1998).
[11] http://www.ais.riec.tohoku.ac.jp/lab/wordlist/index-j.html (in Japanese).
[12] J. S. Bradley, H. Sato, and M. Picard, ''The importance of early reflections for speech intelligibility in rooms,'' J. Acoust. Soc. Am. **113**, 3233–3244 (2003).
[13] S. Sakamoto, Y. Suzuki, S. Amano, K. Ozawa, T. Kondo, and T. Sone, ''New lists for word intelligibility test based on word familiarity and phonetic balance,'' J. Acoust. Soc. Jpn. **54**, 842–849 (1998).
[14] D. F. Hoth, ''Room noise spectra at sbscribers's telephone locations,'' J. Acoust. Soc. Am. **12**, 499 (1941).
[15] J. S. Bradley, ''Speech intelligibility studies in classrooms,'' J. Acoust. Soc. Am. **80**, 846–854 (1986).

A note on measurement of sound pressure with intensity probes[a)]

Peter Juhl[b)]
Physics Department, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Finn Jacobsen[c)]
Acoustic Technology, Ørsted DTU, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kgs. Lyngby, Denmark

(Received 1 December 2003; revised 8 March 2004; accepted 8 June 2004)

The effect of scattering and diffraction on measurement of sound pressure with ''two-microphone'' sound intensity probes is examined using an axisymmetric boundary element model of the probe. Whereas it has been shown a few years ago that the sound intensity estimated with a two-microphone probe is reliable up to 10 kHz when using 0.5 in. microphones in the usual face-to-face arrangement separated by a 12 mm spacer, the sound pressure measured with the same instrument will typically be underestimated at high frequencies. It is shown in this paper that the estimate of the sound pressure can be improved under a variety of realistic sound field conditions by applying a different weighting of the two pressure signals from the probe. The improved intensity probe can measure the sound pressure more accurately at high frequencies than an ordinary sound intensity probe or an ordinary sound level meter. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777851]

PACS numbers: 43.58.Fm, 43.50.Yw, 43.38.Kb [AJZ] Pages: 1614–1620

I. INTRODUCTION

The dominating principle for measurement of sound intensity in air employs two closely spaced pressure microphones and is based on approximating the gradient of the sound pressure by a finite difference.[1] 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.[2] In 1998 it was found that the effect of scattering and diffraction tends to counterbalance the finite-difference error under virtually any sound field condition that can be encountered in practice.[3] It was shown that the practical upper frequency limit of a sound intensity probe based on two 0.5 in. microphones separated by a 12 mm spacer in the face-to-face arrangement is about 10 kHz,[3] which is about an octave higher than the frequency limit determined by the finite-difference approximation.[1,2]

One of the most important indicators for evaluating the reliability of a sound intensity or sound power measurement is the pressure-intensity index (the normalized ratio of the mean square pressure to the intensity in logarithmic form). This quantity, which can be determined with the intensity probe during the measurement,[1] reflects the acoustic conditions of the measurement and provides important information about the influence of phase mismatch.[1,4] Determining the pressure-intensity index obviously involves measuring the sound pressure. However, experimental results have indicated that the sound pressure level measured with a typical intensity probe tends to be too low at high frequencies even though the sound intensity is accurately estimated.[5] Underestimation of the pressure-intensity index may cause the operator to accept inaccurate measurements. Thus, the purpose of this paper is to examine how the sound pressure is estimated by a conventional intensity probe in various sound fields, and, if possible, to improve the estimate.

II. THE FINITE-DIFFERENCE ERROR

The complex sound pressure at the center of the probe is estimated as

$$\hat{p} = (\hat{p}_1 + \hat{p}_2)/2, \tag{1}$$

where $\hat{p}_1$ and $\hat{p}_2$ are the pressure signals as sensed by the microphones of the probe. Here and in what follows the caret indicates an estimated quantity, which may differ from ''true'' value because of diffraction, scattering, and the finite difference approximation. The axial component of the complex particle velocity (the component in the direction of the line that joins the acoustic centers of the two microphones) is

$$\hat{v}_r = (\hat{p}_1 - \hat{p}_2)/(j\rho ck\Delta r), \tag{2}$$

where c is the speed of sound, ρ is the density of air, k is the wave number, and Δr is the acoustic distance between the microphones. This leads to the following expression for the estimated sound intensity:[1]

$$\hat{I}_r = \mathrm{Im}\{\hat{p}_1 \hat{p}_2^*\}/(2\rho ck\Delta r). \tag{3}$$

The finite-difference error of the intensity is due to the combined effect of Eqs. (1) and (2), i.e., the effect of replacing the pressure at the center of the probe with the arithmetic mean of the two pressure signals and replacing the true pressure gradient $\partial p/\partial r$ with the finite difference $\Delta p/\Delta r$ (although the term ''finite-difference error'' seems to refer only

[a)]Portions of this work were presented in ''Sound pressure measurements with sound intensity probes,'' Proceedings of 18th International Congress on Acoustics, Kyoto, Japan, April 2004.
[b)]Electronic mail: pmjuhl@fysik.sdu.dk
[c)]Electronic mail: fja@oersted.dtu.dk

 © 2004 Acoustical Society of America

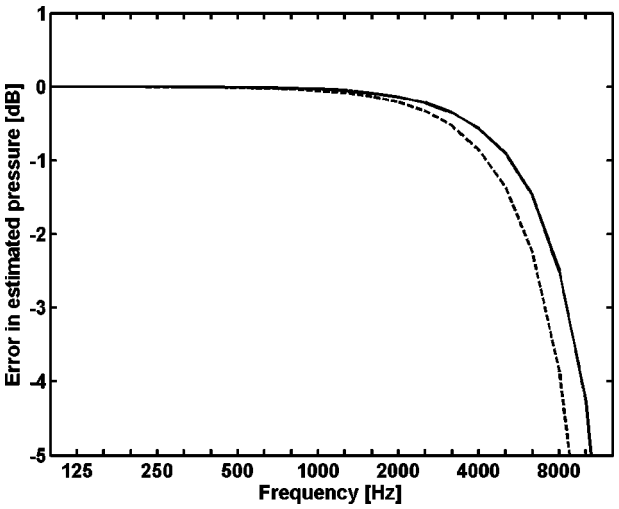


FIG. 1. Finite-difference approximation error in a plane wave of axial incidence with a microphone separation distance of 12 mm.—, sound intensity; ---, sound pressure.

to the latter approximation). For a plane wave of axial incidence the finite-difference error in the estimated intensity (the ratio of the estimate to the true intensity ignoring scattering and diffraction) can be shown to be[2]

$$\hat{I}_r/I_r = \frac{\sin(k\Delta r)}{k\Delta r}, \tag{4}$$

whereas the finite-difference error in the estimated pressure is[6]

$$\hat{p}/p = \cos(k\Delta r/2). \tag{5}$$

As can be seen from Fig. 1, both expressions indicate underestimation at high frequencies. However, the numerical and experimental study described in Ref. 3 demonstrated that for a wide range of sound field conditions scattering and diffraction effects almost completely cancel the finite-difference error of the intensity.

With a 12 mm spacer between the microphones Eq. (5) predicts a finite-difference error of the sound pressure of about 1 dB at 4.3 kHz in a plane wave of axial incidence. However, it may be expected that the increase of the pressure due to scattering and diffraction will compensate for the finite-difference error at least to some extent.

III. A BOUNDARY ELEMENT MODEL

The boundary element method (BEM) based on the Helmholtz integral equation is particularly suited for solving acoustic scattering and diffraction problems. The present work evolves from earlier work. In Ref. 7 an axisymmetric BEM model of a condenser microphone was presented, and it was shown how the outcome of a BEM calculation, which both then and in the present case is a pressure distribution over the ''diaphragm'' of the ''microphone,'' could be related to the electrical output of the device relative to its pressure-field response (i.e., the free-field correction was calculated). At frequencies more than an octave below the resonance of the microphone, which is the frequency range considered here, it was found that the difference between the pressure-field response and the free-field response was almost exclusively due to diffraction. The effect due to the finite impedance of the diaphragm was found to be less than 0.1 dB in this frequency range, and this effect is therefore ignored in the present work. Although the model of the microphone is axisymmetric, nonaxisymmetric sound fields (such as plane wave incidence from other directions than the microphones' axis of rotation) can be handled by taking advantage of the fact that a nonaxisymmetric diaphragm movement does not result in an electrical signal from the microphone.[7] The accuracy of the BEM calculation for a single microphone was established by comparing the calculations with very accurate experimental results.[7]

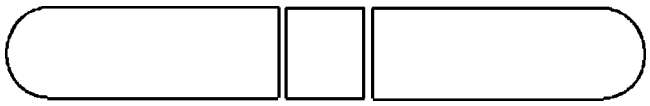


FIG. 2. Geometry of sound intensity probe used in the numerical model.

The BEM model was extended to calculations for an intensity probe in Refs. 3, 8, and 9, and this model has been used in the present work. The probe is modeled as a cylinder with a diameter of 12.7 mm, rounded at both ends. Each ''microphone'' is 31.8 mm long; there is a 1.1-mm gap between the ''diaphragm'' and the spacer; and the ''12 mm spacer'' is actually 10.9 mm long—see Fig. 2. The two microphones and the spacer were modeled using 69 quadratic elements and 141 nodes. At 12.5 kHz, which is the upper frequency limit of the present study, this corresponds to about 22 nodes per wavelength, which is safely above the rule of thumb of using six nodes per wavelength. All frequencies considered in the following are below the first characteristic frequency of the boundary element formulation,[10,11] and therefore no special care has been taken to avoid the nonuniqueness problem.

The model of an intensity probe may give rise to numerical problems because of the close surfaces of the microphones and the spacer. This problem was addressed in Ref. 8, where a method of dealing with such close surfaces was presented and validated. The use of the model for simulation of intensity probe in sound fields of practical relevance was presented in Ref. 3, in which the predictions of the numerical model were found to be in very good agreement with experimental results.

Figure 3, which is adapted from Ref. 12, presents a comparison between measured values of the weighted pressure increase at the two microphones in the probe (the two free-field corrections) and the corresponding calculations for a plane wave at axial incidence. It can be seen that there is excellent agreement between the measurements and the calculations—the maximum deviation is about 0.1 dB.

One of the purposes of the spacer is to ensure a well-defined acoustic distance between the acoustic centers of the two microphones. For a plane wave at axial incidence the phase difference between the two pressure signals should ideally be

$$\Delta\varphi = k\Delta r, \tag{6}$$

where Δr, the (acoustic) separation distance, obviously

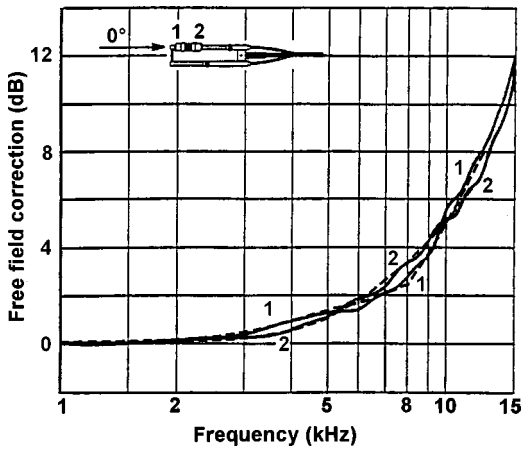


FIG. 3. Free-field correction of the two microphones of a sound intensity probe with 0.5 in. microphones separated by a 12 mm spacer for axial plane wave incidence.—, experimental results reproduced from Ref. 12; ---, numerical results.

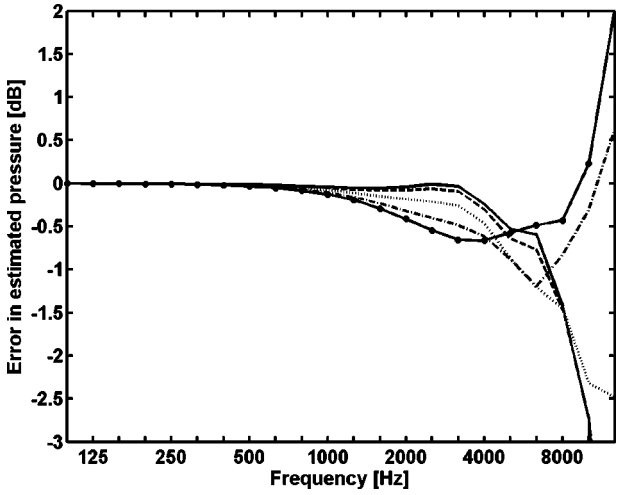


FIG. 5. Error in the estimate of the sound pressure for a plane wave at various angles of incidence:—, 0°; ---, 20°; ···, 40°; -·-·-, 60°; –●–, 80°.

should be independent of the frequency. In Fig. 4, which is adapted from Ref. 12, the effective separation distance predicted by the calculations is compared with measured values. The figure shows twice the value of Δr, and the agreement between measurements and calculations is very good: the maximum deviation occurs at 6 kHz and is about 0.5 mm, which corresponds to a relative deviation of about 2% between the predicted and measured phase difference.

IV. NUMERICAL RESULTS

The results presented in what follows have been calculated for an intensity probe with two 0.5 in. microphones separated by a 12 mm spacer (see Fig. 2). The calculations have been carried out at the one-third octave center frequencies from 100 Hz to 12.5 kHz.

A. The conventional sound pressure estimate

Figure 5 shows the combined effect of the finite-difference error and diffraction on the estimated sound pressure for a plane wave at various angles of incidence. The model of the probe is perfectly symmetrical about a plane though the middle of the spacer with a normal along the axis of rotation. Because of this symmetry the response at a certain angle of incidence, θ (where $\theta=0$ indicates axial incidence), is identical with the response at the supplementary angle of incidence, $\pi-\theta$; therefore angles of incidence only up to 80° are shown. It is clear from Fig. 5 that the pressure increase due to diffraction partly compensates for the finite-difference error—for axial incidence the 1 dB limit is reached just below 8 kHz (compare with Fig. 1). It is also apparent that at 10 kHz, which is the upper limiting frequency for sound intensity measurements with a probe with 0.5 in. microphones separated by a 12 mm spacer,[3] the sound pressure of a plane wave of axial incidence will be underestimated by about 3 dB.

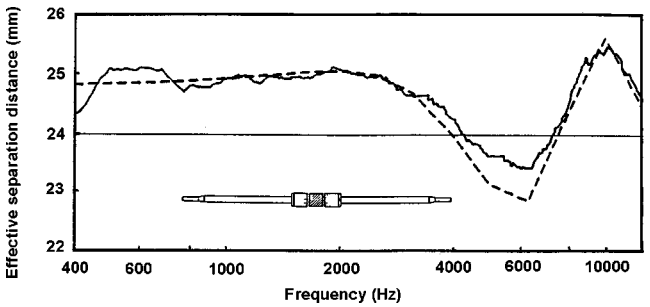


FIG. 4. Effective separation distance (multiplied by 2) between the two microphones of a sound intensity probe with 0.5 in. microphones separated by a 12 mm spacer for axial plane wave incidence.—, experimental results reproduced from Ref. 12; ---, numerical results.

B. An improved estimate

Although the estimate of the pressure given by Eq. (1) works very well in estimating the intensity as shown in Ref. 3, it might be possible to develop a better estimate for the pressure itself. A simple estimate could be based on a weighted average, as follows:

$$\hat{p}_w=(w\hat{p}_1+(1-w)\hat{p}_2), \quad w\in[0;1]. \tag{7}$$

The sum of the two weighting factors must obviously equal unity for the estimate to be valid at low frequencies. In order to find an optimum value for w a function to be minimized must be established. Such a function could be

$$G(w)=\int_0^{\pi/2}|\hat{p}_w|f_w(\theta)d\theta/\left(|p|\int_0^{\pi/2}f_w(\theta)d\theta\right), \tag{8}$$

which ideally should equal one (p is the sound pressure in the undisturbed field). Because of the symmetry of the probe and the fact that two pressure responses from the individual microphones are almost identical, the estimated sound pressure for a plane wave incident at a certain angle, θ, is almost identical with the estimated sound pressure at $\pi-\theta$, in spite of the different weighting of the two signals. The weighting factor $f_w(\theta)$ gives the possibility of favoring estimates that work particularly well in a certain direction, for example, near axial incidence, and this may be advantageous since the probe is pointing towards the sound source in typical sound power measurements. Figure 6 shows the maximum absolute value of $20\log(G(w))$ in the frequency range from 2 kHz to

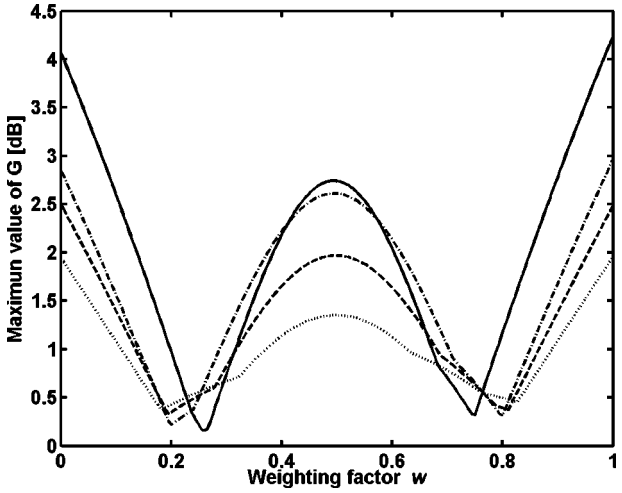


FIG. 6. Maximum value of the logarithmic magnitude of the cost function G in the frequency range 2–10 kHz as a function of the weighting factor w. Weighting function:—, $\delta(\theta)$; ---, $\cos(\theta)$; ⋯, 1; -·-·-, $1-H(\theta-\pi/4)$.

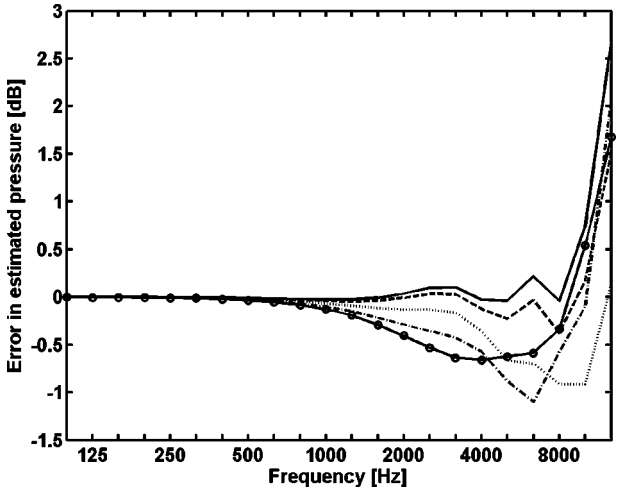


FIG. 7. Error in the estimated sound pressure with $w=0.775$ for a plane wave at various angles of incidence:—, 0°; ---, 20°; ⋯, 40°; -·-·-, 60°; –○–, 80°.

10 kHz as a function of the weighting parameter w calculated for several different weighting functions. At frequencies below 2 kHz all estimates give good results.

Four different weighting functions are examined in Fig. 6. Because the two sound pressure signals are very similar all the curves are almost symmetrical around $w=\frac{1}{2}$. It can be seen that the conventional estimate of the sound pressure corresponding to $w=\frac{1}{2}$ is not the optimal estimate with any of the weighting functions. The solid curve has been calculated with the weighting $\delta(\theta)$, which corresponds to optimizing the response for axial incidence. The dashed line corresponds to a weighting of $\cos(\theta)$, which favors near-axial incidence in the optimization; the dotted curve corresponds to an equal weighting of all angles of incidence between 0 and $\pi/2$; and finally the dashed–dotted curve corresponds to an equal weighting of all angles between 0 and $\pi/4$ in the optimization (H is the Heavyside function). It is apparent that the optimum value of w depends on the sound field, as one would expect. However, it can also be seen that values of w around 0.2 or 0.8 improve the estimate of the sound pressure significantly for all weighting functions compared with the conventional estimate with $w=\frac{1}{2}$.

Figure 7 shows the error of the pressure estimate with $w=0.775$, which seems to be close to the optimum value for all the weighting functions considered here. The figure demonstrates that the estimate with $w=0.775$ performs significantly better than the conventional one (compare with Fig. 5) in the sound fields examined here. In the frequency range of concern (up to 10 kHz) the error is less than 1.1 dB, and for near-axial incidence the error is less that 0.2 dB in the frequency range up to 8 kHz. Moreover, both positive and negative errors occur, and this will probably cause the errors to cancel out at least to some extent in a typical measurement situation with a sound intensity probe.

It is interesting to compare the performance of the improved estimate with that of a single microphone of the same dimensions, *not* mounted on a sound intensity probe. Figure 8 shows such a comparison for axial incidence, for incidence perpendicular to the probe, and for random (diffuse) incidence. The diffuse-field response has been calculated as a weighted average of the squared response at all angles of incidence. It is apparent that the intensity probe with the improved sound pressure estimate is much more omnidirectional than an ordinary microphone.

Sound intensity probes are sometimes used relatively near sources of sound. Figures 9(a), 9(b), and 9(c) show the error in the estimated sound pressure generated by a point source placed on the axis of the probe at distances between 10 and 100 cm from its center, calculated with $w=0.225$, $w=0.775$, and $w=0.5$. It is apparent that $w=0.225$ leads to underestimation and $w=0.775$ leads to overestimation if the point source is near the probe, in agreement with the fact that a weighting factor of 1 simply shifts the measurement position 6 mm closer to the source whereas a factor of 0 does the opposite. In this respect the conventional estimate with $w=0.5$ is clearly better, although even this one is not perfect. In all cases the errors are essentially independent of the frequency.

The measured pressure-intensity index is normally used

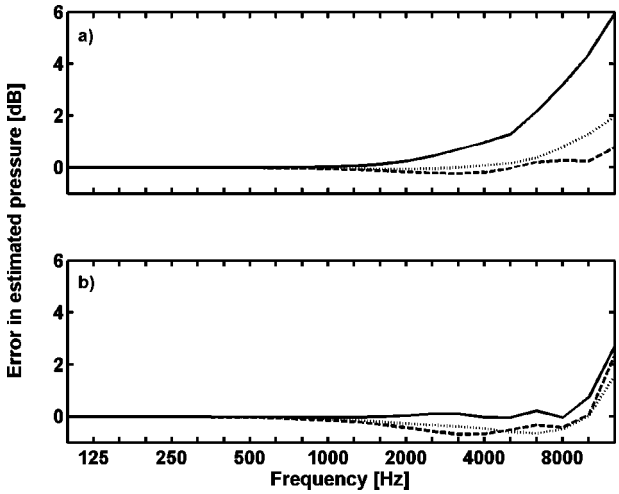


FIG. 8. Error in the estimated sound pressure calculated for (a) a solitary 0.5 in. microphone, and (b) the optimized sound intensity probe, with $w=0.775$.—, 0°; ---, 90°; ⋯, diffuse incidence.

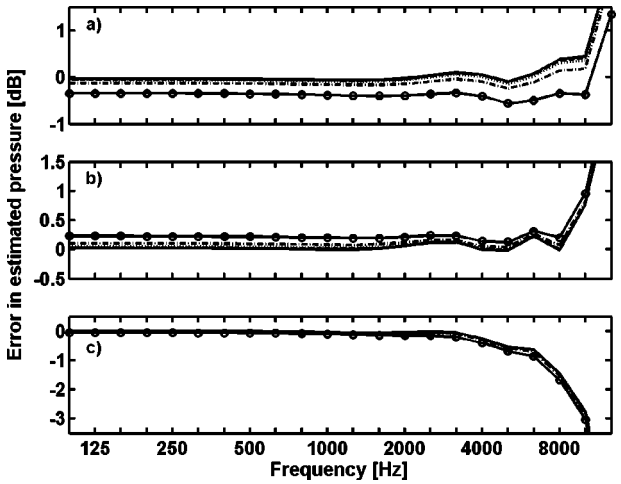


FIG. 9. Error in the estimated sound pressure generated by a point source, estimated with (a) $w=0.225$, (b) $w=0.775$, and (c) $w=0.5$. Distance between the point source and the center of the intensity probe:—, 1 m; ---, 0.75 m; ···, 0.5 m; -·-·-, 0.25 m; –○–, 0.1 m.

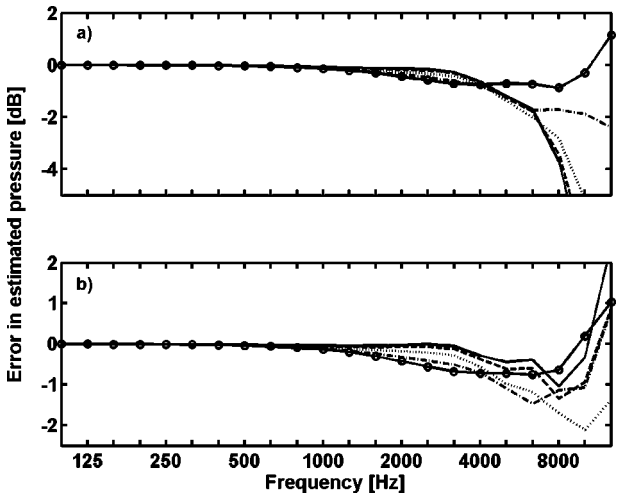


FIG. 10. Error in (a) the conventional estimate of the sound pressure ($w=0.5$) and (b) in the improved estimate ($w=0.775$) for a plane wave at various angles of incidence when phase mismatch corresponding to unmatched microphones has been introduced:—, 0°; ---, 20°; ···, 40°; -·-·-, 60°; –○–, 80°.

for giving an indication of phase mismatch errors in the sound intensity, owing to the fact that phase mismatch gives rise to an error in the measured intensity that is proportional to the mean square pressure.[1] In the circumstances the pressure estimate with $w=0.775$, which overestimates the pressure slightly when the source is near and thus gives a conservative estimate of the phase mismatch error, is a better choice than $w=0.225$. Nearby sources *behind* the intensity probe are unlikely to occur in normal use.

Microphones for measurement of sound intensity must be phase matched fairly 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).[13–15] This is a consequence of the fact that phase mismatch in most of the frequency range is caused by differences between the resonance frequencies and the damping of the two microphones.[16]

The influence of the phase mismatch on measurement of sound pressure with an intensity probe has been examined by artificially changing the phase of one of the pressure signals. The influence of the phase error of a matched microphone set is completely negligible. Figures 10(a) and 10(b), which correspond to Figs. 5 and 7, show the influence of the typical phase mismatch of *unmatched* microphones (25° at 10 kHz) for $w=0.5$ and $w=0.775$. As can be seen the influence of such phase mismatch is fairly moderate, and the improved estimate is less sensitive than the conventional one. This is in good agreement with the fact that the effect of the phase error must disappear altogether when $w=0$.

C. Discussion

The main purpose of this work has been to examine whether it is possible to measure the sound pressure and thus the pressure-intensity index reliably in the frequency range where the sound intensity itself can be measured reliably, that is, up to 10 kHz for an intensity probe with two 0.5 in. microphones separated by a 12 mm spacer. It seems to be possible to improve the conventional estimate considerably in a very simple manner.

The largest error of the improved estimate occurs at 6.3 kHz at an angle of incidence of 60°. Under such conditions the sound pressure will be underestimated by about 1 dB. However, the conventional method underestimates the sound pressure by 1.2 dB under the same conditions and is much worse than the improved one for axial incidence.

It is not possible to improve the performance for perpendicular incidence, simply because the two microphone signals are identical in phase as well as in amplitude. By contrast, it *is* possible to improve the estimate in a diffuse sound field. Although the two pressure signals have the same magnitude in a diffuse sound field and the resulting phase difference is zero the signals are not completely coherent.[17] In such a sound field $w=0.14$ gives slightly better results than $w=0.775$. However, the improvement is modest indeed (less than 0.2 dB), and the performance for axial incidence is worsened significantly: the maximum error is almost doubled. A weighting with $w=0.775$ seems to be the best overall solution.

The spacer length assumed in the calculations has been optimized for measurement of sound intensity, not sound pressure. In some respects it is more difficult to measure the sound pressure at high frequencies: as demonstrated by Fig. 1 the finite-difference error is larger, and whereas large relative errors in the estimated intensity are not very important near perpendicular incidence all directions are equally important in sound pressure measurements. Increasing the length of the spacer would not have any influence on the estimated pressure when the probe is exposed to perpendicular sound incidence, but it would reduce the estimated pressure at high frequencies for sound incidence near the axis of the probe. Such considerations and the results presented in Figs. 7 and 8 lead to the conclusion that it is unlikely that the performance could be improved appreciably by choosing an-

other spacer length—quite apart from the fact that this would be unacceptable if the probe is to be used for measurement of sound intensity.

The results have indicated that a two-microphone arrangement could be advantageous even in ordinary sound pressure measurements. It may seem paradoxical to suggest measuring sound pressure with an intensity probe. The conventional solution in ordinary sound pressure measurements is to use microphones designed to have a flat or nearly flat frequency response under conditions that are similar to those of the measurement. For example, ''free-field microphones'' are designed to have a flat response for axial incidence (at the expense of the response under other conditions),[18] and ''random incidence microphones'' are designed to have a flat response in a diffuse sound field.[18] However, the comparison with an ordinary microphone (one not mounted on an intensity probe) has demonstrated that a two-microphone arrangement with the proposed weighting is far better in the general case. Thus such a two-microphone arrangement may have a potential in conventional sound pressure measurements not related with sound intensity. A two-microphone sound level meter would be far simpler than a sound intensity analyzer and would not require phase-matched microphones.

V. EXPERIMENTAL RESULTS

To examine the validity of the numerical results presented in the foregoing a few experiments have been carried out. A frequency analyzer of type Brüel & Kjær 3550 was used in combination with a sound intensity probe of type B&K 3548, either with 0.5 in. microphones of type B&K 4181 and a 12 mm spacer or with a 0.25 in. microphone of type B&K 4178 on one of the 0.25 in. preamplifiers taken out of the probe. Both microphone sets are free-field microphones. Therefore, it is necessary to compensate for the drop in the pressure sensitivities at high frequencies when the pressure response is needed. This was done using frequency responses measured with an electrostatic actuator, as described in Ref. 3. The weighted estimate of the sound pressure level was calculated from the two power spectra S_{11} and S_{22} and the real part of the cross spectrum C_{12} as follows:

$$L_p(\omega_0) = 10\log(\overline{p_w^2}(\omega_0)), \tag{9}$$

where

$$\overline{p_w^2}(\omega_0) = \frac{w^2}{\pi}\int_{\omega_0-\Delta\omega/2}^{\omega_0+\Delta\omega/2} S_{11}(\omega)\,d\omega + \frac{(1-w)^2}{\pi}\int_{\omega_0-\Delta\omega/2}^{\omega_0+\Delta\omega/2} S_{22}(\omega)\,d\omega + \frac{2w(1-w)}{\pi}\int_{\omega_0-\Delta\omega/2}^{\omega_0+\Delta\omega/2} C_{12}(\omega)\,d\omega. \tag{10}$$

Here the power spectra and the cross power spectrum are two-sided spectra, ω_0 is the center frequency, and $\Delta\omega$ is the bandwidth.

The experimental results shown in Fig. 11 have been determined in an anechoic room about 2 m from a loudspeaker driven with white noise. The ''true'' sound pressure was measured using the 0.25 in. microphone of type B&K 4178, *not* corrected for its pressure response, pointing towards the loudspeaker. Figure 11(a) compares calculated and measured pressure responses for axial incidence using the conventional weighting factor $w=0.5$ and the proposed one, $w=0.775$. There is agreement within ± 1 dB, and the superiority of the improved estimate is apparent. The calculated response of the front microphone is also shown.

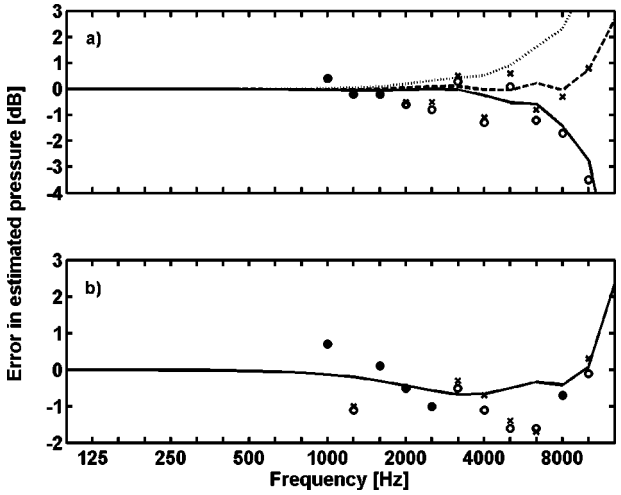


FIG. 11. Error in the estimated sound pressure for (a) axial incidence and (b) perpendicular incidence.—, conventional estimate ($w=0.5$); ○, corresponding measurement; ---, improved estimate ($w=0.775$); ×, corresponding measurement; ⋯, front microphone ($w=1$).

Figure 11(b) shows a similar comparison for perpendicular incidence. In this case the three calculated curves coincide, and the experimental results are also almost unaffected by the weighting factor, which indicates that the physical sound intensity probe is almost symmetrical. The agreement between measurements and calculations is within ± 1.2 dB.

The experimental results shown in Fig. 12 have been obtained in a large reverberation room about 5 m from the

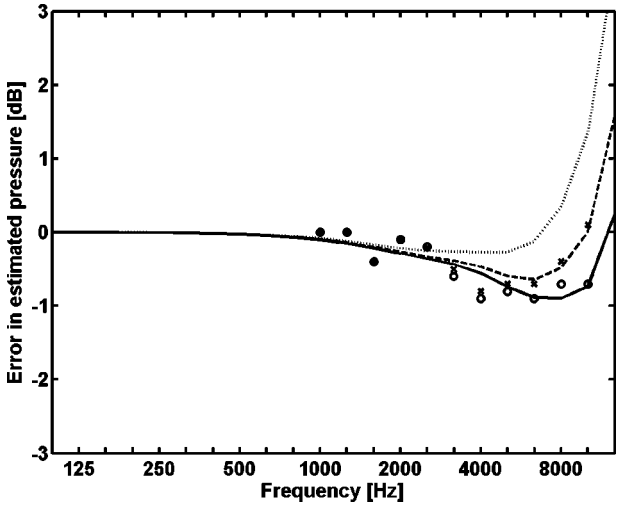


FIG. 12. Error in the estimated sound pressure for diffuse incidence.—, conventional estimate ($w=0.5$); ○, corresponding measurement; ---, improved estimate ($w=0.775$); ×, corresponding measurement; ⋯, front microphone ($w=1$).

loudspeaker. In this case the ''true'' sound pressure was measured with the 0.25 in. microphone of type B&K 4178, this time corrected for its pressure response. This should be accurate up to 10 kHz, as can be seen by scaling the diffuse-field response of 0.5 in. microphones shown in Fig. 8. Figure 12 compares calculated and measured pressure responses using the conventional weighting factor $w=0.5$ and the proposed one, $w=0.775$. There is agreement within ± 0.5 dB. The calculated response of the front microphone is also shown.

VI. CONCLUSIONS

A numerical and experimental study of measurement of sound pressure with two-microphone sound intensity probes has demonstrated that the conventional estimate based on the arithmetic mean of the two pressure signals can be improved significantly by applying different weighting factors on the two signals. The optimum length of the spacer between the microphones for sound intensity measurements is about one microphone diameter. The corresponding optimum values of the weighting factors for sound pressure measurements are about 0.2 and 0.8, respectively, and the latter is slightly better when the source is near the probe. With 0.5 in. microphones and a 12 mm spacer the improved estimate gives tolerable accuracy under a variety of sound field conditions up to 10 kHz, which is about an octave higher than the conventional estimate.

The proposed improvement makes an intensity probe with 0.5 in. microphones more omnidirectional than a single 0.25 in. microphone and may therefore be useful also in sound pressure measurements not related to sound intensity.

[1] F. J. Fahy, *Sound Intensity*, 2nd ed. (E & FN Spon, London, 1995), See sections 5.3, 6.2, 7.2, and 9.3.

[2] G. Rasmussen and M. Brock, ''Acoustic intensity measurement probe,'' in *Proceedings of Recent Developments in Acoustic Intensity Measurement*, edited by M. Bockhoff (Senlis, France, 1981), pp. 81–88.

[3] 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).

[4] F. Jacobsen, ''A simple and effective correction for phase mismatch in intensity probes,'' Appl. Acoust. **33**, 165–180 (1991).

[5] S. E. Keith, G. Krishnappa, V. Chiu, ''The potential for extending the frequency range of sound intensity standards,'' in *Proceedings of the Fifth European Conference on Noise Control*, edited by G. Brambilla, C. Ianiello, and L. Maffei (Naples, Italy, 2003), Paper 169-IP.

[6] T. J. Schultz, ''Acoustic wattmeter,'' J. Acoust. Soc. Am. **28**, 693–699 (1956).

[7] P. Juhl, ''A numerical investigation of standard condenser microphones,'' J. Sound Vib. **177**, 433–446 (1994).

[8] V. Cutanada, P. M. Juhl, and F. Jacobsen, ''On the modeling of narrow gaps using the standard boundary element method,'' J. Acoust. Soc. Am. **109**, 1296–1303 (2001).

[9] V. Cutanda, P. M. Juhl, and F. Jacobsen, ''A numerical investigation of the performance of sound intensity probes at high frequencies,'' in *Proceedings of Fourth International Congress on Sound and Vibration*, edited by M. J. Crocker (St. Petersburg, Russia, 1996), pp. 1897–1904.

[10] H. A. Schenck, ''Improved integral formulation for acoustic radiation problems,'' J. Acoust. Soc. Am. **44**, 41–58 (1968).

[11] P. Juhl, ''A numerical study of the coefficient matrix of the boundary element method near characteristic frequencies,'' J. Sound Vib. **175**, 39–50 (1994).

[12] E. Frederiksen and M. Piil, ''Characteristics of microphone pairs and probes for sound intensity measurements,'' Brüel & Kjærum, Report BA077-11, 1987.

[13] International Electrotechnical Commission (IEC) Standard 61043, ''Instruments for the measurement of sound intensity—Measurements with pairs of pressure sensing microphones,'' 1993.

[14] American National Standard ANSI S1.9-1996, ''Instruments for the measurement of sound intensity,'' 1996.

[15] ''Product data, sound intensity pair—type 4197,'' Brüel & Kjær, Nærum, Denmark, 2000.

[16] E. Frederiksen and O. Schultz, ''Pressure microphones for intensity measurements with significantly improved phase properties,'' Brüel & Kjær Technical Review **4**, 11–19 (1986).

[17] F. Jacobsen and T. Roisin, ''The coherence of reverberant sound fields,'' J. Acoust. Soc. Am. **108**, 204–210 (2000).

[18] R. W. Krug, ''Sound level meters,'' in *Handbook of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1998), Chap. 105.

Constant-beamwidth and constant-powerwidth wideband robust Capon beamformers for acoustic imaging

Zhisong Wang and Jian Li[a)]
Department of Electrical and Computer Engineering, P.O. Box 116130, University of Florida, Gainesville, Florida 32611

Petre Stoica
Division of Systems and Control, Department of Information Technology, Uppsala University, P.O. Box 337, SE-75105 Uppsala, Sweden

Toshikazu Nishida
Department of Electrical and Computer Engineering, P.O. Box 116200, University of Florida, Gainesville, Florida 32611

Mark Sheplak
Department of Mechanical and Aerospace Engineering, P.O. Box 116250, University of Florida, Gainesville, Florida 32611

(Received 12 December 2003; revised 21 March 2004; accepted 25 March 2004)

The standard Capon beamformer (SCB) is sensitive to the mismatch between the assumed and actual array steering vector, which occurs often in practice. Recently a robust Capon beamformer (RCB) was proposed by extending the SCB to the case of uncertain array steering vectors. In certain applications such as acoustic imaging, it is desirable that the beamwidth is constant across the frequency bins. This prevents future corrections for different frequencies and contributes to consistent sound pressure level (SPL) estimation, which means that for an acoustic wideband monopole source with a flat spectrum the acoustic image for each frequency bin stays the same. However, the beamwidth of RCB decreases with frequency, which can lead to inconsistent imaging results across the frequencies. In this paper two beamformers are proposed, namely a constant-beamwidth robust Capon beamformer (CBRCB) and a constant-powerwidth robust Capon beamformer (CPRCB), as extensions of RCB for consistent wideband acoustic imaging. Both CBRCB and CPRCB are more robust against array steering vector errors and finite sample size problems than SCB and have better resolution and interference suppression capability than data-independent beamformers. Moreover, they both can be efficiently implemented. The effectiveness of CBRCB and CPRCB is demonstrated via a number of simulated and experimental examples. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1744751]

PACS numbers: 43.60.Fg, 43.60.Lq, 43.60.Mn [EJS] Pages: 1621–1631

I. INTRODUCTION

The standard Capon beamformer (SCB) (Capon, 1969; Stoica and Moses, 1997) has better resolution and much better interference rejection capability than data-independent beamformers, provided that the array steering vector corresponding to the signal of interest (SOI) is accurately known. However, there are many factors in practice that can degrade the performance of an adaptive beamformer, such as SCB, including look direction errors due to inaccurate knowledge of the source location, array calibration errors due to imprecise knowledge of the microphone responses and positions, and array covariance matrix estimation errors due to a small sample size. Owing to the previous problems, the performance of the Capon beamformer may become worse than that of the standard data-independent beamformers, such as the delay-and-sum (DAS) beamformer. Consequently, Capon beamformers have not been used very much in practice despite their potential benefits. For example, in Humphreys *et al.* (1998) to avoid the potential instability problems of the adaptive methods, the DAS method was chosen instead for acoustic imaging in an anechoic facility using a two-dimensional (2-D) microphone directional array.

Much work has been done to improve the robustness of the Capon beamformer over the past three decades and the literature on robust adaptive beamforming is extensive (see, for example, Van Trees, 2002; Vorobyov *et al.*, 2003; Lorenz and Boyd, 2001; Li *et al.*, 2003, and the many references therein). However, most of the early suggested methods do not *directly* address the problem of the uncertainty of the array steering vector. Only recently have several methods with a clear theoretical background been proposed to tackle the problem of various mismatches in the array steering vector and covariance matrix (see e.g., Vorobyov *et al.*, 2003; Lorenz and Boyd, 2001; Li *et al.*, 2003; Stoica *et al.*, 2003; Wu and Zhang, 1999), which, unlike the early methods, make explicit use of an uncertainty set for the array steering vector. In Wu and Zhang (1999), a polyhedron was used to describe the uncertainty set, whereas spherical and ellipsoi-

a)Author to whom all correspondence should be addressed. Electronic mail: li@dsp.ufl.edu

dal (including flat ellipsoidal) uncertainty sets were considered in Vorobyov *et al.* (2003), Lorenz and Boyd (2001), Li *et al.* (2003), and Stoica *et al.* (2003). The robust Capon beamforming approaches presented in Vorobyov *et al.* (2003) and Lorenz and Boyd (2001) couple the formulation of the original SCB in Capon (1969) with a spherical or ellipsoidal uncertainty set of the array steering vector. The RCB approach in Li *et al.* (2003) and Stoica *et al.* (2003), on the other hand, couples the formulation of the covariance fitting based SCB in Marzetta (1983) with an ellipsoidal uncertainty set. In addition, the RCB comprises a simple way of eliminating the scaling ambiguity when estimating the power of the desired signal while the approaches in Vorobyov *et al.* (2003) and Lorenz and Boyd (2001) did not consider the scaling ambiguity problem.

The aforementioned robust adaptive beamforming approaches were devised under the narrow-band assumption. However, wideband signals are encountered in many applications including, for example, acoustic imaging using microphone arrays (Humphreys *et al.*, 1998; Brooks and Humphreys, 1999; Brooks *et al.*, 1987). For wideband signals, we can use the fast Fourier transform (FFT) to divide the array outputs into many narrow-band frequency bins and apply the narrow-band RCB to each bin separately. However, it is well known that as the frequency increases, the beamwidth of both data-independent and data-adaptive beamformers decreases. This beamwidth variation as a function of frequency will subject the signals incident on the outer portions of the main beam to low-pass filtering and lead to distorted signal spectra or inaccurate SOI power estimation (Goodwin and Elko, 1993; Chou, 1995; Lardies, 1989). Hence it is desirable that the beamwidth of a beamformer remains approximately constant over all frequency bins of interest. In fact, a constant-beamwidth beamformer is desirable in many applications including ultrasonics, underwater acoustics, acoustic imaging and communications, and speech acquisition (Humphreys *et al.*, 1998; Goodwin and Elko, 1993; Lardies, 1989). In the past five decades, many approaches have been proposed to obtain constant-beamwidth beamformers including harmonic nesting (Humphreys *et al.*, 1998; Brooks and Humphreys, 1999; Brooks *et al.*, 1987; Chou, 1995), multibeam (Goodwin and Elko, 1993; Tucker, 1957; Smith, 1970), asymptotic theory based methods (Doles III and Benedict, 1988), and approximation of a continuously distributed sensor via a finite set of discrete sensors (Ward *et al.*, 1995; Abhayapala *et al.*, 2000; Brandstein and Ward, 2001). Among these approaches, harmonic nesting is commonly used for acoustic imaging via microphone arrays. For example, in Humphreys *et al.* (1998), Brooks and Humphreys (1999), and Brooks *et al.* (1987) a shading scheme is used for a directional array consisting of a set of harmonically nested subarrays, each of which is designed for a particular frequency bin. For each array element, shading weights are devised as a function of the frequency. This shading scheme can provide a constant beamwidth for frequencies between 10 and 40 kHz when used with the delay-and-sum (DAS) beamformer. Hereafter, this approach will be referred to as the shaded DAS (SDAS).

In this paper, we show that we can achieve a constant beamwidth across the frequency bins for an adaptive beamformer, by combining our RCB with the shading scheme devised in Humphreys *et al.* (1998), Brooks and Humphreys (1999), and Brooks *et al.* (1987), provided that there are no strong interfering signals near the main beam of the array. We refer to this approach as the constant-beamwidth RCB (CBRCB) algorithm. We also show that we can attain a constant powerwidth, and hence consistent power estimates across the frequency bins, by using RCB with a frequency-dependent uncertainty parameter for the array steering vector; we refer to the so-obtained beamformer as the constant-powerwidth RCB (CPRCB). CBRCB and CPRCB can both be efficiently implemented at a comparable cost with that of SCB by using the Lagrange multiplier methodology. Finally, simulation-based and experimental results are presented to illustrate the effectiveness of CBRCB and CPRCB.

The rest of the paper is organized as follows. In Sec. II, we formulate the problem of interest. In Sec. III, we describe the RCB algorithm and extend it to obtain CPRCB. In Sec. IV, we present the CBRCB algorithm. Simulated and experimental examples comparing the performances of DAS, SCB, SDAS, RCB, CPRCB, and CBRCB are given in Sec. V. Finally, Sec. VI contains the conclusions.

II. PROBLEM FORMULATION

Our RCB approach (Li *et al.*, 2003; Stoica *et al.*, 2003) can be used in many applications including radar, sonar, acoustics, communications, geophysics, astrophysics, and medical imaging. We focus herein on forming acoustic images using a microphone array, which are obtained by determining the sound pressure estimates corresponding to the two- or three-dimensional coordinates of a grid of locations. The signal at each grid location of interest is referred to as the SOI. To determine the wideband SOI sound pressure, we first decompose each of the sensor outputs into many narrow-band frequency bins by applying the FFT. Then for each frequency bin of interest, we determine the corresponding sound pressure estimate. Summing the narrow-band SOI sound pressure estimates over all frequency bins of interest yields the sound pressure estimate of the wideband SOI.

Assume that a wideband SOI impinges on an array with M elements. We divide each sensor output into N nonoverlapping blocks with each block consisting of I samples. We then apply an I-point FFT to each block to obtain I narrow-band frequency bins. The data vector, $\mathbf{y}_i(n)$, for the ith frequency bin and the nth snapshot can be written as

$$\mathbf{y}_i(n)=\mathbf{a}_i(\mathbf{x}_0)s_i(n)+\mathbf{e}_i(n), \quad n=1,...,N, \quad i=1,...,I, \quad (1)$$

where $s_i(n)$ stands for the complex-valued waveform of the SOI that is present at the location coordinate $\mathbf{x}_0$ and the ith frequency bin, $\mathbf{e}_i(n)$ represents a complex $M\times 1$ noise plus interference vector for the coordinate $\mathbf{x}_0$ and the ith frequency bin, and $\mathbf{a}_i(\mathbf{x}_0)$ is the SOI array steering vector, which depends on both $\mathbf{x}_0$ and the ith frequency bin and which satisfies

$$\|\mathbf{a}_i(\mathbf{x}_0)\|^2=M, \quad (2)$$

with $\|\cdot\|$ denoting the Euclidean norm. The nominal or assumed array steering vector $\overline{\mathbf{a}}_i(\mathbf{x}_0)$ has the form

$$\overline{\mathbf{a}_i}(\mathbf{x}_0)=[e^{j2\pi f_i\tau_1}\ \ e^{j2\pi f_i\tau_2}\ \cdots\ e^{j2\pi f_i\tau_M}]^T, \tag{3}$$

where τ_m denotes the propagation time delay between the source at $\mathbf{x}_0$ and the mth microphone, f_i is the center of the ith frequency bin, and $(\cdot)^T$ denotes the transpose. The covariance matrix for the ith frequency bin can be written as

$$\mathbf{R}_i=E[\mathbf{y}_i(n)\mathbf{y}_i^*(n)],\quad i=1,...,I, \tag{4}$$

where $E[\cdot]$ is the expectation operator, and $(\cdot)^*$ denotes the conjugate transpose. In practical applications, $\mathbf{R}_i$ is replaced by the sample covariance matrix $\hat{\mathbf{R}}_i$ [referred to as the cross spectral matrix in Humphreys *et al.* (1998) and Brooks and Humphreys (1999)], where

$$\hat{\mathbf{R}}_i=\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_i(n)\mathbf{y}_i^*(n),\quad i=1,...,I. \tag{5}$$

In aeroacoustic measurements using arrays, the sound pressure response is normally shown. The intensity of the sound pressure response is measured on a logarithmic scale referred to as the sound pressure level (SPL), which is defined as (Blackstock, 2000)

$$\mathrm{SPL}=20\log_{10}(p_{\mathrm{rms}}/p_{\mathrm{ref}})\ \ \mathrm{dB}, \tag{6}$$

where p_{rms} denotes the root-mean-squared pressure in Pa and p_{ref} stands for the reference pressure. For air, the reference pressure is 20 μPa corresponding to the hearing threshold of young people.

Next, we determine a scaling coefficient needed for calculating SPL estimates based on $\hat{\mathbf{R}}_i$. Let $y(l)$, $l=1,2,...,I$, represent a wideband time-domain sequence with I samples, and let $Y(i)$, $i=1,2,...,I$, denote the frequency-domain sequence obtained by applying an I-point FFT to $\{y(l)\}$. Let $y_i(l)$, $l=1,2,...,I$, be the narrow-band time-domain sequence corresponding to the ith frequency bin of the above frequency-domain sequence, in other words, $\{y_i(l)\}$ is obtained by using an inverse FFT on $\{0,...,0,Y(i),0,...,0\}$. Making use of only $Y(i)$, we can determine a root-mean-squared pressure estimate for the ith frequency bin, $\tilde{p}_{\mathrm{rms}}$, as follows:

$$\tilde{p}_{\mathrm{rms}}=\sqrt{\frac{1}{I}\sum_{l=1}^{I}|y_i(l)|^2}=\sqrt{\frac{1}{I^2}|Y(i)|^2}. \tag{7}$$

The second equality is due to the fact that $\Sigma_{l=1}^{I}|y_i(l)|^2=(1/I)|Y(i)|^2$ according to Parseval's theorem. Substituting Eq. (7) into Eq. (6), the SPL estimate for the ith narrow-band frequency bin can be written as $10\log_{10}(|Y(i)|^2/(p_{\mathrm{ref}}^2I^2))$ in dB. Therefore, a scaling coefficient, $1/I^2$, should be used when estimating the sound pressure from the data in the frequency domain. Let σ_i denote the root-mean-squared pressure estimate for the ith narrow-band frequency bin obtained using $\hat{\mathbf{R}}_i$. Then, according to the previous discussion,

$$\mathrm{SPL}=10\log_{10}(\sigma_i^2/(p_{\mathrm{ref}}^2I^2))\ \ \mathrm{dB}. \tag{8}$$

We note in passing that in many applications other than acoustics, the intensities of the sources are often measured by power estimates, which, unlike the SPL estimates, do not require scaling by a reference pressure.

In this paper we distinguish between beampattern and powerpattern and make use of this distinction in the design of the robust constant-beamwidth and constant-powerwidth beamformers. We define the beampattern for the ith frequency bin as

$$\mathrm{BP}_i(\mathbf{x})=|\mathbf{a}_i^*(\mathbf{x})\mathbf{w}(\mathbf{x}_g)|^2, \tag{9}$$

where $\mathbf{w}(\mathbf{x}_g)$ denotes the beamformer's weight vector corresponding to a given location, $\mathbf{x}_g$, and $\mathbf{x}$ is varied to cover each location of interest. Next, we introduce the powerpattern, which at the ith frequency bin is defined as

$$\mathrm{PP}_i(\mathbf{x})=|\mathbf{a}_i^*(\mathbf{x}_g)\mathbf{w}(\mathbf{x})|^2. \tag{10}$$

We remark that the beampattern shows how the beamformer will pass the SOI and interfering signals when it is steered to $\mathbf{x}_g$, whereas the powerpattern shows how the beamformer will pass the signal at $\mathbf{x}_g$ when it is steered to $\mathbf{x}$. $\mathrm{PP}_i(\mathbf{x})$ can be used to measure approximately the normalized power responses corresponding to a series of locations, and hence it is named powerpattern. To see this, we assume that the theoretical covariance matrix for the ith frequency bin has the following form:

$$\mathbf{R}_i=\sigma_i^2\mathbf{a}_i(\mathbf{x}_0)\mathbf{a}_i^*(\mathbf{x}_0)+\mathbf{Q}, \tag{11}$$

where σ_i^2 denotes the SOI power and $\mathbf{Q}$ stands for the interference-plus-noise covariance matrix. If $\mathbf{x}_g=\mathbf{x}_0$ and the signal-to-interference-plus-noise-ratio (SINR) is high, then it follows that $\mathbf{w}^*(\mathbf{x})\mathbf{R}_i\mathbf{w}(\mathbf{x})\approx\sigma_i^2|\mathbf{w}^*(\mathbf{x})\mathbf{a}_i(\mathbf{x}_g)|^2\propto\mathrm{PP}_i(\mathbf{x})$. We next use an example to illustrate the concept of beampattern and powerpattern. Consider a function $|\mathbf{a}_i^*(x_1)\mathbf{w}(x_2)|$ of two coordinate variables, x_1 and x_2. Then the beampattern is a slice of this function for x_2 fixed, whereas the powerpattern is a slice for x_1 fixed. Note that the beampattern and powerpattern of the DAS are identical since its weight vector and steering vector have the same functional form. However, this is not the case for an adaptive beamformer due to the fact that its weight vector depends not only on the corresponding steering vector, but also on the data.

Without loss of generality, we consider herein two-dimensional (2-D) array imaging, in which the beamwidth or powerwidth is defined as the diameter of a circle having the same area as the 3-dB contour of the main lobe of a 2-D beampattern or powerpattern. The beamwidth shows how the nearby signals impact the estimation of SOI. The powerwidth, on the other hand, shows how SOI impacts the estimation of the nearby signals. *From now on, we will concentrate on the ith frequency bin.* For notational simplicity, whenever convenient, we will drop the subscript i and the dependence on $\mathbf{x}$ in the following sections.

III. CONSTANT-POWERWIDTH ROBUST CAPON BEAMFORMING

In this section we first present the RCB algorithm. We also explain some properties of RCB regarding its beamwidth and powerwidth, which are important for consistent wideband imaging. We then exploit these properties and extend RCB by choosing a frequency-dependent uncertainty parameter for the array steering vector to obtain CPRCB.

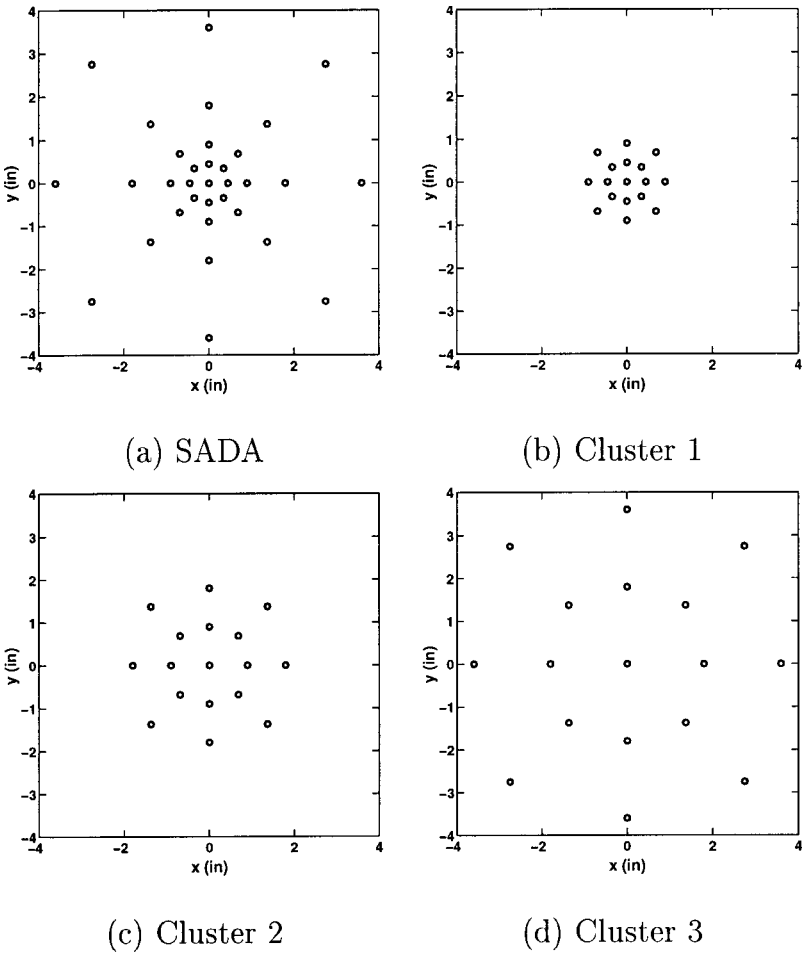


FIG. 1. Microphone layout of the small aperture directional array (SADA) and its three clusters.

The common formulation of the adaptive beamforming problem that leads to SCB, when the array steering vector is assumed to be $\overline{\mathbf{a}}$, is as follows (see, e.g., Capon, 1969; Stoica and Moses, 1997).

(a) Determine the $M\times 1$ vector $\mathbf{w}$ that is the solution to the following linearly constrained quadratic problem:

$$\min_{\mathbf{w}} \mathbf{w}^*\mathbf{R}\mathbf{w} \quad \text{subject to} \quad \mathbf{w}^*\overline{\mathbf{a}}=1. \tag{12}$$

(b) Use $\mathbf{w}^*\mathbf{R}\mathbf{w}$ as an estimate of the sound pressure σ^2.

The solution to Eq. (12) is easily derived:

$$\mathbf{w}=\frac{\mathbf{R}^{-1}\overline{\mathbf{a}}}{\overline{\mathbf{a}}^*\mathbf{R}^{-1}\overline{\mathbf{a}}}, \tag{13}$$

Using Eq. (13) in step (b) above yields the following sound pressure estimate:

$$\tilde{\sigma}^2=\frac{1}{\overline{\mathbf{a}}^*\mathbf{R}^{-1}\overline{\mathbf{a}}}. \tag{14}$$

In the recent papers (Li *et al.*, 2003; Stoica *et al.*, 2003), we have extended the Capon beamformer so as to be able to accurately determine the SPL of SOI even when only an imprecise knowledge of its actual steering vector, $\mathbf{a}_0$, is available. More specifically, we assume that the only knowledge we have about $\mathbf{a}_0$ is that it belongs to the following uncertainty sphere:

$$\{\mathbf{a}\,|\,\|\mathbf{a}-\overline{\mathbf{a}}\|^2\leqslant\varepsilon\}, \tag{15}$$

where $\overline{\mathbf{a}}$ and ε are given ($\overline{\mathbf{a}}$ is the assumed steering vector of SOI, and ε is a user parameter describing the uncertainty of the array steering vector).

To derive the robust Capon beamforming approach, we use the reformulation of the Capon beamforming problem in Marzetta (1983), to which we append the uncertainty set in Eq. (15) (see Li *et al.*, 2003; Stoica *et al.*, 2003). Proceeding in this way we *directly* obtain a robust estimate of σ_0^2, without any intermediate calculation of a vector $\mathbf{w}$:

$$\max_{\sigma^2,\mathbf{a}} \sigma^2 \quad \text{subject to} \quad \mathbf{R}-\sigma^2\mathbf{a}\mathbf{a}^*\geqslant 0,$$
$$\|\mathbf{a}-\overline{\mathbf{a}}\|^2\leqslant\varepsilon. \tag{16}$$

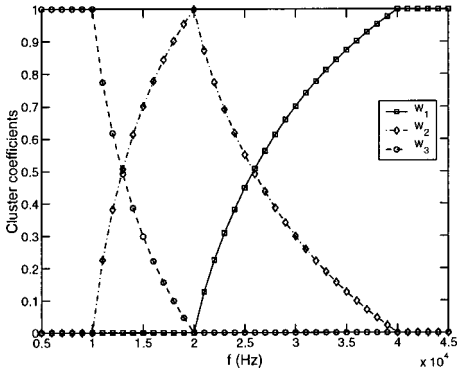


FIG. 2. Cluster shading weights for SADA, as functions of the frequency, with w_1, w_2, and w_3 corresponding to clusters 1, 2, and 3, respectively.

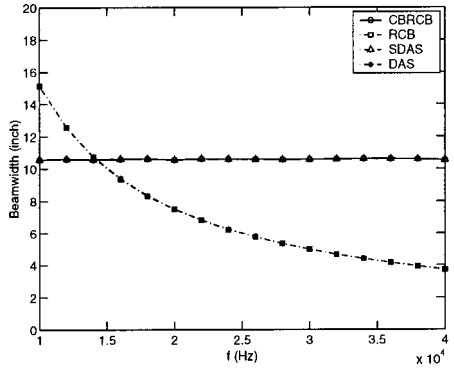


FIG. 3. Comparison of the beamwidths for the CBRCB, RCB, SDAS, and DAS methods with $N=64$. We used $\varepsilon=2.0$ for RCB and CBRCB.

Note that the first line above can be interpreted as a covariance fitting problem: given $\mathbf{R}$ and $\mathbf{a}$, we wish to determine the largest possible SOI term, $\sigma^2\mathbf{a}\mathbf{a}^*$, that can be a part of $\mathbf{R}$ under the natural constraint that the residual covariance matrix be positive semidefinite.

For any given $\mathbf{a}$, the solution $\hat{\sigma}_0^2$ to Eq. (16) is given by the counterpart of Eq. (14) with $\overline{\mathbf{a}}$ replaced by $\mathbf{a}$, as shown in Stoica *et al.* (2003). Hence Eq. (16) can be reduced to the following problem

$$\min_{\mathbf{a}} \mathbf{a}^*\mathbf{R}^{-1}\mathbf{a} \quad \text{subject to} \quad \|\mathbf{a}-\overline{\mathbf{a}}\|^2 \leqslant \varepsilon. \tag{17}$$

To exclude the trivial solution $\mathbf{a}=\mathbf{0}$ to (17), we assume that

$$\|\overline{\mathbf{a}}\|^2 > \varepsilon. \tag{18}$$

Because the solution to (17) [under (18)] will evidently occur on the boundary of the constraint set, we can reformulate (17) as the following quadratic problem with a quadratic equality constraint:

$$\min_{\mathbf{a}} \mathbf{a}^*\mathbf{R}^{-1}\mathbf{a} \quad \text{subject to} \quad \|\mathbf{a}-\overline{\mathbf{a}}\|^2 = \varepsilon. \tag{19}$$

The solution $\breve{\mathbf{a}}$ to the above optimization problem can be determined by using the Lagrange multiplier method and it has the form [see Li *et al.* (2003) and Stoica *et al.* (2003) for details]

$$\breve{\mathbf{a}} = \left(\frac{\mathbf{R}^{-1}}{\lambda}+\mathbf{I}\right)^{-1}\overline{\mathbf{a}} \tag{20}$$

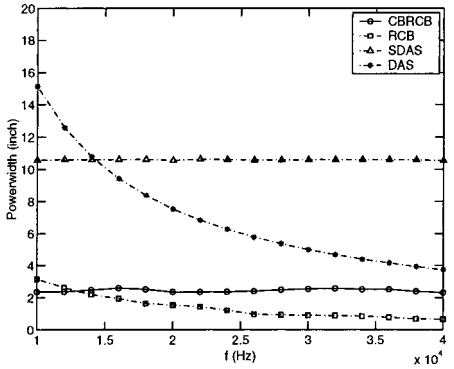
$$= \overline{\mathbf{a}} - (\mathbf{I}+\lambda\mathbf{R})^{-1}\overline{\mathbf{a}}, \tag{21}$$

where we have used the matrix inversion lemma (Stoica and Moses, 1997) to obtain the second equality. Using (21) in the equality constraint of (19), the Lagrange multiplier $\lambda \geqslant 0$ is then obtained as the solution to the constraint equation:

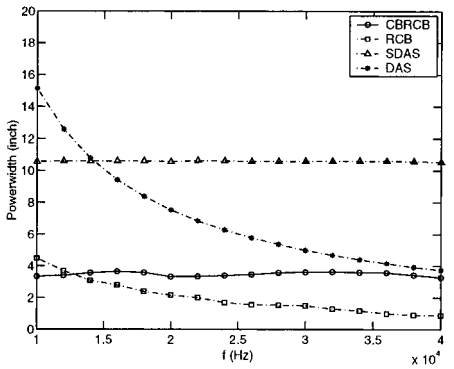
$$\|(\mathbf{I}+\lambda\mathbf{R})^{-1}\overline{\mathbf{a}}\|^2 = \varepsilon, \tag{22}$$

whose left-hand side is a monotonically decreasing function of λ for $\lambda \geqslant 0$.

Next, $\hat{\sigma}^2$ is determined by using Eq. (14) with $\overline{\mathbf{a}}$ replaced by $\breve{\mathbf{a}}$ in Eq. (20), after the norm constraint $\|\breve{\mathbf{a}}\|^2 = M$ is imposed to eliminate a scaling ambiguity problem (see Li *et al.*, 2003; Stoica *et al.*, 2003):



(a) $\varepsilon = 1.0$ for RCB and $\varepsilon = 0.5$ for CBRCB



(b) $\varepsilon = 2.0$ for RCB and $\varepsilon = 1.0$ for CBRCB

FIG. 4. Comparison of the powerwidths for the CBRCB, RCB, SDAS, and DAS methods with $N=64$ when (a) $\varepsilon=1.0$ for RCB and $\varepsilon=0.5$ for CBRCB and (b) $\varepsilon=2.0$ for RCB and $\varepsilon=1.0$ for CBRCB.

$$\hat{\sigma}^2 = \frac{\|\breve{\mathbf{a}}\|^2}{M\breve{\mathbf{a}}^*\mathbf{R}^{-1}\breve{\mathbf{a}}}. \tag{23}$$

The beamwidth of RCB decreases with the frequency but generally it does not depend on the choice of ε if the SOI is sufficiently separated from the interferences. On the other hand, the powerwidth of RCB depends on the signal-to-noise-ratio (SNR), the frequency, and ε. Since the steering vector and hence its uncertainty set are both functions of the frequency [see Eqs. (3) and (15)], it is natural to consider altering the uncertainty parameter ε with the frequency. Intuitively, a larger ε will yield a larger powerwidth. By choosing a frequency-dependent parameter ε for RCB, we obtain the constant-powerwidth robust Capon beamformer (CPRCB), which is able to provide consistent SPL estimates across the frequency bins for the source of interest. However, the beamwidth of CPRCB changes with the frequency in the same way as that of RCB.

Although it is difficult to yield an analytical formula for choosing ε as a function of frequency to guarantee a nearly constant powerwidth, such a choice can be readily made numerically via a contour plot of the powerwidths of RCB with respect to the frequency and ε. Given a desired powerwidth, we can determine ε as a function of frequency from the contour plot. (See Sec. V for details.)

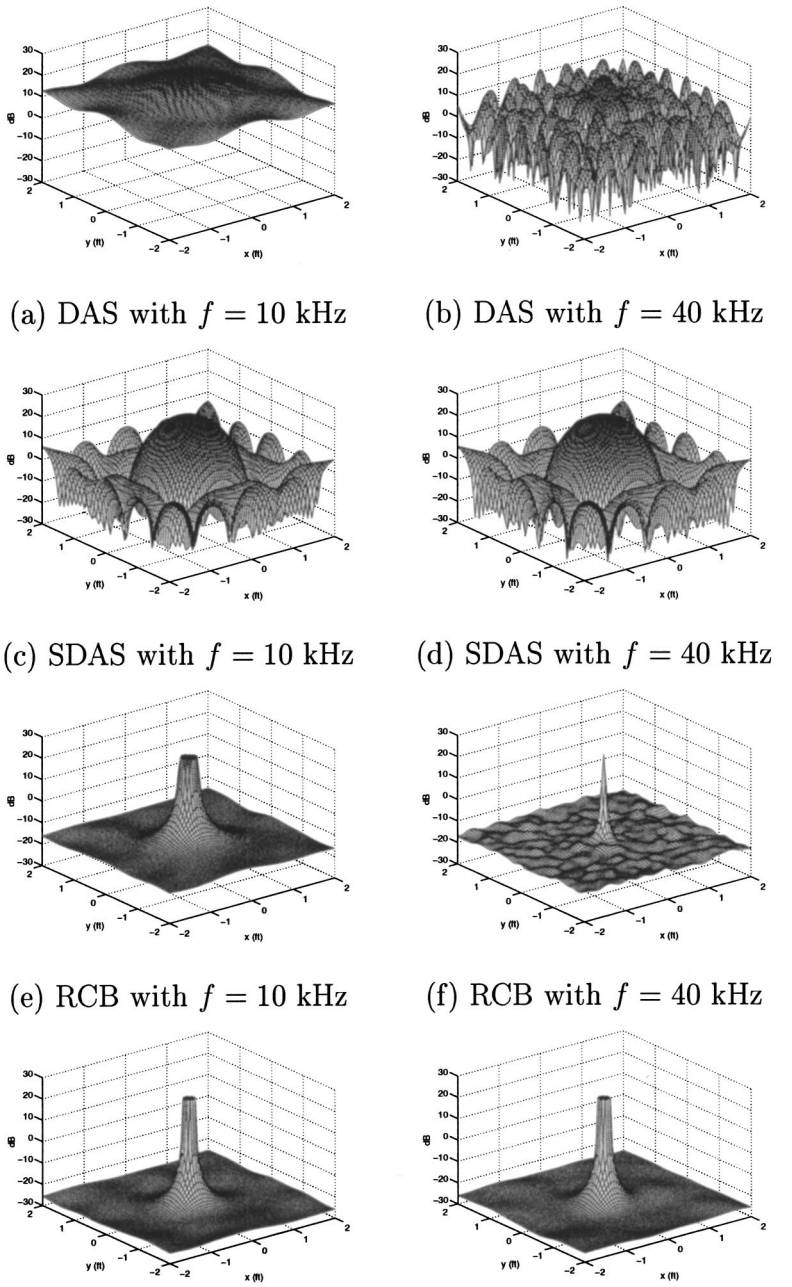


FIG. 5. Comparison of the acoustic imaging results obtained via the DAS, SDAS, RCB, and CBRCB methods with $N=64$, for the narrow-band frequency bins at $f=10$ kHz for (a), (c), (e), and (g), and $f=40$ kHz for (b), (d), (f) and (h), respectively. For RCB, $\varepsilon=2.0$. For CBRCB, $\varepsilon=1.0$. The z axes show the SPL.

IV. CONSTANT-BEAMWIDTH ROBUST CAPON BEAMFORMING

The shading scheme in Humphreys *et al.* (1998) and Brooks and Humphreys (1999) is based on the idea that at each frequency bin, a specific set of array element shading weights should be used which are appropriate for that particular frequency bin. This scheme provides a simple way of calculating the array element shading weights to achieve constant beamwidth for DAS between 10 and 40 kHz. Our RCB can be readily combined with the shading scheme of Humphreys *et al.* (1998) and Brooks and Humphreys (1999) to obtain a constant beamwidth for the desired frequency band, as explained below. We refer to this approach as the constant-beamwidth RCB (CBRCB) algorithm.

Let $\mathbf{v}$ denote the $M\times 1$ vector containing the array ele-

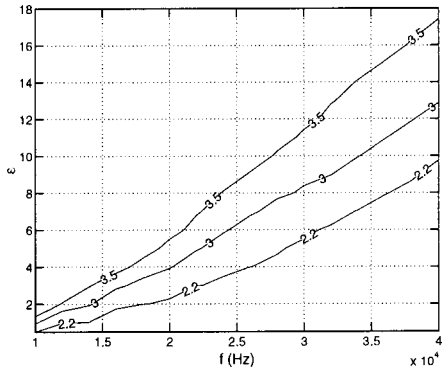
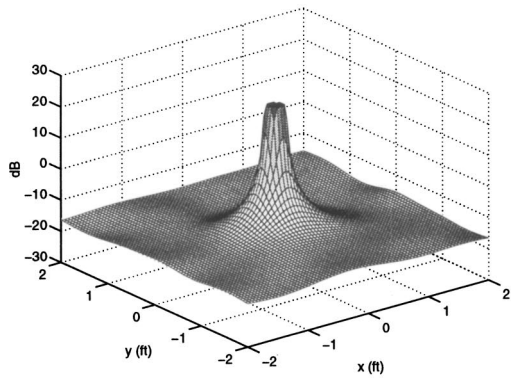
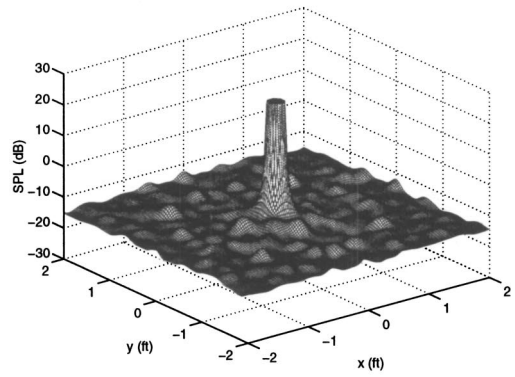


FIG. 6. Contour plots of the powerwidth, versus ε and the frequency, for the RCB method. The numbers on the contours are the 3-dB powerwidths in inch.

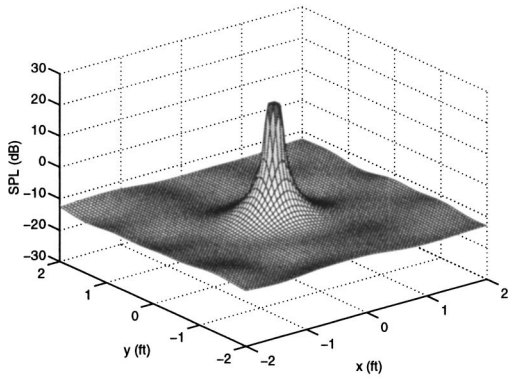


(a) $f = 10$ kHz and $\varepsilon = 1.3$

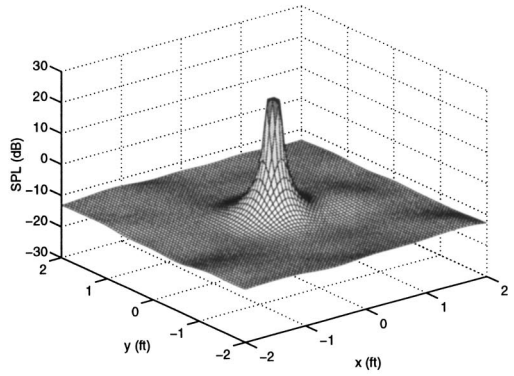


(b) $f = 40$ kHz and $\varepsilon = 13$

FIG. 7. Acoustic imaging results obtained via the CPRCB method with $N = 64$ when (a) $f = 10$ kHz and $\varepsilon = 1.3$ and (b) $f = 40$ kHz and $\varepsilon = 13$.



(a) $f = 10$ kHz



(b) $f = 40$ kHz

FIG. 8. Acoustic imaging results obtained via the CBRCB method with $N = 64$ and $\varepsilon = 0.65$ when (a) $f = 10$ kHz and (b) $f = 40$ kHz.

ment shading weights for the given frequency bin. The assumed array steering vector for CBRCB can now be written as

$$\tilde{\mathbf{a}} = \mathbf{v} \odot \overline{\mathbf{a}}, \tag{24}$$

where $\odot$ denotes elementwise multiplication (Stoica and Moses, 1997). Accordingly, the covariance matrix for the CBRCB is tapered as follows:

$$\tilde{\mathbf{R}} = \mathbf{R} \odot (\mathbf{v}\mathbf{v}^T). \tag{25}$$

Since both $\mathbf{R}$ and $\mathbf{v}\mathbf{v}^T$ are positive semidefinite matrices, $\tilde{\mathbf{R}}$ is also positive semidefinite (see Stoica and Moses, 1997; Guerci, 1999). Note that RCB can be viewed as a special case of CBRCB with all the elements of $\mathbf{v}$ being 1.

Similarly to Eq. (16) in the previous section, the CBRCB has the form

$$\begin{aligned} &\max_{\sigma^2, \mathbf{a}} \sigma^2 \quad \text{subject to} \quad \tilde{\mathbf{R}} - \sigma^2 \mathbf{a}\mathbf{a}^* \geqslant 0, \\ &\|\mathbf{a} - \tilde{\mathbf{a}}\|^2 \leqslant \varepsilon, \end{aligned} \tag{26}$$

which can be solved like the RCB problem in Li *et al.* (2003).

We remark that the ability of CBRCB to retain a constant beamwidth across the frequencies makes it suitable for a large body of applications such as ultrasonics, underwater acoustics, acoustic communications, as well as speech acquisition. Furthermore, CBRCB can also achieve constant powerwidth, which is essential for consistent imaging. However, CBRCB is less powerful and flexible than CPRCB for applications where constant powerwidth is demanded. First, CPRCB has more degrees of freedom for interference suppression than CBRCB since at each frequency the shading scheme involved in the latter tends to deactivate some elements of the full array. In addition, CPRCB can be used with arbitrary arrays, while a special underlying array structure is required for CBRCB due to the particular shading scheme employed. As mentioned earlier, SDAS can also be utilized to yield constant beamwidth. Nevertheless, SDAS is data independent and hence has poorer resolution and worse interference suppression capability than CBRCB and CPRCB.

V. SIMULATED AND EXPERIMENTAL EXAMPLES

We provide some simulation-based and experimental examples to compare the performances of the DAS, SDAS, SCB, RCB, CPRCB, and CBRCB approaches. In all of the examples, we use the small aperture directional array (SADA) (Humphreys *et al.*, 1998; Brooks and Humphreys, 1999) for a scenario with one wideband point calibration source, which is 5 ft above the array at the broadside. SADA consists of 33 microphones arranged in four circles of eight microphones each and one microphone at the array center. The diameter of each circle is twice that of the closest circle it encloses. The maximum radius of the array is 3.89 in. Figure 1 shows the microphone layout of the SADA and its three subarrays used in the shading scheme [referred to as

clusters in Humphreys *et al.* (1998) and Brooks and Humphreys (1999)]. Note that some microphones are shared by different clusters. Each cluster of SADA has the same directional characteristics for a given wavenumber-length product kD_n, where k is the wavenumber and D_n is the diagonal distance between the elements of the nth cluster. The wavenumber-length products at 10 kHz for cluster 3, at 20 kHz for cluster 2, and at 40 kHz for cluster 1 are the same. According to the array coordinate frame, the array is located in the x-y plane, with center location at (0,0,0). Note that inch is used as the unit for the 3-D coordinates.

In what follows we assume that the distance between the array and the source is known and plot the 2-D images by scanning the locations on a plane parallel to the array and situated 5 ft above. In fact, even if we only had an approximate knowledge of this distance, the imaging results would still be similar. A simple explanation is as follows. Assume that the array has an aperture of A and is located at a distance of L from a narrow-band point source with a wavelength of λ. According to Born and Wolf (1970) and Borcea *et al.* (2002), if $L \gg A$, the array range resolution is $\lambda(L/A)^2$, which is much larger than $\lambda(L/A)$, the cross range resolution. Hence the distance is not a key issue here. We assume that $\mathbf{a}$ belongs to the uncertainty set

$$\|\mathbf{a}-\overline{\mathbf{a}}\|^2 \leqslant \varepsilon, \tag{27}$$

where ε is a user parameter chosen to account for the steering vector uncertainty. Note that this form of uncertainty set used in the CPRCB can cover all kinds of array errors, including calibration errors, look direction errors, or array covariance estimation errors due to a small snapshot number (sample size). The uncertainty set for the CBRCB is the same as above except that $\tilde{\mathbf{a}}$ is used instead of $\overline{\mathbf{a}}$. Figure 2 shows the SADA cluster shading weights as a function of the frequency bins, with w_1, w_2, and w_3 corresponding to clusters 1, 2, and 3, respectively. Since some array elements are shared by different clusters, the shading weights of those elements are the sum of the corresponding cluster shading weights.

In the simulated examples below, we consider an array that is identical to SADA, with a wideband monopole source (flat spectrum from 0 Hz to 70 kHz) located at (0,0,60) (except for Fig. 9) in the array coordinate frame and a spatially white Gaussian noise with SNR equal to 20 dB and the SPL equal to 20 dB for each frequency. We use an 8192-point FFT on the non-overlapping blocks (each containing 8192 samples) of simulated data to convert the wideband signal into 8192 narrow-band frequency bins.

Figure 3 compares the 3-dB beamwidths as functions of the frequency, corresponding to the CBRCB, RCB, SDAS, and DAS methods when $N=64$. Note that herein the beamwidth of RCB coincides with that of DAS and the beamwidth of CBRCB coincides with that of SDAS. CBRCB and SDAS achieve constant beamwidth over the frequency band from 10 to 40 kHz, whereas the beamwidths of RCB and DAS are frequency dependent and decrease appreciably with the frequency. We used $\varepsilon=2.0$ for RCB and CBRCB. Other choices of ε for RCB and CBRCB yield the same results and hence they are not shown here. We remark that one can not achieve constant beamwidth for RCB by varying ε. For example, it can be shown that the beampattern of RCB is independent of ε if $\mathbf{Q}$ in Eq. (11) is white Gaussian noise, i.e., $\mathbf{Q}=\sigma_n^2\mathbf{I}$, where σ_n^2 denotes the noise power and $\mathbf{I}$ is an identity matrix.

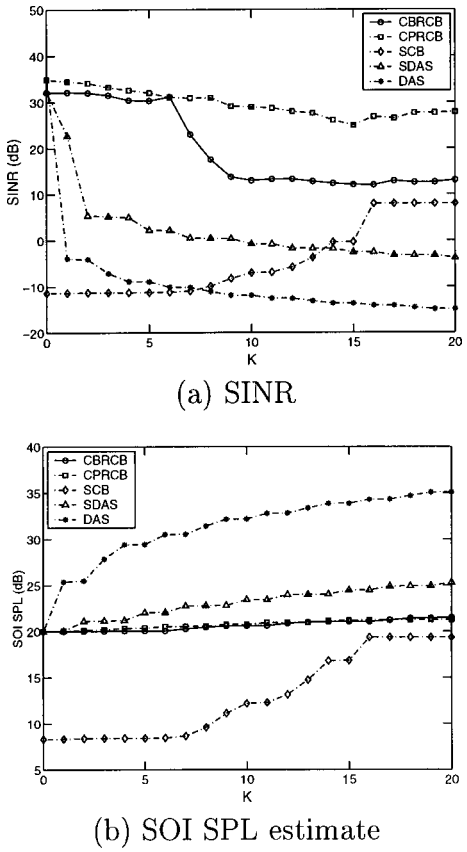


FIG. 9. Comparison of the SINR and SOI SPL estimates obtained via the CBRCB, CPRCB, SCB, SDAS, and DAS methods, versus the number of interferences K, for the narrow-band frequency bin at $f=20$ kHz. For CPRCB, $\varepsilon=2.0$. For CBRCB, $\varepsilon=1.0$. We consider a look direction error case where the assumed source location is (0,0,60) but the actual point source is located at (0.2,0.2,60) with SNR equal to 20 dB. The INRs are equal to 40 dB.

Figure 4 compares the 3-dB powerwidths of the CBRCB, RCB, SDAS, and DAS methods, as functions of the frequency, when $N=64$. Note that the powerwidths of the DAS and RCB methods decrease drastically as the frequency increases, while SDAS and CBRCB can both achieve approximately constant powerwidth. In addition, CBRCB has much smaller powerwidth than SDAS. As can be seen from the figure, we can also adjust the powerwidth for CBRCB and RCB by choosing different values of ε.

Figure 5 compares the acoustic imaging results or sound pressure level (SPL) estimates obtained via the DAS, SDAS, RCB, and CBRCB methods for the narrow-band frequency bins at 10 and 40 kHz, with $N=64$. The z axes show the

TABLE I. Comparison of the CBRCB and CPRCB approaches.

Approach	Constant beamwidth	Constant powerwidth	Loss of DOF
CBRCB	Yes	Yes	Yes
CPRCB	No	Yes	No

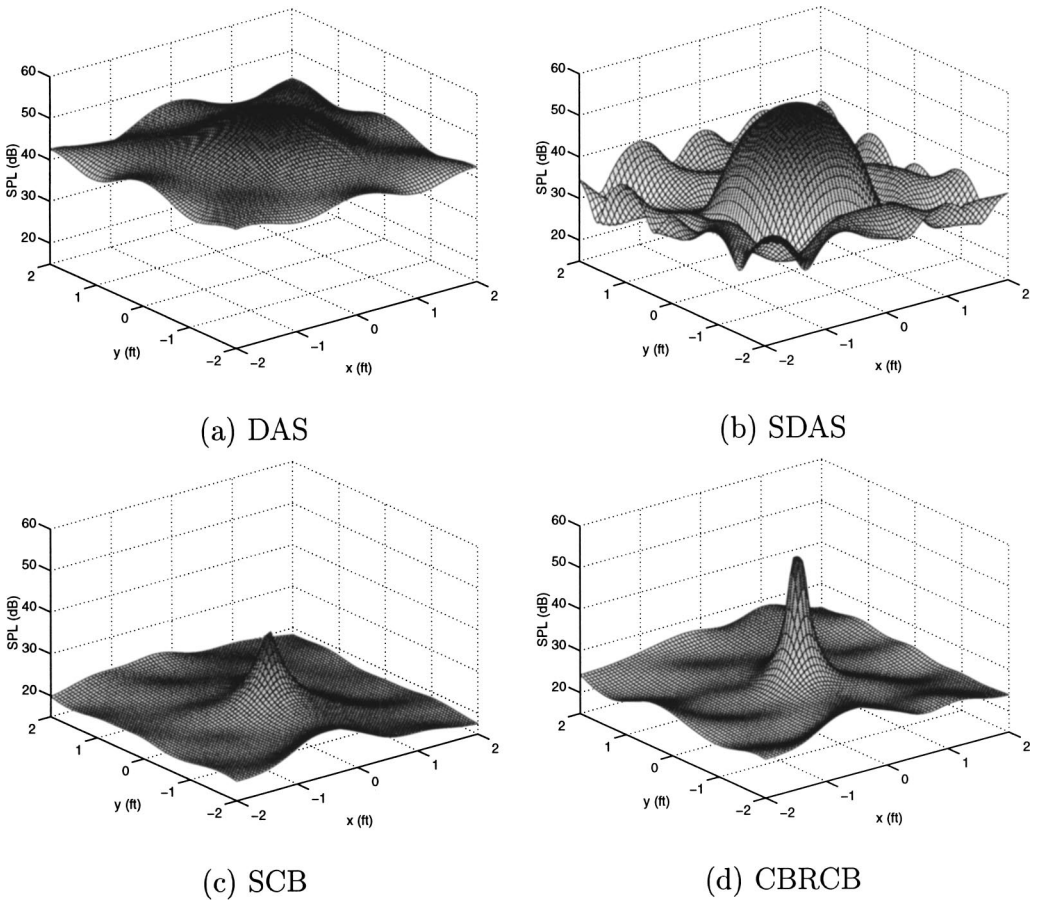
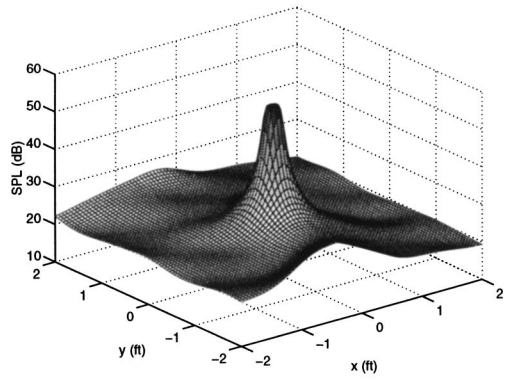


FIG. 10. Comparison of the acoustic imaging results obtained via the DAS, SDAS, SCB, and CBRCB methods for the narrow-band frequency bin at 10 kHz with $N=64$. For CBRCB, $\varepsilon=0.65$.

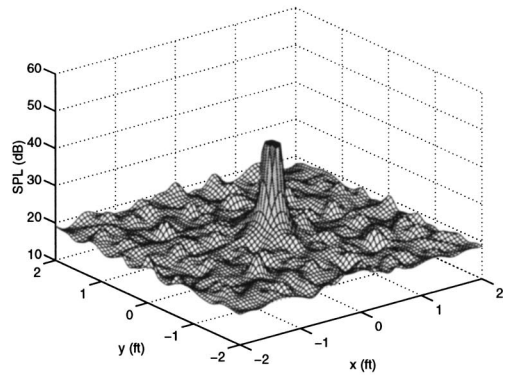
SPL. We used $\varepsilon=2.0$ for RCB and $\varepsilon=1.0$ for CBRCB. Note that we choose ε for CBRCB to be one-half of that for RCB due to the fact that the squared norm of the steering vector for CBRCB is about one-half of that of RCB. As can be seen, the DAS method has poor resolution and high sidelobes and its images vary considerably with the frequency. RCB cannot be used to obtain consistent imaging results over different frequency bins, either, though it has much better resolution than DAS. It is worth noting that both SDAS and CBRCB maintain approximately the same SPL estimates across the frequency bins, but the latter has much better resolution and lower sidelobes and hence better interference rejection capability than the former. It is obvious that CBRCB significantly outperforms the other methods. According to the previous discussions and the results shown in Figs. 3 and 4, it is the constant powerwidth rather than the constant beamwidth that contributes to the better performance of CBRCB as compared to SDAS.

Figure 6 shows the contours of the 3-dB powerwidth of RCB as ε and the frequency vary, when $N=64$. As can be seen, the contours are almost linear with respect to the frequency and ε. Therefore, given a desired powerwidth, we can readily determine ε as a function of the frequency from the corresponding contour plot. Then CPRCB will have a constant powerwidth across the frequency bins. In Fig. 7, we show the imaging results obtained via the CPRCB approach by choosing $\varepsilon=1.3$ when $f=10\,\mathrm{kHz}$ and $\varepsilon=13$ when $f=40\,\mathrm{kHz}$ from the 3-in. powerwidth contour in Fig. 6. The similarity of the SOI SPL estimates obtained with CPRCB at these two frequencies, especially near the powerwidth area, verifies the consistency of CPRCB in powerpattern across the frequencies. Figure 8 shows the imaging results obtained via the CBRCB approach with $\varepsilon=0.65$, for $f=10\,\mathrm{kHz}$ and $f=40\,\mathrm{kHz}$. Again we note the consistency in the imaging results. Therefore, both CPRCB and CBRCB are suitable for applications where consistent SPL estimates are desirable. However, the sidelobes in Fig. 7(b) are higher and rougher than those in Fig. 8(b). Despite this fact, CPRCB does not perform worse than CBRCB (see the next example).

We consider a look direction error case where the assumed source location is (0,0,60) but the actual source is located at (0.2,0.2,60) with SNR equal to 20 dB. Also we consider a varying number of interferences from $K=0$ to $K=20$, which are situated on a circle with a radius of 20 in. and have an INR equal to 40 dB. The circle is on a plane parallel to the array and situated 60 in. above. We assume that the theoretical covariance matrix $\mathbf{R}$ is known here. Figure 9 compares the SINR and SOI SPL estimates obtained via the CBRCB, CPRCB, SCB, SDAS, and DAS methods versus the number of interferences K, for the narrowband frequency bin at $f=20\,\mathrm{kHz}$. For CPRCB, $\varepsilon=2.0$. For CBRCB, $\varepsilon=1.0$. Note that SCB is very sensitive to the steering vector mismatch and suffers from severe performance degradation in SINR and SOI SPL estimates. Although DAS and SDAS are robust against array errors, they have poor capacity for interference suppression. Consequently, their

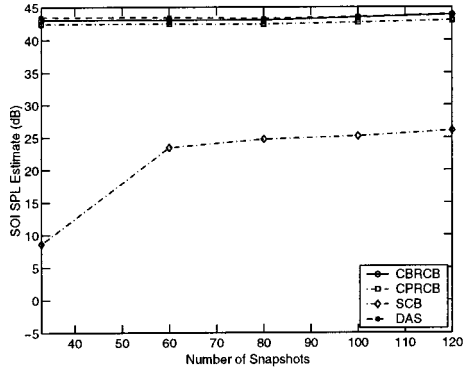


(a) $f = 10$ kHz and $\varepsilon = 1.3$

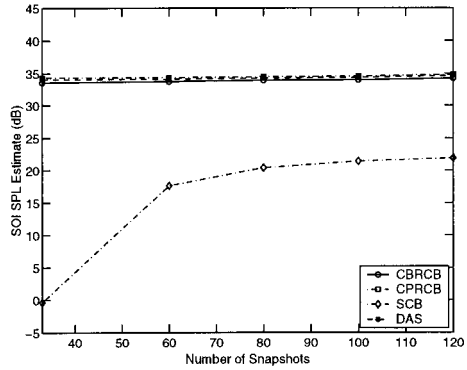


(b) $f = 40$ kHz and $\varepsilon = 13$

FIG. 11. Acoustic imaging results obtained via the CPRCB method with $N=64$ when (a) $f=10$ kHz and $\varepsilon=1.3$ and (b) $f=40$ kHz and $\varepsilon=13$.



(a) $f = 10$ kHz



(b) $f = 40$ kHz

FIG. 12. Comparison of the SOI SPL estimates, versus N, obtained via the CBRCB, CPRCB, SCB, and DAS methods for the narrow-band frequency bins at (a) $f=10$ kHz and (b) $f=40$ kHz, respectively. For CPRCB, $\varepsilon=1.3$ when $f=10$ kHz and $\varepsilon=13$ when $f=40$ kHz. For CBRCB, $\varepsilon=0.65$.

SINRs and SOI SPL estimates are unsatisfactory. CBRCB and CPRCB outperform the other approaches due to their robustness to steering vector errors, better resolution, and much better interference rejection capability than the data-independent beamformers. As can be seen, CPRCB has higher SINR than CBRCB. This is due to the fact that the former has more degrees of freedom (DOFs) for interference suppression than the latter. It might seem surprising that the performance of SCB improves as the number of interferences K increases. There is a simple explanation for this. When K is small, SCB has a sufficient number of DOFs and the SOI is suppressed as interference. As K increases, SCB focuses more on suppressing the interferences than the SOI since the INR is much higher than the SNR. Table I compares the CBRCB and CPRCB approaches in terms of constant beamwidth, constant powerwidth, and loss of DOF.

Finally, we apply several methods to the experimental data with test 510 and run 3 collected at the Quiet Flow Facility (QFF), NASA Langley Research Center, using the SADA array. The wideband signal is located near (0,0,60) in the array coordinate frame. We use a 1024-point FFT on non-overlapping blocks (each containing 1024 samples) to convert the wideband signal into 1024 narrow-band frequency bins. We estimate the sound pressure level (SPL), which is defined in Eq. (8).

Figure 10 compares the acoustic imaging results obtained via the DAS, SDAS, SCB, and CBRCB methods, with $N=64$, for the narrow-band frequency bin at 10 kHz. For the case of a single dominant source, DAS acts as a matched filter and gives good power estimate at the true source location. Note that both DAS and SDAS have poor resolution and high sidelobes but they are robust against array errors. The SCB approach, on the other hand, is very sensitive to steering vector mismatch. We see that it suffers from severe performance degradation and substantially underestimates the signal SPL. Empirically, we found that even if very fine imaging grids (e.g., with a grid step of 0.01 in.) are used, the performance of SCB still does not improve. Thus the mismatch is not due to pointing errors. We have also tried using large numbers of snapshots (e.g., $N=200$) for SCB but obtained similar inferior imaging results. The discussions above indicate that the performance degradation of SCB is most likely due to array calibration errors or the presence of weak highly correlated multipath signals (Wang *et al.*, 2004). Figure 10(d) shows the imaging results obtained via CBRCB. Note that CBRCB is not only robust to array errors but also has a good resolution, outperforming DAS, SDAS, and SCB. Also we found that the source coordinate estimates are (0.37, −0.44,60), which is the peak location of the CBRCB image.

Figure 11 is similar to Fig. 7 except that the experimental data are used here. We also choose $\varepsilon=1.3$ when $f=10$ kHz and $\varepsilon=13$ when $f=40$ kHz, as in Fig. 7. As can be seen, the SPL at 10 kHz is about 10 dB higher than that at 40 kHz. Hence the calibration source presumably does not have a flat spectrum between 10 and 40 kHz. The experimental

imaging results of CBRCB are similar and hence not shown.

Figure 12 compares the SPL estimates at the aforementioned estimated source location, as a function of the number of snapshots, obtained via the CBRCB, CPRCB, SCB, and DAS methods. Note that the CBRCB, CPRCB, and DAS methods are not sensitive to the small snapshot number effect, and that their SPL estimates vary very little with N. SDAS performs similarly to DAS, CBRCB, and CPRCB and hence it is not shown. In contrast, although the performance of SCB improves somewhat as the number of snapshots increases, its convergence speed becomes very slow after $N=60$ and it still considerably underestimates the SOI SPL even when $N=120$.

VI. CONCLUSIONS

We have presented two methods, CPRCB and CBRCB, as extensions of the RCB for consistent wideband acoustic imaging. We have shown that these two approaches retain the appealing properties of RCB, including much better robustness to steering vector mismatches than SCB and better resolution and much better interference rejection capability than the data-independent beamformers. CBRCB combines RCB with a shading scheme proposed for DAS and yields constant beamwidth over the frequency band used to design the shading scheme, whereas CPRCB extends RCB by choosing a frequency-dependent array steering vector uncertainty parameter to attain constant powerwidth and hence consistent SPL estimates across the frequency band of interest. CBRCB requires a special underlying array structure while CPRCB does not. In addition, the former has fewer degrees of freedom for interference suppression than the latter. Our simulated and experimental examples have demonstrated that CBRCB is the suitable choice for applications requiring constant beamwidth such as speech acquisition, while CPRCB is the preferred choice for applications demanding constant powerwidth for consistent intensity (e.g., SPL or power) estimation such as acoustic imaging.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation Grant Nos. CCR-0104887 and ECS-0097636, the Swedish Science Council (VR), and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT). The Small Aperture Directional Array data sets used in the generation of results shown in this work were provided under the NASA Langley Research Center Grant No. NAG-1-2133.

Abhayapala, T., Kennedy, R., and Williamson, R. (**2000**). ''Nearfield broadband array design using a radially modal expansion,'' J. Acoust. Soc. Am. **107**, 392–403.

Blackstock, D. T. (**2000**). *Fundamentals of Physical Acoustics* (Wiley, New York).

Borcea, L., Papanicolaou, G., Tsogka, C., and Berryman, J. G. (**2002**). ''Imaging and time reversal in random media,'' Inverse Probl. **18**, 1247–1279.

Born, M., and Wolf, E. (**1970**). *Principles of Optics* (Academic, New York).

Brandstein, M., and Ward, D. (eds.) (**2001**). *Microphone Arrays: Signal Processing Techniques and Applications* (Springer Verlag, New York).

Brooks, T., Marcolini, M., and Pope, D. (**1987**). ''A directional array approach for the measurement of rotor source distributions with controlled spatial resolution,'' J. Sound Vib. **112**, 192–197.

Brooks, T. F., and Humphreys, W. M. (**1999**). ''Effect of directional array size on the measurement of airframe noise components,'' AIAA Paper 99-1958, 5th AIAA Aeroacoustics Conference, Bellevue, WA.

Capon, J. (**1969**). ''High resolution frequency-wavenumber spectrum analysis,'' Proc. IEEE **57**, 1408–1418.

Chou, T. (**1995**). ''Frequency-independent beamformer with low response error,'' ICASSP, Detroit, MI, Vol. 5, pp. 2995–2998.

Doles, III, J., and Benedict, F. (**1988**). ''Broad-band array design using the asymptotic theory of unequally spaced arrays,'' IEEE Trans. Antennas Propag. **36**, 27–33.

Goodwin, M. M., and Elko, G. W. (**1993**). ''Constant beamwidth beamforming,'' ICASSP, Minneapolis, MN, Vol. 1, pp. 169–172.

Guerci, J. R. (**1999**). ''Theory and application of covariance matrix tapers for robust adaptive beamforming,'' IEEE Trans. Signal Process. **47**, 977–985.

Humphreys, W. M., Brooks, T. F., Hunter, W. W., and Meadows, K. R. (**1998**). ''Design and use of microphone directional arrays for aeroacoustic measurement,'' AIAA Paper 98-0471, 36st Aerospace Sciences Meeting and Exhibit, Reno, NV.

Lardies, J. (**1989**). ''Acoustic ring array with constant beamwidth over a very wide frequency range,'' Acoust. Lett. **13**, 77–81.

Li, J., Stoica, P., and Wang, Z. (**2003**). ''On robust Capon beamforming and diagonal loading,'' IEEE Trans. Signal Process. **51**, 1702–1715.

Lorenz, R., and Boyd, S. (**2001**). ''Robust minimum variance beamforming,'' IEEE Trans. Signal Process. available at http://www.stanford.edu/~boyd/rmvb.html.

Marzetta, T. L. (**1983**). ''A new interpretation for Capon's maximum likelihood method of frequency-wavenumber spectrum estimation,'' IEEE Trans. Acoust., Speech, Signal Process. **31**, 445–449.

Smith, R. (**1970**). ''Constant beamwidth receiving arrays for broad band sonar systems,'' Acustica **23**, 21–26.

Stoica, P., and Moses, R. L. (**1997**). *Introduction to Spectral Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Stoica, P., Wang, Z., and Li, J. (**2003**). ''Robust Capon beamforming,'' IEEE Signal Process. Lett. **10**, 172–175.

Tucker, D. (**1957**). ''Arrays with constant beam-width over a wide frequency range,'' Nature (London) **180**, 496–497.

Van Trees, H. L. (**2002**). *Detection, Estimation, and Modulation Theory, Part IV, Optimum Array Processing* (Wiley, New York).

Vorobyov, S. A., Gershman, A. B., and Luo, Z.-Q. (**2003**). ''Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem,'' IEEE Trans. Signal Process. **51**, 313–324.

Wang, Y., Li, J., Stoica, P., Sheplak, M., and Nishida, T. (**2004**). ''Wideband RELAX and wideband CLEAN for aeroacoustic imaging,'' J. Acoust. Soc. Am. available at ftp://www.dsp.ufl.edu/ywwang/jasa.pdf.

Ward, D., Kennedy, R., and Williamson, R. (**1995**). ''Theory and design of broad-band sensor arrays with frequency invariant far-field beam patterns,'' J. Acoust. Soc. Am. **97**, 1023–1034.

Wu, S. Q., and Zhang, J. Y. (**1999**). ''A new robust beamforming method with antennae calibration errors,'' IEEE Wireless Communications and Networking Conference, New Orleans, LA, Vol. 2, pp. 869–872.

Modeling the growth rate of distortion product otoacoustic emissions by active nonlinear oscillators

Renata Sisto[a)]
Dipartimento Igiene del Lavoro, ISPESL, Via Fontana Candida, 1, 00040 Monte Porzio Catone (Roma), Italy

Arturo Moleti[b)]
Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica, 1, 00133 Roma, Italy

(Received 12 March 2004; revised 26 May 2004; accepted 3 June 2004)

In this work, growth-rate curves of the $2f_1$-f_2 distortion product otoacoustic emission (DPOAE) are analyzed in a population of 30 noise exposed subjects, including both normal-hearing and hearing impaired subjects. A particular embedded limit-cycle oscillator equation is used to model the cochlear resonant response at the cochlear places of the primary and secondary tone frequencies (f_2 and $2f_1$-f_2). The parameters of the oscillator equation can be directly interpreted in terms of effectiveness of the cochlear feedback mechanisms associated with the active filter amplification. A two-sources paradigm is included in the model, in agreement with experimental evidence and with the assumptions of more detailed full cochlear models based on the transmission line formalism. According to this paradigm, DPOAEs are nonlinearly generated at the cochlear place that is resonant at frequency f_2, and coherently reflected at the $2f_1$-f_2 place. The analysis shows that the model, which had been previously used to describe the relaxation dynamics of transient evoked otoacoustic emissions (TEOAEs), also correctly predicts the observed growth rate of the DPOAE response as a function of the primary tones amplitude. A significant difference is observed between normal and impaired ears. The comparison between the growth rate curves at different frequencies provides information about the dependence of cochlear tuning on frequency. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1775278]

PACS numbers: 43.64.Jb [BLM] Pages: 1632–1638

I. INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals measured in the outer ear either with (evoked OAEs, EOAEs) or without (spontaneous OAEs, SOAEs) acoustic stimulation. A variety of OAE techniques are now available for investigating the properties of the auditory system (Probst *et al.*, 1991; Shaffer *et al.*, 2003). Experimental evidence provided by OAE observations shows that the transiently evoked cochlear response is essentially that of an active nonlinear amplifier characterized by resonances at characteristic frequencies, sometimes with measurable limit-cycle spontaneous oscillations. Long lasting emissions, which dominate the transient evoked OAE (TEOAE) signals, may effectively be modeled by a particular embedded limit-cycle oscillator equation (Sisto and Moleti, 1999), which is different from the classical van der Pol equation, and correctly predicts both the saturation of the stimulus–response curve and the exponential decay of the spectral line amplitude, as well as the existence, synchronization and suppression of SOAEs. In this work, the model will be applied to the analysis of $2f_1$-f_2 DPOAEs, as regards the growth rate of the response as a function of the primary-tone amplitude.

The ''two-mechanisms model'' for DPOAE generation seems to be well established (Talmadge *et al.*, 1998). This model predicts that $2f_1$-f_2 DPOAEs are the superposition of two separate emissions: nonlinear distortion near the f_2 place, and linear coherent reflection near the $2f_1$-f_2 place. Experimental confirmation of this model has come from the analysis of the DPOAE fine structure in humans and animals, which is due to the interference between the two emissions (Talmadge *et al.*, 1999; Mauermann *et al.*, 1999a; Kalluri and Shera, 2001, Withnell *et al.*, 2003). The effect of multiple internal reflections on the DPOAE fine structure has also been considered (Dhar *et al.*, 2002). The DPOAE growth-rate has been studied by many authors both in humans (Popelka *et al.*, 1993; Abdala, 2000) and in small mammals (Whitehead *et al.*, 1995a, b; Withnell and Yates, 1998). Some authors investigated the relationship between DPOAE growth-rate and hearing loss (e.g., Dorn *et al.*, 2001), finding that the optimal difference between the level of the two stimuli, which is capable of eliciting maximum amplitude DPOAEs (Gaskill and Brown, 1990; Whitehead *et al.*, 1995a, b; Kummer *et al.*, 2000) also provides the best correlation with cochlear hearing loss. It has been noted that, in some cases, the disappearance of a DPOAE fine structure may be a sensitive indicator of hearing loss (Mauermann *et al.*, 1999b). Methods for the estimation of the hearing threshold from DPOAE I/O functions have been proposed (Boege and Janssen, 2002; Oswald and Janssen, 2003). The relation with loudness perception (Neely *et al.*, 2003) has also been recently discussed.

In this work we will focus on the interpretation of the

[a)]Electronic mail: r.sisto@dil.ispesl.it
[b)]Electronic mail: arturo.moleti@roma2.infn.it

 © 2004 Acoustical Society of America

DPOAE growth curves in terms of the response at frequency $f_{DP}=2f_1-f_2$ of two effective oscillators: the oscillator of frequency f_2, nonlinearly driven by the primary tones f_1 and f_2, and the oscillator of frequency f_{DP}, driven by the tone f_{DP} that has been previously nonlinearly generated by the oscillator of frequency f_2. In doing so, we will use linear units for the stimulus and response amplitudes, and a constant difference between the levels of the two stimuli, to simplify the comparison of the experimental results with the predictions of the model. This comparison may provide, in principle, indirect information on the tuning properties of the cochlea. The analysis of the DPOAE suppression tuning curves (STCs) (Kummer *et al.*, 1995, Abdala, 2001, Gorga *et al.*, 2003), and that of the TEOAE latency-frequency relation (Moleti and Sisto, 2003), have already been proposed to get OAE-based estimates of the cochlear quality factor Q as a function of frequency, comparing normal-hearing and hearing impaired subjects, and adults and neonates. The comparison between different estimates of Q should include also psychoacoustical estimates, but this is beyond the scope of the present work. Indeed, the interpretation of such comparisons would need assuming a comprehensive model of the auditory function, because different techniques could be sensitive to different aspects of the mechanisms that are responsible for the remarkable frequency resolution capability of human hearing.

II. METHOD

Audiometric recordings and DPOAE growth rate curves of 60 ears from 30 adult male subjects (age=21–41 yr, average=30 yr) exposed to industrial noise have been analyzed. Pure tone audiograms were recorded, for all subjects, in an acoustically shielded room. The audiometric threshold was measured with an accuracy of 5 dB. Sound conductive hearing loss was preliminarily excluded by otoscopy and tympanometry. The audiometric test frequencies were 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz. All exposed subjects had been exposed to the same level of noise (A-weighted personal noise exposure level normalized to a nominal 8 h working day, $L_{EX,8H}=87.5\pm0.5$ dB A). They had worked for different periods of time (1 to 14 yr, average=6 yr), using hearing protectors, performing the same tasks in the same industrial environment, so their personal daily noise exposure levels were almost exactly the same, but the exposure duration was variable among them.

The population was divided in two audiometric classes, according to the measured hearing threshold level in a frequency range including that involved in the generation of the DPOAE elicited by $f_2=4$ kHz. Ears with HL$\geqslant$20 dB at any frequency between 2 kHz and 6 kHz were defined as ''high-frequency impaired'' (29 ears). Ears with HL$\leqslant$10 dB at all frequencies between 2 kHz and 6 kHz were defined as ''normal-hearing'' (21 ears). A subset of 10 ears with intermediate hearing threshold levels was not included in the analysis, to get a more significant difference between the audiometric characteristics of the two classes. However, the average hearing threshold difference between the two classes varied between 5 dB (at 2 kHz) and 15 dB (at 6 kHz), corresponding to mild hearing loss. The average hearing threshold pattern of the two classes is shown in Fig. 1. The average exposure time was not significantly different between the two classes.

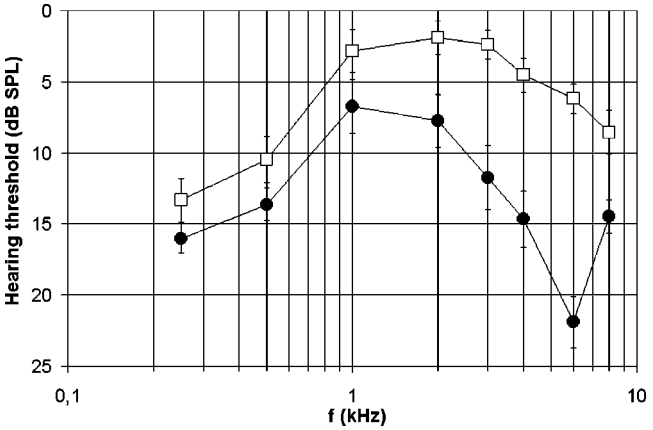


FIG. 1. Average hearing threshold level in the two classes of normal-hearing (21 ears) and hf-impaired ears (29 ears).

DPOAE recordings were recorded for all subjects in an acoustically shielded room using the ILO-96 V.5 system (Otodynamics, Ltd.), with a fixed frequency ratio $f_2/f_1=1.22$, and a fixed amplitude ratio A_1-$A_2=5$ dB. The growth rate was measured in the range $A_2=(40-75\ \text{dB})$, for $f_2=4$ kHz.

In a different population of 9 nonexposed young male subjects the DPOAE growth rate was also measured for $f_2=2$ kHz. According to the same criterion, extended to 1 kHz to include all the generation places of both DPOAEs, 8 normal-hearing ears were selected.

III. MODEL

We use a simple embedded oscillator model, in which the oscillator is directly interfaced to the outer ear. The model is based on the properties of the nonlinear oscillator equation:

$$\ddot{\xi}+\left(a\xi^2+b-\frac{c}{\langle\xi^2\rangle}\right)\dot{\xi}+\omega_0^2\xi=F, \tag{1}$$

where ξ is the oscillator displacement, ω_0 its resonance angular frequency, a, b and c the oscillator constant parameters and F can be any forcing term associated with the stimulus. This oscillator equation was originally proposed (Sisto and Moleti, 1999) because the van der Pol oscillator, which is generally used in embedded cochlear models (Tubis and Talmadge, 1998), cannot explain the exponential decay regime observed in the relaxation dynamics of click evoked OAEs. The difference, with respect to the classical van der Pol oscillator, is in the linear damping and the nonlinear anti-damping terms, which replace the van der Pol linear anti-damping term, while the quadratic nonlinear damping term is the same for both oscillators. The small value of the linear damping coefficient b, as well as the presence of the anti-damping term, is dependent on the properties of the cochlear active filter mechanism mediated by the outer hair cells. The response to a click of this oscillator consists of a short, strongly damped phase, followed by a phase with ex-

ponential decay, and, finally, the system reaches the equilibrium amplitude at the SOAE level (Sisto and Moleti, 1999). In the case of sinusoidal stimulation of angular frequency ω_0, the amplitude of the response of this oscillator is linearly growing with the stimulus amplitude:

$$\xi \approx \frac{F}{\omega_0 b}, \tag{2}$$

as long as

$$\xi \ll \sqrt{\frac{b}{a}}, \tag{3}$$

while for larger values of the displacement ξ, saturation is predicted, with

$$\xi \propto \left(\frac{F}{a\omega_0}\right)^{1/3}. \tag{4}$$

The cubic DPOAE at frequency $f_{DP}=2f_1-f_2$ is generally assumed (e.g., Talmadge *et al.*, 1998) to come from two different sources, nonlinear generation of the combination tone at the f_2 tonotopic place, and subsequent reflection near the f_{DP} place. It is possible to use Eq. (1) to predict the DPOAE response as the sum of two contributions, associated with the response of two oscillators of frequency f_2 and f_{DP}. This is done by assuming that the oscillator of frequency f_{DP} is driven by the sinusoidal stimulus at frequency f_{DP} that is nonlinearly generated at the f_2 place and transmitted to the f_{DP} place. The first contribution is proportional to the product of the amplitude of the excitation at frequency f_2 and the square of the excitation at frequency f_1. The maximum value of this product is found near the f_2 tonotopic place, thus it is reasonable to schematize the primary nonlinear generation of the tone f_{DP} as it were produced by a single oscillator of frequency f_2.

In the case of the $2f_2$-f_1 DPOAE, the external force is proportional to the driver pressure at the two frequencies f_2 and f_1:

$$F=A(p_{\omega_1}+p_{\omega_2}), \tag{5}$$

where A is a conversion factor associated with the conversion of the outer ear pressure stimulus into a differential pressure on the cochlear membrane sides and its transmission to the tonotopic place x_2. It is assumed to be the same for the two nearby frequencies of the DPOAE stimulus.
In the saturated regime, at the cochlear place x_2:

$$a\omega_2\xi_2^3(\omega_2) \cong Ap_{\omega_2}, \tag{6}$$

$$a\omega_1\xi_2^3(\omega_1) \cong Ap_{\omega_1}; \tag{7}$$

thus

$$\xi_2(\omega_2) \cong \left(\frac{Ap_{\omega_2}}{a\omega_2}\right)^{1/3}, \tag{8}$$

$$\xi_2(\omega_1) \cong \left(\frac{Ap_{\omega_1}}{a\omega_1}\right)^{1/3}. \tag{9}$$

The leading order solutions of $\xi_1(\omega_1)$, $\xi_2(\omega_2)$ force the oscillator at x_2 at the frequency ω_{DP} with the term

$$F_2(\omega_{DP})=a\omega_2\xi_2^2(\omega_1)\xi_2^*(\omega_2). \tag{10}$$

At first order, the equation for the response at frequency ω_{DP} of the oscillator at the x_2 place is

$$\ddot{\xi}_2+\left(a\xi_2^2+b-\frac{c}{\langle\xi_2^2\rangle}\right)\dot{\xi}_2+\omega_2^2\xi_2=a\omega_2\xi_2^2(\omega_1)\xi_2^*(\omega_2). \tag{11}$$

Below saturation level,

$$|\xi_2(\omega_{DP})| \approx \frac{|F_2(\omega_{DP})|}{\sqrt{(\omega_2^2-\omega_{DP}^2)^2+\omega_{DP}^2b^2}}, \tag{12}$$

where

$$\omega_{DP}=\omega_2-2\Delta\omega, \quad \Delta\omega=\omega_2-\omega_1. \tag{13}$$

Thus:

$$\begin{aligned}|\xi_2(\omega_{DP})| &\cong \frac{|F_2(\omega_{DP})|}{\sqrt{4^2\Delta\omega^2\omega_{DP}^2\left(1+\frac{\Delta\omega}{\omega_{DP}}\right)^2+\omega_{DP}^2b^2}} \\ &\cong \frac{|F_2(\omega_{DP})|}{\sqrt{4^2\omega_{DP}^2\Delta\omega^2\left[\left(1+\frac{\Delta\omega}{\omega_{DP}}\right)^2+\frac{b^2}{16\Delta\omega^2}\right]}} \\ &\cong \frac{|F_2(\omega_{DP})|}{4\omega_{DP}\Delta\omega},\end{aligned} \tag{14}$$

and, substituting $F_2(\omega_{DP})$, as given by Eq. (10), one gets

$$\begin{aligned}|\xi_2(\omega_{DP})| &= \frac{a\omega_2}{4\omega_{DP}\Delta\omega}\left(\frac{Ap_{\omega_2}}{a\omega_2}\right)^{1/3}\left(\frac{Ap_{\omega_1}}{a\omega_1}\right)^{2/3} \\ &= \frac{A\omega_2}{4\omega_{DP}\Delta\omega}\left(\frac{p_{\omega_2}}{\omega_2}\right)^{1/3}\left(\frac{p_{\omega_1}}{\omega_1}\right)^{2/3}.\end{aligned} \tag{15}$$

For $p_1=Mp_2$,

$$\begin{aligned}|\xi_2(\omega_{DP})| &= \frac{Ap_{\omega_2}\omega_{DP}\left(1+\frac{2\Delta\omega}{\omega_{DP}}\right)}{4\omega_{DP}\Delta\omega}\left(\frac{M}{\omega_1}\right)^{2/3}\left(\frac{1}{\omega_2}\right)^{1/3} \\ &= \frac{Ap_{\omega_2}\left(1+\frac{2\Delta\omega}{\omega_{DP}}\right)^{2/3}}{4\omega_{DP}\Delta\omega}\left(\frac{M}{\left(1+\frac{\Delta\omega}{\omega_{DP}}\right)}\right)^{2/3}.\end{aligned} \tag{16}$$

The excitation at frequency f_{DP} of the oscillator at the place x_2, transmitted to the place x_{DP}, acts as a forcing term F_{DP} for the oscillator at the x_{DP} place. Representing with a constant factor α the attenuation due to cochlear transmission, the forcing term is

$$F_{DP}(\omega_{DP})=\alpha\omega_{DP}^2\xi_2(\omega_{DP}). \tag{17}$$

Far from saturation, the response amplitude at the tonotopic place x_{DP} will be

$$|\xi_{DP}(\omega_{DP})|=\frac{\alpha\omega_{DP}\xi_2(\omega_{DP})}{b}. \tag{18}$$

Substituting Eq. (16),

$$|\xi_{\mathrm{DP}}(\omega_{\mathrm{DP}})| = \frac{\alpha A p_{\omega_2}\left(1+\frac{2\Delta\omega}{\omega_{\mathrm{DP}}}\right)^{2/3}}{4b\Delta\omega}\left(\frac{M}{\left(1+\frac{\Delta\omega}{\omega_{\mathrm{DP}}}\right)}\right)^{2/3}$$

$$= \frac{\alpha A p_{\omega_2}}{b\omega_{\mathrm{DP}}}\left(\frac{6M}{5}\right)^{2/3}, \tag{19}$$

where

$$\Delta\omega \approx 0.2\omega_1 = 0.25\omega_{\mathrm{DP}} \tag{20}$$

has been used.

In embedded oscillator models, there is proportionality between the excitation of the oscillator and the outer ear pressure:

$$\omega^2 \xi = Rp. \tag{21}$$

Thus, the pressure measured as DPOAE of frequency f_{DP} in the ear canal is given by

$$p_{\mathrm{DP}} = \frac{\omega_{\mathrm{DP}}^2}{R}[\xi_2(\omega_{\mathrm{DP}}) + \xi_{\mathrm{DP}}(\omega_{\mathrm{DP}})]. \tag{22}$$

Both terms contribute linearly to the first part of the growth-rate curve, the nonlinear response of the f_2 place, and the response of the tonotopic site f_{DP}:

$$p_{\mathrm{DP}}^{\mathrm{I}} \cong \frac{A p_{\omega_2}}{R}\left(\frac{6M}{5}\right)^{2/3}(1+\alpha Q) = C(1+\alpha Q)p_{\omega_2}, \tag{23}$$

where Q is the quality factor of the resonator of frequency f_{DP}, which, in the linear damping regime, is given by $Q = \omega_{\mathrm{DP}}/b$.

At high levels of stimulation, the response of the tonotopic place at frequency f_{DP} reaches the saturation level too, thus, in this range, the second contribution tends to a constant saturation value, and the first one becomes dominant:

$$p_{\mathrm{DP}}^{\mathrm{II}} \cong \frac{A p_{\omega_2}}{R}\left(\frac{6M}{5}\right)^{2/3} + p_{\mathrm{DPsat}} = Cp_{\omega_2} + p_{\mathrm{DPsat}}. \tag{24}$$

Summarizing, if the primary oscillator at f_2 is in the saturated regime, this contribution to the DPOAE response results to be proportional to the stimulus intensity. The second contribution is proportional to the first one, as long as saturation of the local oscillator at f_{DP} does not occur. Thus it can be predicted that the f_{DP} DPOAE response is the sum of two terms, which are both proportional to the external forcing term in a wide stimulus amplitude interval, the ''saturated-linear range,'' such that the response of the primary oscillators is saturated, while that of the f_{DP} oscillator is not. For very high levels of stimulation, the f_{DP} oscillator enters the ''fully saturated'' regime, and the slope of the growth curve is expected to decrease. In the saturated-linear range, the slope of the growth curve is expected to be

$$\frac{\partial p_{\mathrm{DP}}^{\mathrm{I}}}{\partial p_2} = C(1+\alpha Q), \tag{25}$$

while in the fully saturated range it should decrease to

$$\frac{\partial p_{\mathrm{DP}}^{\mathrm{II}}}{\partial p_2} = C. \tag{26}$$

The value of the coefficients A and R may be estimated only using a full cochlear model; thus it would be a difficult and model-dependent task to give an absolute estimate of Q starting from Eqs. (23) and (24). Nevertheless, by using differential techniques, the comparison with experimental data may provide relative information on the parameter Q, whose value is associated with the active filter mechanism mediated by the outer hair cells. For example, by measuring the change in slope of the growth curve in two populations of normal and hearing impaired subjects, one can look for systematic differences between the average quality factors of the two populations:

$$\frac{\left[\frac{\partial p_{\mathrm{DP}}^{\mathrm{I}}}{\partial p_2} - \frac{\partial p_{\mathrm{DP}}^{\mathrm{II}}}{\partial p_2}\right]_{\mathrm{normal}}}{\left[\frac{\partial p_{\mathrm{DP}}^{\mathrm{I}}}{\partial p_2} - \frac{\partial p_{\mathrm{DP}}^{\mathrm{II}}}{\partial p_2}\right]_{\mathrm{impaired}}} = \frac{Q_{\mathrm{normal}}}{Q_{\mathrm{impaired}}} = \frac{b_{\mathrm{impaired}}}{b_{\mathrm{normal}}}, \tag{27}$$

where the forward–backward transmission parameters included in the constant C have been implicitly assumed to be the same in the two populations.

Similarly, by computing the ratio between the change in slope of the average growth-rate curves relative to different DPOAE frequencies, one gets information on the frequency dependence of the average quality factor.

The details of the transmission along the cochlear membrane and interference effects between the two sources, which are taken into account only by full cochlear models (e.g., Talmadge *et al.*, 1998), are not predicted by this oversimplified two-sources embedded model. Despite these limitations, we show that, using differential parameters and/or averaging over a consistent number of subjects, the model predictions could be successfully compared with experimental data.

IV. RESULTS

The experimental average DPOAE I/O functions for the normal-hearing and the high-frequency impaired classes of ears are shown in Fig. 2(a), using the conventional dB–dB plot. The same growth curves, expressed in adimensional linear pressure units normalized to the standard pressure $p_0 = 20\ \mu$Pa, are shown in Fig. 2(b). We will always refer to such linear–linear plots when discussing slopes and amplitudes of the DPOAE curves, while the log–log plot is also provided for a comparison with earlier studies. In the ears with hearing loss there is a significantly lower level and slope of the DPOAE response. The data shown in Fig. 2(b) may be represented by two linear regimes, with a smooth change in the slope of the response–stimulus curve, occurring at a stimulus sound pressure of order $1500p_0$ ($\approx$65 dB SPL). As suggested in the previous section, these two linear regimes correspond to the saturated-linear range described by Eqs. (23) and (25), and to the fully-saturated range described by Eqs. (24) and (26). The slopes are esti-

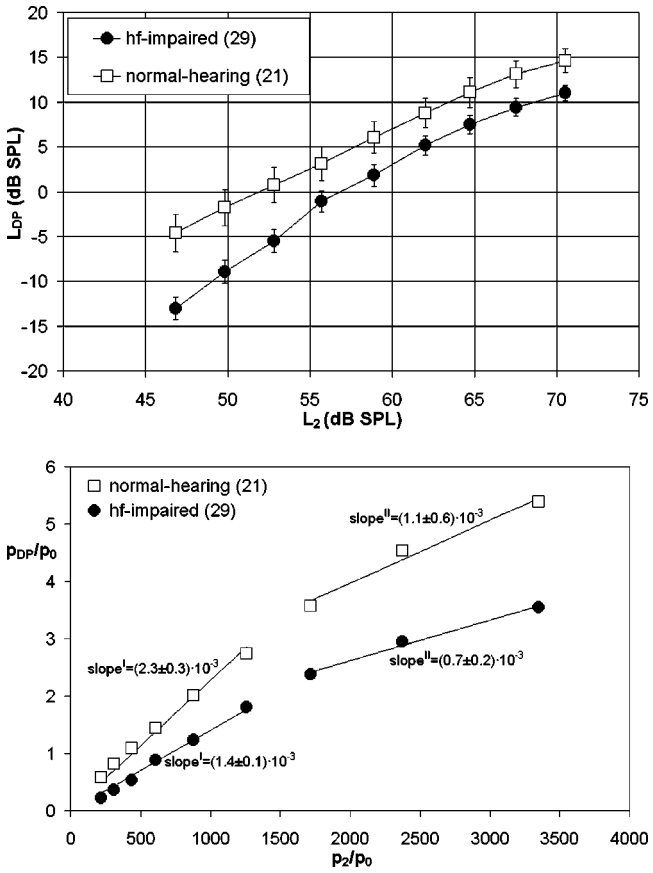


FIG. 2. Growth rate of the DPOAE response ($f_2=4$ kHz) averaged over two classes of normal-hearing and hf-impaired ears. In (a) pressure levels L are expressed in dB SPL units, while in (b) sound pressure is given in adimensional linear units of $p_0=20\ \mu$Pa. Best-fit straight lines are also shown.

mated by fitting the data to two independent linear functions, imposing that the low amplitude fit functions pass through the origin of the axes. The statistical errors on the slopes are also shown.

It should be stressed here that in the present model the change in slope is assumed to be dependent on the quality factor of the cochlear place x_{DP}. While it is still true that the emission comes mainly from the place that is resonant at frequency f_2, our estimates of the quality factor are actually relative to f_{DP}.

Using Eq. (27), the estimated ratio between the average quality factors at $f_{DP}=2.6$ kHz, ($f_2=4$ kHz) of normal and impaired ears is obtained:

$$Q(n)/Q(i)=1.7\pm 1.1. \tag{28}$$

In 8 nonexposed normal ears, the DPOAE growth-rate curve was measured for two different stimulus frequencies: $f_2=2$ kHz and 4 kHz. The two average growth-rate curves are shown in Fig. 3. The slopes of the curves in the two regimes described by Eqs. (25) and (26) are reported with the experimental errors. The comparison between the slopes of the two curves, using the above model, yields an estimate of the frequency dependence of the quality factor of the cochlear resonances. The estimated ratio between the quality factors at the two f_{DP} frequencies is

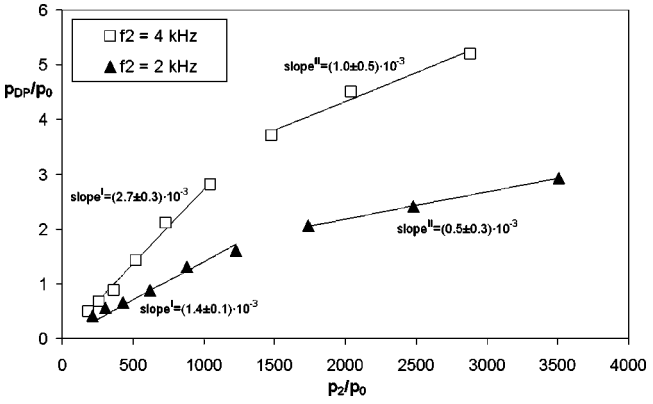


FIG. 3. A comparison between the average DPOAE growth rate curves recorded for $f_2=4$ kHz and $f_2=2$ kHz in 8 normal-hearing ears of nonexposed subjects. Pressure is expressed in adimensional linear units of $p_0=20\ \mu$Pa. Best-fit straight lines are also shown.

$$Q(f_{DP}=2.6\ \text{kHz})/Q(f_{DP}=1.3\ \text{kHz})=1.8\pm 1.2. \tag{29}$$

V. DISCUSSION

The results show that the DPOAE growth rate is sensitive to mild hearing loss, associated with exposure to industrial noise. The behavior shown in Fig. 2(b) may be interpreted using our simple two-sources model of the DPOAE generation. The uncertainty on the slopes, and consequently on the estimated Q ratios, is quite large. Nevertheless, the results still deserve to be interpreted. The observed difference between impaired and normal hearing ears was expected, due to the lower effectiveness of the cochlear active filter in the impaired ears. In fact, the active feedback mechanism implies a negative contribution to linear damping, which is associated, in this model, with a lower value of the coefficient b of Eq. (1), as well as with the presence of the negative term associated with the existence of spontaneous emissions. Thus the damage of the outer hair cells induced by noise is expected to produce an increase of the coefficient b. Incidentally, it should be noted that the observed behavior of the growth rate curves would not be expected by a model using the van der Pol oscillator instead of the oscillator of Eq. (1). In fact, the van der Pol equation admits no linear damping regime, because the limit-cycle regime associated with the existence of SOAEs comes from the balance between a quadratic nonlinear damping term and a linear antidamping term. The necessity of a linear damping term, whose coefficient amplitude b is reduced by the active filter mechanisms, is suggested by this work, confirming earlier analyses of the relaxation dynamics of evoked OAEs (Sisto and Moleti, 1999; Sisto *et al.*, 2001). A recent experimental study of the DPOAE STCs in normal-hearing and hearing-impaired subjects found no correlation between the measured Q_{10} at 4 kHz and the hearing threshold (Gorga *et al.*, 2003). Even if both measurements are affected by large uncertainties, we note that this result does not agree with the findings of the present work, but we would like to stress that the Q_{10} estimated from the DPOAE STCs is not easily comparable to the Q defined in the present work, unless a precise cochlear model of the DPOAE generation is used also for the interpretation of the DPOAE STCs. In our case, we assume to be

sensitive to changes in the quality factor of the cochlear place x_{DP}, driven by a small amplitude secondary tone of frequency f_{DP}, while the STCs' experiments are sensitive to the bandwidth of the resonator at the cochlear place x_2, driven by the high amplitude primary tones of frequencies f_1 and f_2.

Our estimate of the ratio between the tuning at $f_{DP}=2.6$ and 1.3 kHz in normal ears shows an increase of Q with frequency, which is in qualitative agreement with previously published OAE-based estimates of the cochlear tuning curve. For example, Moleti and Sisto (2003) found $Q(2.6\text{ kHz})/Q(1.3\text{ kHz})\approx 2$, using TEOAE latency data of a population of normal hearing adult subjects, while Kummer *et al.* (1995) found a slower dependence on frequency [$Q_{10}(4\text{ kHz})\approx 4$, $Q_{10}(2)\approx 3$, $Q_{10}(1)\approx 2.5$] from the analysis of DPOAE STCs. A recent study, including both behavioral and otoacoustic estimates of cochlear tuning (Shera *et al.*, 2002) finds high absolute values of Q, with a power-law dependence on frequency with an exponent≈ 0.3, which would correspond to a ratio $Q(2f)/Q(f)\approx 1.25$ for any frequency f.

In scale-invariant models the quality factor is independent of frequency (Talmadge *et al.*, 1998). The experimental results of this work further support the evidence that this symmetry is only approximately valid in the real cochlea.

VI. CONCLUSIONS

A simple embedded oscillator cochlear model has been applied to the study of DPOAE growth rate in subjects exposed to noise. As in the case of the TEOAE relaxation dynamics, the new oscillator equation provides a simple theoretical framework for the interpretation of the experimental data. In the preliminary analysis of a composite audiometric-OAE database which is currently being accumulated, significant differences were observed between the amplitude and slope of the growth rate curves of two populations of normal hearing and hearing impaired ears, both exposed to industrial noise. The parameters of the proposed oscillator could be useful to describe the effects of cochlear damage in exposed populations, namely the decreased effectiveness of the nonlinear active filter mechanism.

ACKNOWLEDGMENTS

This work was supported by ISPESL in the framework of the Finalized Project on Otoacoustic Emissions of the Italian Ministry of Health. We wish to thank Dr. V. Monechi, Ms. A. Citroni, Dr. S. Pellegrini, and Ms. S. Chelotti for having collected the experimental data.

Abdala, C. (**2000**). "Distortion product otoacoustic emission ($2f1$-$f2$) amplitude growth in human adults and neonates," J. Acoust. Soc. Am. **107**, 446–456.

Abdala, C. (**2001**). "Maturation of the cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels," J. Acoust. Soc. Am. **110**, 1465–1476.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**, 1810–1818.

Dhar, S., Talmadge, C. L., Long, G., and Tubis, A. (**2002**). "Multiple internal reflection in the cochlea and their effect on DPOAE fine structure," J. Acoust. Soc. Am. **112**, 2882–2897.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). "Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. **110**, 3119–3131.

Gaskill, S. A., and Brown, A. M. (**1990**). "The behavior of the acoustic distortion product, $2f1$-$f2$, from the human ear and its relation to auditory sensitivity," J. Acoust. Soc. Am. **88**, 821–839.

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.

Kalluri, R., and Shera, C. (**2001**). "Distortion-product source unmixing: a test for the two-mechanism model for DPOAE generation," J. Acoust. Soc. Am. **109**, 622–637.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**1995**). "Suppression tuning characteristics of the $2f1$-$f2$ distortion-product otoacoustic emission in humans," J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). "Optimal $L(1)$-$L(2)$ primary tone level separation remains independent of test frequency in humans," Hear. Res. **146**, 47–56.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999a**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. I. Fine structure and higher-order DPOAE as a function of the frequency ratio $f2/f1$," J. Acoust. Soc. Am. **106**, 3473–3483.

Mauermann, M., Uppenkamp, S., van Hengel, P. W., and Kollmeier, B. (**1999b**). "Evidence for the distortion product frequency place as a source of distortion product otoacoustic emission (DPOAE) fine structure in humans. II. Fine structure for different shapes of cochlear hearing loss," J. Acoust. Soc. Am. **106**, 3484–3491.

Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.

Neely, S. T., Gorga, M. P., and Dorn, P. A. (**2003**). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **114**, 1499–1507.

Oswald, J. A., and Janssen, T. (**2003**). "Weighted DPOAE input/output-functions: a tool for automatic assessment of hearing loss in clinical application," Z. Med. Phys. **13**, 93–98.

Popelka, G., Osterhammel, P., Nielsen, L., and Rasmussen, A. (**1993**). "Growth of distortion product otoacoustic emissions with primary tone level in humans," Hear. Res. **71**, 12–22.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. (**2003**). "Sources and mechanisms of DPOAE generation: implications for the prediction of auditory sensitivity," Ear Hear. **24**, 367–379.

Shera, C. A., Guinan, Jr., J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Sisto, R., and Moleti, A. (**1999**). "Modeling otoacoustic emissions by active non linear oscillators," J. Acoust. Soc. Am. **106**, 1893–1906.

Sisto, R., Moleti, A., and Lucertini, M. (**2001**). "Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects," J. Acoust. Soc. Am. **109**, 638–647.

Talmadge, C. L., Long, G., Tubis, A., and Dhar, S. (**1999**). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **105**, 275–292.

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.

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.

Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). "Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L2 below L1," J. Acoust. Soc. Am. **97**, 2346–2358.

Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L.,

and Martin, G. K., (**1995b**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L1, L2 space,'' J. Acoust. Soc. Am. **97**, 2359–2377.
Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input–output functions,'' Hear. Res. **123**, 87–96.
Withnell, R. H., Shaffer, L. A., and Talmadge, C. L. (**2003**). ''Generation of DPOAEs in the guinea pig,'' Hear. Res. **178**, 106–117.

Frequency specificity of distortion-product otoacoustic emissions produced by high-level tones despite inefficient cochlear electromechanical feedback[a)]

Sirley Carvalho, Thierry Mom, Laurent Gilain, and Paul Avan[b)]
Laboratory of Sensory Biophysics, School of Medicine, Clermont-Ferrand, France

(Received 3 March 2004; revised 7 June 2004; accepted 7 June 2004)

Distortion product otoacoustic emissions (DPOAEs) are thought to stem from the outer hair cells (OHCs) around the normally narrow place tuned to the primary tone stimuli. They are thus said to be frequency-specific: their local absence should accurately pinpoint local OHC damage. Yet the influence of impaired tuning on DPOAE frequency specificity is poorly documented. Mice with local damage to OHCs were examined. Their DPOAEs were frequency-specific in that audiometric notches were accurately tracked. The same cochleae were further impaired by ischemia or furosemide injection inducing strial dysfunction with flat loss of sensitivity and tuning, while the preexisting pattern of damaged OHCs remained unaltered. Despite the loss of cochlear activity, DPOAEs produced by high-level ($\geqslant$70 dB SPL) primaries remained large in about the same interval where they had been initially normal, i.e., that with nondamaged OHCs, albeit with a slight frequency shift, of -1.1 kHz on average. Thus, the ability of DPOAEs to map structurally intact OHCs cannot be a mere consequence of cochlear tuning as it largely persists in its absence. The key element for this correct mapping is likely part of intact OHC structures (e.g., stereocilia bundles) and must have some tuning of its own. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777873]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 1639–1648

I. INTRODUCTION

Because the processing of sound by the cochlea entails nonlinearities (Eguiluz *et al.*, 2000), basilar membrane motion in response to pairs of primary tones at frequencies f1 and f2 contains intermodulation components at arithmetical combinations of f1 and f2. These components can be detected as otoacoustic emissions (OAEs; Kemp, 1978) in the ear-canal sound pressure, where they are termed distortion-product OAEs (DPOAEs).

Being the strongest in mammalian ears, the DPOAEs produced at $2f1-f2$ have been the most studied ones. A vast body of evidence (reviewed in Kemp, 2002) points to outer hair cells (OHCs) as being at their origin and accordingly they are widely used for noninvasive assessment of cochlear function (Martin *et al.*, 1990). There may be more to DPOAEs than a mere hearing-screening technique: owing to their electromotile properties (Brownell *et al.*, 1985), OHCs are thought to be at the core of a subtle cochlear mechanism first hypothesized by Gold (1948) and involving a sequence of mechano-electrical, electro-mechanical and coupling events (review in Patuzzi, 1996) organized in a regenerative feedback loop. This looped mechanism is believed to be responsible for a triad of characteristic properties of normal cochlear micromechanics, i.e., at the same time for its exquisite sensitivity, sharp tuning and compressive nonlinearity, with DPOAEs as a by-product of either the whole feedback loop or some of its stages.

As their measurement is noninvasive and very fast, DPOAEs provide a unique tool for probing the aforesaid triad. Hearing sensitivity depends on the gain achieved through the OHC-based feedback, the impairment of which affects DPOAEs so that they vanish when the hearing loss due to OHC damage exceeds a screening limit around 30–40 dB HL (e.g., Martin *et al.*, 1990; Avan and Bonfils, 1993; Hamernik *et al.*, 1996, Kemp, 2002). Screening ability is not specific of DPOAEs as transient-evoked OAEs perform in a quite similar manner. However, owing to their being generated by narrow-band stimuli, DPOAEs are thought to be frequency-specific in that they should provide local information regarding their sources, that is, where maximum interaction occurs between f1 and f2 tones. There is evidence that this happens around the places tuned to f2 or f1 (Martin *et al.*, 1987; Allen and Fahey, 1992; Avan *et al.*, 1998). Moreover, in the normal cochlea, the tuning is so sharp that the excitation pattern of a pure tone along the basilar membrane barely spreads beyond a few hundred μm at low levels (Ren, 2002). Consequently, the place where DPOAEs are generated should be equally narrow, thus it should be possible to correlate presence or absence of DPOAEs at $2f1-f2$ with intact or damaged nonlinear elements around a well-defined place tuned to the primary stimuli. A useful presentation making implicit use of this notion is the so-called DPgram representing DPOAE amplitudes against either f2, or the numerically quite similar geometric mean of f1 and f2.

To assess whether DPOAEs are frequency-specific, correlations between audiograms and DPgrams have been studied: several early publications reported them to be significant

[a)]A preliminary account of this work was presented at the 26th Midwinter Meeting of the Association for Research in Otolaryngology, 23–27 February 2003, Daytona Beach, FL.

[b)]Author to whom all correspondence should be addressed: Laboratory of Sensory Biophysics, P.O. Box 38, School of Medicine, 63001 Clermont-Ferrand, France. Electronic mail: paul.avan@u-clermont1.fr

(Martin *et al.*, 1990; Avan and Bonfils, 1993). Further works have confirmed the high correlations among DPgrams, auditory brainstem response (or ABR) audiograms and presence or absence of OHCs at the place tuned to the frequency of interest as revealed on cytocochleograms (to mention only a few examples, Hamernik *et al.*, 1996; Le Calvez *et al.*, 1998; Harding *et al.*, 2002). Actually, correlations between DPgrams and audiometry were found every time hearing loss was shown to result from effectively damaged or missing OHCs (Avan *et al.*, 2003). Conversely, strial dysfunction in the presence of intact OHCs (Mills and Rubel, 1994; Mom *et al.*, 1997) produced impaired audiograms coexisting with extant DPOAEs when induced by high-level stimuli >60 dB SPL. At first, a second source of DPOAEs was invoked, termed ''passive'' in view of its apparent lack of vulnerability in contrast with ''active'' cochlear operation (Mills, 1997). Later, a series of reports showed that a single source could account for DPOAE properties regardless of stimulus level (Lukashkin *et al.*, 2002) and that extant DPOAEs produced by higher level stimuli still probed OHCs, and likely resulted from nonlinearities in their transduction channels or stereocilia (Avan *et al.*, 2003). These authors argued that strial dysfunction had no reason to affect DPOAE presence provided stimulus levels were high enough to make up for the loss of cochlear gain and ensure that nonlinear elements could be driven into the nonlinear part of their characteristics. Thus, although the idea to perform a sort of objective ''audiometry'' with DPOAEs produced by high-level tones has to be given up when cochlear dysfunction is of strial origin, these DPOAEs likely keep probing the presence of undamaged OHCs.

If this idea of using high-level DPOAEs is to find some useful application, e.g., in clinical situations, the notion of their frequency specificity has to be addressed. As explained above, mapping OHCs with the help of DPgrams relies upon the premise that DPOAEs are produced in a narrow interval along the cochlear partition, as the accuracy of the picture given by DPOAEs of their sources cannot be better than the width over which their generation site can spread. It is indeed widely accepted that excitation patterns along the basilar membrane smear basalward in relation to saturation, at increasing primary levels, or absence of active elements, in a damaged cochlea, so that inefficient activity should result in auditory filter broadening (Moore and Glasberg, 1983; Rosen and Stock, 1992). In spite of that, a number of experiments suggest that the close correspondence between DPgrams and audiograms hardly suffers from increasing the levels of primary tones to 70 dB SPL or more (Bonfils *et al.*, 1997; Avan *et al.*, 2003), and that several features of OAEs fail to react to auditory filter broadening with increasing level (Leeuw and Dreschler, 1998).

Another related issue arises from recent results suggesting that the relation of DPOAE tuning to cochlear tuning may not be straightforward. Tuning at the DPOAE level is studied by building suppression tuning curves (STCs), representing the level of a suppressor tone required to reduce DPOAE level by some predetermined criterion against suppressor frequency (Brown and Kemp, 1984; Abdala, 1998; Kummer *et al.*, 1995; Martin *et al.*, 1998). In a normal cochlea, their V-shaped profiles resemble those of basilar membrane, single nerve fiber and psychophysical tuning, which raised some hopes that DPOAE STCs could be used for noninvasive assays of cochlear tuning (e.g., Abdala, 1998). Following a similar premise, Shera *et al.* (2002) derived estimates of human cochlear tuning from measurement of otoacoustic emission group delays, using single tones as stimuli instead of pairs of tones for DPOAEs. Yet, the essential relation of DPOAEs to cochlear tuning is challenged by discrepancies between disrupted cochlear amplification implying broadened neural tuning curves with marked decrease of Q10dB, and preserved DPOAE STCs (Martin *et al.*, 1998; Howard *et al.*, 2003). Among other interpretations, DPOAEs might stem from inherently tuned sources more robust than the electromotility-based feedback loop as a whole.

The goal of this work is to check to what extent DPOAEs are frequency-specific at high stimulus levels, whether it requires intact cochlear tuning and how it varies when it gets lost. The first experimental stage used cochleae with local anatomical damage to OHCs, from either genetically defective or acoustically traumatized mice of the strain CD1. These mice are known for exhibiting a large variety of patterns of basal OHC losses whereas the apical cochlea remains normal (Shone *et al.*, 1991). The maps of healthy versus absent OHCs and normal versus impaired function match each other (Le Calvez *et al.*, 1998). CD1 mice whose audiograms were found to be normal were exposed to a loud tone with a level and duration such that an acute acoustic trauma was produced above 6 to 14 kHz. Studies of how DPOAEs react to acoustic trauma have led to results similar to those with OHC loss of genetic origin (e.g., Harding *et al.*, 2002). Whether genetic or due to acoustic trauma, this first stage, called *OHC loss/damage* hereafter, was designed so that DPOAEs were strongly decreased or absent at high frequencies even at high stimulus levels, and hearing thresholds were elevated in the corresponding range. Conversely, low-frequency responses were normal and the transition between the two regions was sharp. Thus an equally sharp anatomical transition most likely existed between healthy apical OHCs and damaged or absent more basal OHCs (Le Calvez *et al.*, 1998). DPOAEs were combined with standard electrophysiological tests to check how accurately the localized OHC lesions could be mapped, and how cochlear tuning was affected inside the damaged frequency interval.

At a second stage, the function was further compromised in the same cochleae but this time by a totally different means. Either ischemia or injection of a loop diuretic was used, resulting in a decrease in endocochlear potential (EP) and likely a dramatic loss of cochlear tuning at all frequencies (Evans and Klinke, 1982; Sewell, 1984; Ruggero and Rich, 1991). This second impairment affected the initially intact apex of the cochlea without the least modifying the pattern of damaged OHCs and the boundary between intact apical and damaged basal OHCs: neither ischemia nor furosemide injection can damage extant OHCs, at least, for ischemia, within the first hour post-onset (Billett *et al.*, 1989). At this stage, termed *strial/EP abnormality* hereafter, the places where OHCs were damaged from the beginning, as well as those with intact OHCs, no longer benefited from the

electromotility-based feedback loop that normally enhances basilar membrane motion and cochlear tuning. In sharp contrast with the basal cochlea at the OHC loss/damage stage, DPOAEs could persist from the areas with intact OHCs, provided high enough stimulus levels were used [70 dB SPL or more (Mom *et al.*, 2001; Avan *et al.*, 2003)]. The DPgrams obtained at such levels were compared to DPgrams collected at much lower stimulus levels at the first stage, i.e., at a time when part of the cochlea was still fully active and normally tuned.

II. MATERIAL AND METHODS

Mice belonging to strain CD1 (Swiss albino from the local Charles-River laboratories) were used. In young animals (8 to 16 weeks of age in the present work), the progressive degeneration of sensory cells affects only the basal OHCs and is such that audiograms and DPgrams (see protocols below) often exhibit sharp, concomitant transitions between normal and pathological intervals (Shone *et al.*, 1991; Le Calvez *et al.*, 1998).

Mice were placed in a sound-insulated booth and deeply anaesthetized by an i.p. injection of ketamine chlorhydrate (7 mg for a 30-g mouse, renewed as needed), initially administered with levomepromazin chlorhydrate (0.1 mg). The core temperature was kept at 39 °C by a regulated heating pad. First, a punctured myringotomy was performed under microscopic control in order to ensure long-term middle-ear pressure stability. Then, the tapered tip of an acoustic probe containing a sensitive microphone (ER10B, Etymotic Research) and two plastic sound tubes connected to the outputs of two earphones (ER2, Etymotic Research) was sealed in the external auditory canal of the right ear, using cyanoacrylate glue.

A. Principle of DPOAE recordings

A computer-controlled CubeDis system collected the DPOAEs (Allen, 1990; software Cubdisp v2.43, Mimosa Acoustics). The protocol was identical to that of previous publications (Mom *et al.*, 2001; Avan *et al.*, 2003). Briefly, after automatic in-the-ear calibration, two primary tones at frequencies f1 and f2 ($f2/f1 = 1.20$) were emitted with the same levels ($L1 = L2$) in the right ear canal by the two earphones. The processor board of the CubeDis system extracted the spectral component at $2f1 - f2$ from the acoustic pressure detected by the probe microphone, and a few components close to $2f1 - f2$ provided an estimation of the noise floor. For every DPgram, f2 was swept from about 20 to 3 kHz (10 descending steps/octave, 0.5 to 1 s/step). Primary levels were increased from 40 to 80 dB SPL in 10-dB steps, and occasionally, in 1- to 5-dB steps at the last stage of an experiment.

B. Principle of electrophysiological recordings

1. Toneburst evoked ABRs

Frequency-specific auditory brainstem response (ABR) thresholds were evaluated with the help of tone bursts (3–40 kHz, rise and decay time 1 ms, plateau duration 4 ms) generated by an arbitrary waveform synthesizer (Wavetek 70) at a repetition rate of 50/s. They were sent to a high-voltage amplifier with built-in equalizer driving a $\frac{1}{2}$-in. condenser microphone (Bruel and Kjaer 4134) serving as emitter. Its tapered tip was sealed in the right ear canal. Stimulus levels varied in 5-dB steps from 0 dB SPL to 10 dB above the level required to get a first repeatable ABR response. Acoustic calibration was performed in the ear canal by means of a thin probe microphone (Bruel and Kjaer 4182) ensuring reliable measurements up to 40 kHz. Stimulus frequencies were 3, 5, 10, 15, 20, 30 and 40 kHz, and around the transition from normal to elevated ABR thresholds, finer 0.5- to 1-kHz steps were used in order to improve the definition of the normal-to-impaired boundary.

Electric responses were recorded between subcutaneous stainless steel electrodes at the vertex and ipsilateral mastoid, with the lower back serving as ground. The signal was preamplified (Grass P511, $\times 200\,000$, bandpass 0.3–3 kHz) and 200 to 500 epochs were averaged by a CED 1401+ interface driven by Sigavg software (v.6.04, Cambridge Electronic Design). When the stimulus level neared the auditory threshold, only the ABR wave with the longest latency was clearly distinguished from the noise background. The stimulus level was varied in 5-dB steps and the ABR threshold at every frequency was defined as the smallest stimulus level needed for producing at least one repeatable wave $> 0.15\ \mu$V.

2. ABR suppression tuning curves

Tone bursts were emitted at frequency f_{TB} with a level set at 5 dB above ABR threshold at f_{TB}, thus allowing easy identification of a stable ABR response after averaging. A continuous pure tone with varying frequency f_{mask} and level L_{mask} was then added up to the tone bursts in the external ear canal. The tone frequency f_{mask} was varied in 0.5-kHz steps around f_{TB} and the level L_{mask} just sufficient to abolish the ABR waves was evaluated. The plot of this level against f_{mask} represented a neural STC.

C. Experimental protocols

1. First stage, OHC loss/damage

This stage often spontaneously occurred in young CD1 mice owing to their genetic defect, showing up as a sharp drop of DPOAE levels—from near normal values at lower frequencies, down to the noise floor—as well as an equally sharp increase in hearing thresholds, around some frequency between 6 and 14 kHz. A sample of mice presenting this pattern was included ($n = 6$).

Other young CD1 mice exhibited normal DPgrams and audiograms up to 40 kHz. A sample of such mice was submitted to a loud pure tone presented in their right ear for 5 to 10 min (95 SPL, 6 to 12 kHz) in order to produce notches in the DPgrams at frequencies as similar as possible to those of the genetically defective mice. The loud tone was produced by a Wavetek 70 waveform generator, amplified by a custom-made voltage amplifier and played in the ear canal through the Bruel and Kjaer earphone already used for ABR and CAP measurements. Calibration of the tone level was performed with the help of a Bruel and Kjaer probe microphone 4182 in a custom-made 0.5 cm^3 artificial ear, thought

to be fairly representative of the setup in the mouse's ear. The effects of the loud tone were assessed immediately after ending the tone exposure, by DPgram and ABR measurements. If they were such that, at L1 = L2 = 70 dB SPL, DPOAE amplitude did not drop below −10 dB SPL at least at three frequencies of the DPgram, the tone exposure was deemed insufficient and was immediately repeated at 100 dB SPL for 10 min. When needed, this second exposure always led to a sufficient effect. DPgrams were then collected at 70 dB SPL every 5 min for the next 20 min in order to check that early recovery from auditory fatigue was not large enough for the lowest DPOAEs to reach above −10 dB SPL, otherwise the animal would be discarded. Eventually, $n = 11$ animals with deep, stable notches in their DPgrams were included.

The required sharp transition between normal and pathological DPOAEs showed up on the DPgrams as a ''low-pass filter'' profile whose roll-off was characterized with the help of parameter DP-rolloff, defined as the first frequency such that the DPOAE amplitude started dropping by more than 5 dB within one-tenth of an octave. Likewise, on the ABR audiograms, ABR-rolloff was defined as the first frequency such that the ABR threshold reached 15 dB nHL or more. Next, two frequencies were chosen, one well inside the range with normal ABR thresholds and one in the elevated range, well above ABR-rolloff. Neural STCs were built at these two frequencies, and their Q10dBs were measured.

2. Second stage, strial/EP abnormality

To complement the hearing loss due to local OHC pathology, broadband cochlear dysfunction was produced in the same cochleae either by complete cochlear ischemia (in a subsample of mice; $n = 3$ from group OHC loss/damage; $n = 8$ from group strial/EP abnormality) or by injection of loop diuretics furosemide ($n = 3$ from group OHC loss/damage; $n = 3$ from group strial/EP abnormality). Cochlear ischemia was obtained by anaesthetic overdose inducing immediate heart failure. Furosemide intoxication required an injection of at least 100 mg/kg furosemide (subcutaneous; furosemide requires a high *p*H to be active: the use of i.p.-administered ketamine chlorhydrate for anesthesia generates a low *p*H in the peritoneal cavity, such that i.p.-injected furosemide precipitates, thereby getting inactivated). Repeated controls of the DPgrams using 60-dB primary tones evaluated how efficiently ischemia or loop diuretics knocked down the cochlear gain: once achieved, DPOAEs were expected to vanish at all frequencies. This happened within a few minutes after ischemia, and after about 30 min for furosemide. ABR thresholds were immediately assessed, then the series of measurements specific of this stage was carried out as follows.

DPgrams were plotted for L1 = L2 varying stepwise (2-dB steps) from 60 up to at least 75 dB SPL. If a transition between low- and high-level DPOAEs showed up on a DPgram so as to generate a ''low-pass filter'' profile, its roll-off was characterized with the help of the already defined parameter DP-rolloff. ABR audiograms could seldom be processed as they had been at the first stage with OHC damage because no ABR wave could be found in ischemized animals, and the effect of furosemide on the cochlea is known to be such that inner hair cells as well as OHCs are impaired: contrary to DPgrams, ABRs reflected the combination of both impairments. Thus, ABR threshold increases sometimes largely exceeded 60 dB, thereby precluding neural STCs to be reliably built. Eventually, STCs were available only in two ears and, even then, only at frequencies where ABR thresholds had been normal at the first stage with OHC damage.

If the sharp transition between DPOAEs in healthy versus damaged frequency intervals observed at the stage with OHC loss/damage (and to be ascertained here) were a reflection of cochlear tuning, it should get blurred or even vanish in the presence of strial/EP abnormality. If, instead of depending on cochlear tuning, this sharp transition resulted from some property of physically intact OHCs regardless of their actual ability to exert efficient feedback on the cochlea, it should persist even though its characteristics, for example its position along the frequency map, might be somewhat altered.

The care and use of all animals were approved according to the rules promulgated by the French Department of Agriculture (authorization 63–105 to author PA).

III. RESULTS

A. OHC loss/damage

1. DPgrams and ABR thresholds

In all included mice, localized OHC pathology was such that ABR thresholds, while being normal at low frequencies, were elevated by more than 30 dB at high frequencies. The transition between normal and impaired intervals happened suddenly at ABR-rolloff. Figure 1(a) depicts averaged ABR thresholds across frequencies in the subsample of seven ears exhibiting an ABR-rolloff at 9–10 kHz: they remained within the normal range up to this limit, then sharply increased to 30–40 dB nHL between 10 and 12 kHz. Well above ABR-rolloff, the observed threshold elevations depended little on frequency and stabilized between 30 and 50 dB. This held irrespective of the cause for deafness, whether genetic or due to tonal overexposure, thus the seven ABR audiograms used to generate Fig. 1(a) were a mixture of the two origins. Similar, horizontally shifted profiles were found in the mice with ABR-rolloffs falling between 6–9 and 10–14 kHz.

Likewise, DPOAEs were normal at low frequencies and exhibited a marked decrease at all primary levels above some DPOAE roll-off frequency DP-rolloff, so that a sharp boundary separated the normal and impaired regions. In Figs. 1(b)–1(d) (respectively, for 70, 60 and 50 dB SPL primary levels), average DPOAEs obtained from the same ears as Fig. 1(a) ($n = 7$,ABR–rolloff = 9–10 kHz) departed from their pretrauma values (shown in thin lines, open symbols) above a DP-rolloff equal to ABR-rolloff. As for ABRs, none of the characteristics of DPgrams depended on whether the local OHC pathology was due to genetics or overexposure: the slope of the 70 dB SPL-DPgram above DP-rolloff, characteristic of how sharply the normal and damaged regions differed from each other, was −13.3 dB/kHz (s.d. 5.7) in genetically impaired mice, and −11.7 dB/kHz (s.d. 1.7) in

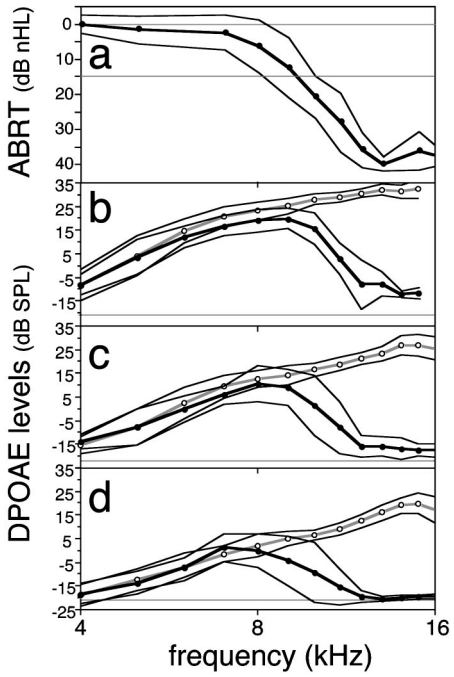


FIG. 1. Averaged responses of the subsample of ears with ABR-rolloffs around 9–10 kHz at the stage with OHC loss/dysfunction ($n=7$). (a) ABR thresholds, means ± one s.d. Thin horizontal line: traditional clinical limit between normal and impaired thresholds. (b,c,d) DPgrams (means ± one s.d.) in response to 70, 60 and 50 dB SPL primary tones respectively. At the beginning of an experiment, DPOAEs were normal (shaded lines, open circles). With OHC loss/dysfunction, DPOAEs (solid lines, closed circles) departed from normal above 10 kHz at all primary levels, dropped below −10 dB SPL above 12 kHz, while remaining strictly normal up to 8 kHz. Thin horizontal lines in panels (b)–(d): average background noise.

overexposed mice. Likewise, roll-off frequencies were similarly scattered between 6 and 14 kHz in both cases. This prompted us to pool all DPOAE data and, for what follows, all data, irrespective of OHC pathology.

It can be seen on Figs. 1(b)–(d) that DPgrams at all primary levels performed equally well for separating normal and impaired frequency intervals, and that DP-rolloffs did not depend on primary level. A Pearson regression analysis of ABR versus DP roll-offs at 70 dB SPL in all mice revealed a highly significant correlation described by the following linear regression:

$$\text{DP-rolloff (kHz)} = 1.65 + 0.90 \times \text{ABR-rolloff (kHz)},$$

$$r^2 = 0.79; p < 0.001.$$

The slope of this regression line is near 1, showing that the two roll-off frequencies closely paralleled each other. Its intercept of 1.65 kHz may probably be ascribed to the different definitions of roll-off frequencies on DPgrams versus ABR diagrams: that of ABRs relied upon the standard 15-dB limit inspired from the clinical definition of hearing loss whereas the DPgram roll-offs were placed with reference to the frequency where DPgrams tended to start plunging downward most clearly.

Neural STCs were V-shaped at frequencies with normal ABR thresholds: one example is shown in Fig. 2, where the probe frequency was set at 10 kHz and the probe level at 5 dB nHL (open triangle). STCs tended to lose their sharp tips

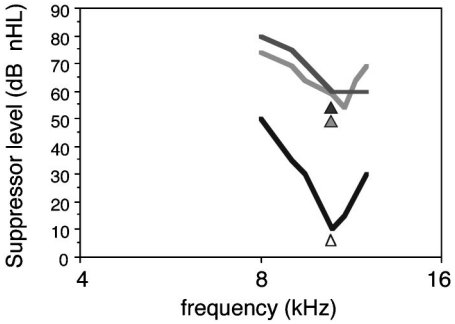


FIG. 2. Suppression tuning curves in a normal ear (bold black line) and in two pathological ears (shaded lines). The test tone-burst was set at 10 kHz and at the levels indicated by triangles (i.e., 5 dB above ABR threshold).

whenever ABR threshold elevations exceeded 30–40 dB (two examples on Fig. 2 with a 10-kHz probe frequency, the probe levels being set at 45 and 50 dB nHL, respectively). From an average of 4.7 (s.d. 1.0) when normal, Q10dBs decreased to 2.3 (s.d. 0.5).

This stage of OHC loss/damage did not involve any novel protocol, thus standard theories of cochlear physiopathology provide simple guidelines toward the interpretation of its outcome. Due to either sound overexposure or natural genetic impairment, OHCs basalward of some place were absent or damaged, in other words the nonlinear elements at the origin of DPOAEs were disabled. Wherever the OHC-based feedback loop was disrupted, the sensitivity and tuning were worse than normal, hence the blunt neural STCs. The borders between frequency intervals with versus without DPOAEs, and with normal versus increased ABR thresholds were clear-cut, suggesting that so was the topographical limit between intact and damaged areas.

B. Strial/EP abnormality

When general cochlear dysfunction was produced in the previously locally damaged cochleae, either by total cochlear ischemia or by injection of loop diuretics, the formerly normal DPOAEs, i.e., below DP-rolloff, exhibited large decreases while of course the previously missing DPOAEs above DP-rolloff remained very low or absent. Low-frequency DPOAEs in response to primary tone levels $\leqslant 60$ dB SPL were particularly affected, and were often hardly detectable above the noise floor. The example depicted in Fig. 3(a) shows that the DPOAEs that had been little affected by tonal overexposure at the OHC loss/damage stage of the experiment (bold line, below 11 kHz, with reference to the bolder line before exposure) dropped down after furosemide administration (shaded line and open circles). In all animals, this resulted in flat DPgrams, barely differing from the equally flat noise floor line [thin line, Fig. 3(a)].

It took only about 1 min after ischemia onset for ABR responses to vanish so that their thresholds could no longer be assessed. After furosemide injection, ABR threshold elevations exceeded 60 dB in all animals, but it has to be kept in mind that such losses encompass two components, a ''mechanosensory'' one in relation to OHC dysfunction, and a neural one due to inner hair cells losing some of their

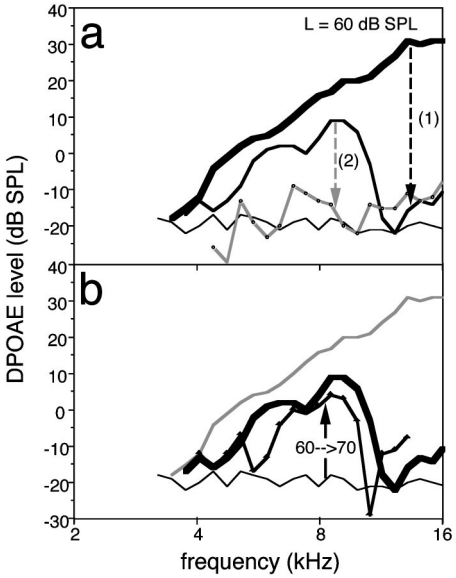


FIG. 3. (a) Example of DPgrams in an initially normal ear (bold line) with primary levels fixed at 60 dB SPL. After pure-tone overexposure (dashed arrow, labeled 1), the DPgram (thinner black line) exhibited a sharp drop, down to noise floor above 11 kHz (thinnest line). At the next stage reached with the help of furosemide injection (150 mg/kg subcutaneous, dashed shaded arrow labeled 2), the DPgram (shaded line and open circles) dropped to the noise floor even at low frequencies. (b) DPgrams from panel (a) serve as references (initial, shaded line and after overexposure, bold line). When primary levels increased from 60 to 70 dB SPL, the DPgram (thin black line, closed triangles) almost matched the reference one of the stage with OHC loss/dysfunction.

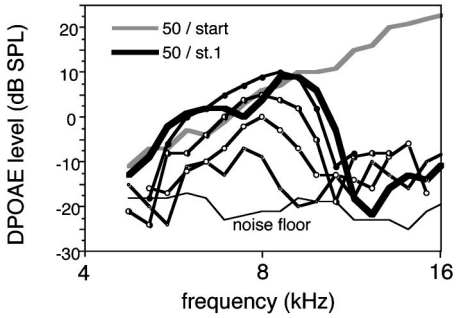


FIG. 4. Examples of initially normal DPgram with 50 dB SPL primaries (shaded line), DPgram at the stage with OHC loss/dysfunction (bold line), and stacked DPgrams 30 min post-mortem in response to increasing primary levels, from 65 to 71 dB SPL in 2-dB steps.

ability to respond as a result of furosemide-induced strial dysfunction. It would thus be difficult to infer from ABR threshold elevations how much cochlear gain was lost, except that this loss was of course well below the total threshold elevation. The high levels required for ABRs to be found allowed neural STCs to be obtained only in two ears at low frequencies. Their Q10 was <2.2, while having been >3.5 initially.

Whether originating from genetically impaired or overexposed mice and whether treated by ischemia or furosemide, all ears behaved so similarly in terms of what follows that the outcomes will be merged. When primary levels were $\geqslant 70$ dB SPL, DPOAE responses to strial/EP insults greatly differed from those of lower-level DPOAEs, in that residual DPOAEs were systematically found well above noise floor at low frequencies. Compared with Fig. 3(a), Fig. 3(b) shows that when the primary level reached 70 dB SPL, DPOAEs (closed triangles) resurfaced in the whole frequency range where they had been present at the previous stage, that is, below 11 kHz. They fell short of the initial 60 dB SPL responses (bold line) by only about 5 dB. This return of DPOAEs at higher primary levels restored the transition between low and high frequencies that had been characteristic of the OHC loss/damage stage. How it happened is depicted in Fig. 4. The initial DPgram at 50 dB SPL is shown in shaded lines, then the DPgram at the OHC loss/damage stage in bold lines. Thirty min after ischemia was produced, starting at 60 dB SPL from a hardly significant DPgram (not shown) buried in the noise floor around -20 dB SPL (thin line), DPOAEs began exhibiting some low-frequency responses at 65 dB SPL (lowest solid line). With stepwise increase of primary levels (2-dB steps, up to 71 dB SPL), the DPgrams matched increasingly better the reference DPgram at 50 dB SPL at the OHC loss/damage stage (bold line). The slope of the transition between present and absent DPOAEs even increased with increasing primary level, to reach the prefurosemide slope for primary levels of 71 dB SPL. In the example of Fig. 4, the frequencies at which the transition occurred did not change in the presence of strial/EP abnormality. Yet, in several other ears, a slight downward shift of the transition at the strial/EP abnormality stage was observed. On average, the transition between normal and impaired OHC areas moved by -1.1 kHz.

To reveal the general trends of DPgram changes when strial/EP abnormality was produced, each individual DPgram was horizontally shifted in such a way that the zero of its frequency scale was DP-rolloff measured at the OHC loss/damage stage. Once all DP-rolloffs were lined up at zero of the new scale (in octave frequencies), individual DPgrams before versus after strial/EP abnormality were averaged and the averages overlaid on each other (Fig. 5). According to visual inspection, the average DPgram at 70 dB SPL at the strial/EP abnormality stage exhibited the best amplitude match with the DPgram around 50 dB SPL at the OHC loss/damage stage (not shown on Fig. 5 for sake of clarity). Likewise, the average DPgram at 75 dB SPL at the strial/EP abnormality stage (shaded line) tended to level up with the average DPgram around 60 dB SPL at the OHC loss/damage stage (bold line). It confirms that the larger the primary level, the smaller the DPOAE level decrease, in keeping with what Avan *et al.* (2003) reported for ischemia and Mills and Rubel (1994) for furosemide. Besides, the average downward frequency shift of the transition from intact to damaged OHC areas is evident in Fig. 5. Small residual DPOAEs were often recorded above the cutoff frequency when using primary levels >70 dB SPL at the strial/EP abnormality stage, but they never exceeded 0 dB SPL while they should normally have been well above 30 dB SPL in this range. Residual higher-frequency DPOAEs were thus so grossly pathological that the genuine existence of a cutoff frequency can hardly be called into question since the lower-frequency part of DPgrams was rather near-normal.

At frequencies where large DPOAEs remained in response to high-level primaries, DPOAE growth functions

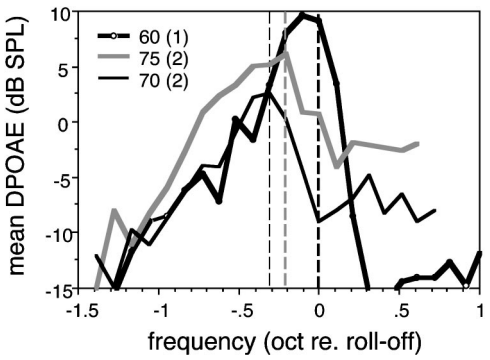


FIG. 5. Averaged DPgrams across all mice, referred to an octave frequency scale centered on each individual roll-off frequency at the first stage with OHC loss/dysfunction. At this first stage, the average DPgram with 60 dB SPL primary tones [bold line, labeled 60 (1)] thus exhibits a roll-off frequency at 0, marked by the bold dashed vertical line. At the second stage with strial/EP abnormality, the average DPgram in response to 70 dB SPL primary levels [thin black line, labeled 70 (2)] falls shy of the reference DPgram at the previous stage at 60 dB SPL, but the average DPgram in response to 75 dB SPL primary levels [shaded line, labeled 75 (2)] almost matches it as regards amplitudes. Yet DPgrams at the stage with strial/EP abnormality are shifted downward by 0.2–0.3 oct (thinner dashed vertical lines mark their roll-off frequencies). At all frequencies, standard deviations (not shown) were <5 dB, mainly reflecting the variability of initial DPgrams.

against stimulus levels exhibited an increase in slope, for example, from an average of 1.05 dB/dB (s.d. 0.23) around 8 kHz and 70 dB SPL in ears with normal auditory thresholds at this frequency before strial dysfunction was induced, to an average of 2.22 dB/dB (s.d. 0.58) in the presence of strial/EP abnormality. This was likely the result of a loss in mechanical compression at the basilar membrane level, ascribed together with the already reported loss in sensitivity and tuning, to the loss of cochlear amplification.

In summary, averaged as well as individual DPgrams revealed three conspicuous features at the strial/EP abnormality stage: first, and quite surprisingly, the initial transition between low- and high-frequency intervals and due to OHC loss/damage remained present, now separating intervals with large versus very small residual DPOAEs at high primary levels. Second, this transition remained as steep as that initially existing between intervals with healthy versus damaged OHCs. Last, the two transitions before and after strial/EP abnormality were found almost at the same frequency, except that a slight downward shift, of 1.1 kHz on average, was apparent in the presence of strial/EP abnormality.

IV. DISCUSSION

At the initial stage of the experiments, cochlear impairments concerned a restricted, sharply defined frequency interval with either absent or damaged OHCs. In CD1 mice with spontaneously occurring high-frequency deafness, basal OHCs were missing after having degenerated (Shone *et al.*, 1991; Le Calvez *et al.*, 1998). Acute damage induced by tonal overexposure was less drastic. The exposure characteristics (short durations and relatively low levels) ensured that damage was restricted to OHCs, and more precisely those tuned to about half an octave above exposure frequency (Cody and Johnstone, 1981). As a result, the rootlets of their stereocilia bundles are weakened (Liberman and Dodds, 1987) and mechanotransduction channels are temporarily impaired as well (Patuzzi, 1998). Excitotoxic effects affecting inner hair cells and auditory-nerve dendrites have been described only at higher levels (Puel *et al.*, 1998). Although genetically defective mice had no OHC in the basal part of the cochlea, whereas overexposed ones only had temporary impairment of their OHC stereocilia, their DPOAE characteristics turned out to be similar.

At the OHC loss/damage stage, threshold elevation and concomitant reduction of neural tuning were readily ascribed to local disruption of OHC-based feedback loop. If present, residual DPOAEs obtained at the highest primary levels did not exceed -10 dB SPL, i.e., they were about 40 dB below normal ones. It confirms that even when produced by high-level stimuli, DPOAEs correctly located the regions with versus without intact OHCs. The quasi-absence of DPOAEs induced by high stimulus levels at high frequencies suggests that they depend on the very structures impaired by genetic defect or sound overexposure (Avan *et al.*, 2003) thus points to OHC stereocilia as their source.

At the next stage with EP/strial abnormality, the impairment of cochlear function that was previously restricted to high frequency regions extended to lower frequencies. Cochlear ischemia due to heart failure and furosemide administration differ by their time course, exact mechanism and reversibility, yet they both induce a drop of EP in relation to strial dysfunction and their effects on DPOAEs were found to be similar. The consequences on cochlear gain, tuning [when it could be assessed, also see Ruggero and Rich (1991)] and compressive nonlinearity (as indirectly derived from the growth rate of residual DPOAEs) were concomitant and were those expected from disruption of the OHC-based cochlear feedback. Note that the overall functional effects of the OHC loss/damage (restricted to high frequency regions) and EP/strial abnormality stage (affecting all frequencies) concur with the current theory of cochlear micromechanics (Patuzzi, 1996) in that they do *not* depend on the reason why OHC function is impaired: whether due to lack of OHCs, impairment of mechanotransduction channels or failure of the EP ''battery'' that drives bidirectional transduction, cochlear gain, tuning and compression are altogether affected as a result of Gold's regenerative feedback loop (Gold, 1948) being disrupted.

Thus, at the EP/strial abnormality stage and as far as auditory thresholds and tuning were concerned, no transition and thus no contrast was expected, and none was found, between the high-frequency region with damaged OHCs and the low-frequency region with intact OHCs. In sharp contrast, DPOAEs produced by high-level stimuli distinguished between OHC loss/damage and EP/strial abnormality: they remained present where and only where OHCs were intact even though unable to function for lack of normal EP.

The finding that in spite of a strial/EP abnormality, DPOAEs could still be detected is not a novel one. Numerous publications have reported that it happens at stimulus levels$\geqslant$70 dB SPL, notably after ischemia or furosemide administration (Mills and Rubel, 1994; Mom *et al.*, 1997,

2001). That they cannot be produced by other structures than OHCs is already ruled out by the outcome of the basal OHC loss/damage stage: there were no DPOAEs at high frequencies. It has been argued that high-level DPOAEs might still be generated by OHCs but come from other places than those usually tuned to f2, and particularly the basal cochlea (Mills, 1997). However, this model cannot explain here why the residual DPOAEs produced by high-level stimuli after strial/EP abnormality came, within about 1 kHz from the same well-defined frequency range as that of DPOAEs from low-level stimuli when the corresponding part of the cochlea was still ''active'' with structurally intact OHCs. This observation precludes any significant involvement of different places in the generation of DPOAEs at the strial/EP abnormality stage versus ''active'' DPOAEs at the OHC loss/damage stage, all the more since, from the beginning, the basal cochlea had already lost most of its ability to produce DPOAEs even at high stimulus levels. Conversely, the results line up with Lukashkin *et al.*'s model (2002) assuming that only one OHC-related source produces DPOAEs irrespective of stimulus level. The most parsimonious explanation is that the same nonlinear structures were involved in the generation of DPOAEs all the time and that the only serious consequence of the cochlear amplifier becoming ineffective at the strial/EP abnormality stage was that the input level to the nonlinear system had to be increased in the outer ear canal to make up for the loss of gain at the cellular level. Once the nonlinear system received enough input, its ability to distort was restored (Avan *et al.*, 2003).

Even in this framework, the hardly changed sharpness of the boundary between regions with and without intact OHCs came as a surprise, because at the strial/EP abnormality stage, basilar membrane motion could not exhibit sharp tuning. This has been shown in numerous reports of basilar membrane or auditory nerve unit responses (Evans and Klinke, 1982; Sewell, 1984; Ruggero and Rich, 1991; Ruggero, 1992), and in the present work, it was inferred from the decreased Q10dBs of ABR-based STCs. Actually, even at the beginning, the sharpness of DPgram transitions did not show any change with increasing stimulus level [e.g., Figs. 1(b)–(d)], and this is not less surprising because the excitation pattern along the basilar membrane is known to spread more and more toward the cochlear base at increasing SPL (Ruggero, 1992; Nilsen and Russell, 2000). This should have made it necessary to use a special rationale to compare DPgrams at high versus low primary-tone levels, and yet simple visual comparisons were enough to bring forward their close similarities [Figs. 1(b)–(d), 3–5].

The presence of DPOAEs coming from places with well-defined boundaries, inside which OHCs are intact although not efficient enough to contribute to cochlear tuning, suggests that the elements producing DPOAEs must have some tuning of their own. The relation of DPOAE frequency specificity to tuning is the following. The extent of the area where vibrations at frequencies f1 and f2 overlap determines where a DPOAE comes from, as it has to be produced where maximum interaction at the level of nonlinear elements takes place between primary tones. One expects this extent to be normally limited by cochlear tuning mechanisms, because a few hundred μm away from the peak of basilar membrane motion in response to a tone, vibration amplitudes rapidly decrease (Ren, 2002). When tuning loses efficiency, excitation patterns get broader, so that places tuned to more than one octave above or below the frequency of interest respond significantly (e.g., Kim *et al.*, 1991). Then, considering a cochlea with no OHCs basalward to some frequency limit F, primary frequencies$>$F could stimulate OHCs normally responding to frequencies$<$F, so that DPOAEs on the DPgram should extend beyond F. This was not the case here (Fig. 5). A possible explanation would be that the DPOAE-generating elements have a resonance of their own. As already suggested, OHC stereocilia and/or their mechano-electrical transduction channels might be good candidates for being the actual sources of DPOAEs: the nonlinearity of stereocilia bundle mechanical properties and notably their stiffness (Strelioff *et al.*, 1985) is patent, as well as that of their transduction function (Kros *et al.*, 1992), with the resulting ability to generate intermodulation (Jaramillo *et al.*, 1993). It may be significant that a mechanical model of hair-cells built from *in vitro* data has shown that the OHC-tectorial membrane system, coupled by OHC stereocilia, exhibits a resonance frequency that increases from 1.2 kHz at the apex to 22 kHz at the base and therefore matches the observed best frequency of the cochlear partition at the same place (Strelioff *et al.*, 1985, their Fig. 4). The mechanotransduction channels of OHCs may still exhibit residual function despite strial failure because even though the EP drops to zero, a residual voltage difference between endolymph and interior of OHCs persists for a while, large enough for a residual cochlear microphonic potential to persist (Mom *et al.*, 1997, 2001), even though too small for electromechanical feedback to be efficient.

Although, most of the time, DPOAE tuning properties closely resemble those of cochlear tuning (Martin *et al.*, 1987) so that the tuning properties of both the DPOAEs and the whole cochlea are often similarly compromised, it has already been pointed out that cochlear and DPOAE tunings do not always suffer the same degree of impairment (Martin *et al.*, 1998; Howard *et al.*, 2003). We suggest here that this is possibly because DPOAE sources have a robust mechanical tuning of their own.

V. CONCLUSION

Put together, a number of recent findings take exception to the view that DPOAEs are ''active'' signals *per se*: the persistence of DPOAEs above 65 dB SPL in mutant mice without attached tectorial membrane (Lukashkin *et al.*, 2004) and in prestin-null mice (Liberman *et al.*, 2004) despite obvious impairment to their cochlear function are actually serious blows. According to Avan *et al.* (2003) and to the present results, as long as OHCs keep having mechanically functional stereocilia around a place tuned to f1 and f2, the presence of DPOAEs from this place seems warranted at high stimulus levels, while the absence of OHC stereocilia—admittedly, this would be by far the most frequent pattern of cochlear damage, but not the only one—compromises DPOAEs regardless of stimulus level. Likewise, the presence of intermodulation distortion in isolated saccular hair cells

(Jaramillo *et al.*, 1993) leads to believe that DPOAEs are just a by-product of one stage of stereocilia function.

In accordance with these findings, we propose that DPOAEs primarily result from a passive mechanism whose chief requirement is a truism, namely, to be nonlinear and receive enough input to produce detectable distortion. That DPOAEs can nonetheless be of interest as tokens of ''activity'' can be ascribed to the huge increase in input level to the nonlinear distortion-generating elements along with cochlear amplification and feedback: this is critical for DPOAEs to be elicited by weak sounds in the ear canal of an active ear. Besides, it must be kept in mind that the passive OHC stereocilia participate in a key stage of Gold's regenerative cochlear loop and, as such, own their share of the eventual ''active'' outcome of this loop. Last, the present results also suggest that some tuning exists at this stage, even when the cochlea itself exhibits no tuning as a result of its feedback loop being not fully functional. Exploration of this tuning through the suppression properties of DPOAEs in the presence of a third, interfering tone is currently under way (Avan *et al.*, in preparation).

ACKNOWLEDGMENTS

S.C. was funded by Grant No. 200510/97-5 from CNPq in Brazil. The technical help of Christine Puchol and Sandrine Plantade is gratefully acknowledged.

Abdala, C. (**1998**). ''A development study of distortion product otoacoustic emission ($2f1-f2$) suppression in humans,'' Hear. Res. **121**, 125–138.

Allen, J. B. (**1990**). ''User manual for the Cubdis distortion product measurement system,'' AT&T Bell Labs.

Allen, J. B., and Fahey, P. F. (**1992**). ''Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane,'' J. Acoust. Soc. Am. **92**, 178–188.

Avan, P., and Bonfils, P. (**1993**). ''Frequency specificity of human distortion product otoacoustic emissions,'' Audiology **32**, 12–26.

Avan, P., Bonfils, P., Gilain, L., and Mom, T. (**2003**). ''Physiopathological significance of distortion-product otoacoustic emissions at $2f1-f2$ produced by high-versus low-level stimuli,'' J. Acoust. Soc. Am. **113**, 430–441.

Avan, P., Magnan, P., Smurzynski, J., Probst, R., and Dancer, A. (**1998**). ''Direct evidence of cubic difference tone propagation by intracochlear acoustic pressure measurements in the guinea-pig,'' Eur. J. Neurosci. **10**, 1764–1770.

Billett, T. E., Thorne, P. R., and Gavin, J. B. (**1989**). ''The nature and progression of injury in the organ of Corti during ischemia,'' Hear. Res. **41**, 189–197.

Bonfils, P., Avan, P., Landais, P., Erminy, M., and Biacabe, B. (**1997**). ''Statistical evaluation of hearing screening by distortion product otoacoustic emissions,'' Ann. Otol. Rhinol. Laryngol. **106**, 1052–1062.

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f1-f2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227**, 194–196.

Cody, A. R., and Johnstone, B. M. (**1981**). ''Acoustic trauma: single neuron basis for the 'half-octave shift,' '' J. Acoust. Soc. Am. **70**, 707–711.

Eguíluz, V. M., Ospeck, M., Choe, Y., Hudspeth, A. J., and Magnasco, M. O. (**2000**). ''Essential nonlinearities in hearing,'' Phys. Rev. Lett. **84**, 5232–5235.

Evans, E. F., and Klinke, R. (**1982**). ''The effects of intracochlear and systemic furosemide on the properties of single cochlear nerve fibers in the cat,'' J. Physiol. (London) **331**, 409–428.

Gold, T. (**1948**). ''Hearing II. The physical basis for the action of the cochlea,'' Proc. R. Soc. B. (Edinburgh) **135**, 492–498.

Hamernik, R. P., Ahroon, W. A., and Lei, S. F. (**1996**). ''The cubic distortion product otoacoustic emissions from the normal and noise-damaged chinchilla cochlea,'' J. Acoust. Soc. Am. **100**, 1003–1012.

Harding, G. W., Bohne, B. A., and Ahmad, M. (**2002**). ''DPOAE level shifts and ABR threshold shifts compared to detailed analysis of histopathological damage from noise,'' Hear. Res. **174**, 158–171.

Howard, M. A., Stagner, B. B., Foster, P. K., Lonsbury-Martin, B. L., and Martin, G. K. (**2003**). ''Suppression tuning in noise-exposed rabbits,'' J. Acoust. Soc. Am. **114**, 279–293.

Jaramillo, F., Markin, V. S., and Hudspeth, A. J. (**1993**). ''Auditory illusions and the single hair cell,'' Nature (London) **364**, 527–529.

Kemp, D. T. (**1978**). ''Stimulated acoustic emissions from within the human auditory system,'' J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**2002**). ''Exploring cochlear status with otoacoustic emissions,'' in *Otoacoustic Emissions, Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme, New York), pp. 1–47.

Kim, D. O., Parham, K., Sirianni, J. G., and Chang, S. O. (**1991**). ''Spatial response profiles of posteroventral cochlear nucleus neurons and auditory-nerve fibers in unanesthetized decerebrate cats: response to pure tones,'' J. Acoust. Soc. Am. **89**, 2804–2817.

Kros, C. J., Rusch, A., and Richardson, G. P. (**1992**). ''Mechano-electrical transducer currents in hair cells of the cultured neonatal mouse cochlea,'' Proc. R. Soc. London, Ser. B **249**, 185–193.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f1-f2$ distortion product emissions in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Le Calvez, S., Avan, P., Gilain, L., and Romand, R. (**1998**). ''CD1 hearing-impaired mice. I: Distortion product otoacoustic emission levels, cochlear function and morphology,'' Hear. Res. **120**, 37–50.

Leeuw, A. R., and Dreschler, W. A. (**1998**). ''The relation between otoacoustic emissions and the broadening of the auditory filter for higher levels,'' Hear. Res. **126**, 1–10.

Liberman, M. C., and Dodds, L. W. (**1987**). ''Acute ultrastructural changes in acoustic trauma: serial-section reconstruction of stereocilia and cuticular plates,'' Hear. Res. **26**, 45–64.

Liberman, M. C., Guinan, J. J., and Zuo, J. (**2004**). ''Distortion-product otoacoustic emissions (DPOAEs) without prestin-mediated OHC motility,'' Abstr. Assoc. Res. Otolaryngol., edited by P. Santi, Vol. 27, p. 341.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). ''One source for distortion product otoacoustic emissions generated by low- and high-level primaries,'' J. Acoust. Soc. Am. **111**, 2740–2748.

Lukashkin, A. N., Lukashkina, V. A., Legan, K., Richardson, G. P., and Russell, I. J. (**2004**). ''Role of the tectorial membrane revealed by otoacoustic emissions recorded from wild type and transgenic TectaΔENT/ΔENT mice,'' J. Neurophysiol. **91**, 163–171.

Martin, G. K., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. L. (**1998**). ''Effect of loop diuretics on the suppression tuning of distortion-product otoacoustic emission in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.

Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contours and pure-tone exposures,'' Hear. Res. **28**, 191–208.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). ''Distortion product emissions in humans. III: influence of sensorineural hearing loss,'' Ann. Otol. Rhinol. Laryngol. **99**, 30–42.

Mills, D. M. (**1997**). ''Interpretation of distortion product otoacoustic emissions measurements: I. two stimulus tones,'' J. Acoust. Soc. Am. **102**, 413–429.

Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hear. Res. **77**, 183–199.

Mom, T., Avan, P., Romand, R., and Gilain, L. (**1997**). ''Monitoring of functional changes after transient ischemia in gerbil cochlea,'' Brain Res. **751**, 20–30.

Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). ''Origin of cubic difference tones generated by high-level stimuli: effect of ischemia and auditory fatigue on the gerbil cochlea,'' J. Acoust. Soc. Am. **110**, 1477–1488.

Moore, B. C. J., and Glasberg, B. R. (**1983**). ''Suggested formulae for calculating auditory filter band-widths and excitation patterns,'' J. Acoust. Soc. Am. **74**, 750–753.

Nilsen, K. E., and Russell, I. J. (**2000**). ''The spatial and temporal represen-

tation of a tone one the guinea pig basilar membrane,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11751–11758.
Patuzzi, R. B. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.
Patuzzi, R. B. (**1998**). ''A four state kinetic model of the temporary threshold shift after loud sound based on inactivation of hair cell transduction channels,'' Hear. Res. **125**, 39–70.
Puel, J. L., Ruel, J., d'Aldin, C. Gervais, and Pujol, R. (**1998**). ''Excitotoxicity and repair of cochlear synapses after noise-trauma induced hearing loss,'' Neuro Report **22**, 2109–2114.
Ren, T. Y. (**2002**). ''Longitudinal pattern of basilar membrane vibration in the sensitive cochlea,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 17101–17106.
Rosen, S., and Stock, D. (**1992**). ''Auditory filter bandwidths as a function of level at low frequencies (125 Hz–1 kHz),'' J. Acoust. Soc. Am. **92**, 773–781.
Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.
Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.
Sewell, W. F. (**1984**). ''The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats,'' Hear. Res. **14**, 305–314.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Shone, G., Raphael, Y., and Miller, J. M. (**1991**). ''Hereditary deafness occurring in cd/1 mice,'' Hear. Res. **57**, 153–156.
Strelioff, D., Flock, A., and Minser, K. E. (**1985**). ''Role of inner and outer hair cells in mechanical frequency selectivity of the cochlea,'' Hear. Res. **18**, 169–175.

Otoacoustic emissions without somatic motility: Can stereocilia mechanics drive the mammalian cochlea?

M. C. Liberman[a)]
Eaton-Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, and the Department of Otology & Laryngology, Harvard Medical School, Boston, Massachusetts 02115

Jian Zuo
Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, Tennessee 38105

J. J. Guinan, Jr.
Eaton-Peabody Laboratory, Massachusetts Eye & Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, and the Department of Otology & Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 26 March 2004; revised 24 May 2004; accepted 3 June 2004)

Distortion product otoacoustic emissions (DPOAEs) evoked by low-level tones are a sensitive indicator of outer hair cell (OHC) function. High-level DPOAEs are less vulnerable to cochlear insult, and their dependence on the OHC function is more controversial. Here, the mechanism underlying high-level DPOAE generation is addressed using a mutant mouse line lacking prestin, the molecular motor driving OHC somatic motility, required for cochlear amplification. With prestin deletion, attenuated DPOAEs were measurable at high sound levels. DPOAE thresholds were shifted by $\sim$50 dB, matching the loss of cochlear amplifier gain measured in compound action potentials. In contrast, at high sound levels, distortion products in the cochlear microphonic (CM) of mutants were not decreased *re* wildtypes (expressed *re* CM at the primaries). Distortion products in both CM and otoacoustic emissions disappeared rapidly after death. The results show that OHC somatic motility is not necessary for the production of DPOAEs at high SPLs. They also suggest that the small, physiologically vulnerable DPOAE that remains without prestin-based motility is due directly to the mechanical nonlinearity associated with stereociliary transduction, and that this stereocilia mechanical nonlinearity is robustly coupled to the motion of the cochlear partition to the extent that it can drive the middle ear. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1775275]

PACS numbers: 43.64.Kc, 43.64.Jb, 43.64.Ld [BLM] Pages: 1649–1655

I. INTRODUCTION

When presented with two tones, a listener hears combination tones at mathematical combinations of the input frequencies, f_1 and f_2, the most prominent of which is at $2f_1$-f_2. The nonlinearity responsible for these combination tones originates in the auditory periphery: when two tones are presented, response components at $2f_1$-f_2 can be measured in auditory nerve discharge (Goldstein and Kiang, 1968), in ear-canal sound pressure (e.g., Mountain, 1980) and in basilar membrane motion (Robles *et al.*, 1997). These nonlinear responses, termed distortion product otoacoustic emissions (DPOAEs) when measured in the ear canal, are extremely labile: i.e., they are suppressed by manipulations which reduce cochlear sensitivity (e.g., Mountain, 1980; Siegel *et al.*, 1982). For this reason, they are widely used as a noninvasive assay of cochlear function (Lonsbury-Martin and Martin, 1990).

Somatic electromotility of outer hair cells (OHCs), which can be demonstrated *in vitro* (Brownell *et al.*, 1985), drives an amplification process, *in vivo*, which is required for mechanical sensitivity in the mammalian cochlea (Liberman *et al.*, 2002). Thus OHCs, and their feedback enhancement of cochlear vibration, must contribute to the amplification of DPOAEs. However, the source of the nonlinearity which produces DPOAEs is less clear. Indeed, different sources may contribute at high vs low sound levels (Mills and Rubel, 1994). Candidate sources in the cochlear transduction process include nonlinearities in (1) mechano-electric transduction by stereocilia (Corey and Hudspeth, 1979), or (2) electro-mechanical transduction by somatic motility (Santos-Sacchi, 1991). Alternatively, distortion may arise, not through transduction *per se*, but via the mechanical nonlinearities associated with the gating of transduction channels (Jaramillo *et al.*, 1993), which could be mechanically coupled to the cochlear partition such as to make the entire vibration system nonlinear.

A direct test of the source(s) of nonlinearity generating DPOAEs is available via the mouse line with targeted deletion of prestin, a member of the sulfate transporter family expressed in OHCs, and proposed as the molecular motor driving cochlear amplification (Zheng *et al.*, 2000). Prestin

[a)]Address for correspondence: M. Charles Liberman, Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, Massachusetts 02114. Electronic mail: mcl@epl.meei.harvard.edu

deletion reduces cochlear sensitivity *in vivo* by 35–60 dB, as it eliminates OHC electromotility *in vitro* (Liberman *et al.*, 2002). A recent study of prestin mutants did not focus on DPOAEs, because DPOAE thresholds were close to the level of system distortion. With an improved acoustic delivery system, we now revisit the question of DPOAE generation without prestin, and show that distortion components at $2f_1$-f_2, although attenuated, remain in both cochlear microphonic potentials and ear-canal sound pressure.

These observations are not inconsistent with the conventional view that, in the normal ear, mammalian DPOAEs originate when stereocilia-based nonlinearities in forward transduction create distortion components in OHC receptor potentials, which are then reverse-transduced into mechanical distortions via prestin-mediated somatic motility. However, the persistence of ear-canal DPOAEs in the absence of OHC somatic motility indicates that a source of mechanical nonlinearity must remain. We suggest that this source is the mechanical manifestation of the nonlinearities in forward transduction, i.e., the nonlinear mechanics of the stereocilia bundle. If true, mammalian stereocilia must be sufficiently well coupled to the motion of the cochlear partition that they can drive the middle ear to produce DPOAEs in the ear canal.

II. METHODS

A. Animals, experimental groups, anesthesia and surgery

The creation of the mouse line with targeted deletion of the prestin gene, has been described elsewhere (Liberman *et al.*, 2002). In the present study we used mice from a subsequent generation ($F3$-$F5$) of this line, propagated by matings of heterozygous carriers of the deletion. Animals were studied at 1–2 months of age. Groups of wildtype, heterozygous and homozygous littermates were studied during the same experimental session. All procedures were approved by the IACAUC of the Massachusetts Eye and Ear Infirmary.

Animals were anesthetized with ketamine (100 mg/kg i.p.) and xylazine (20 mg/kg i.p.) with booster injections (1/2 original dose) as needed. Surgical preparation involved removal of the right pinna, followed by exposure of the right bulla, which was opened with a #11 scalpel for the recording of cochlear potentials.

B. Cochlear potentials

The compound action potential (CAP) and cochlear microphonic (CM) were recorded from the right ear of each mouse via a fine, Teflon-coated silver wire on the round window membrane referred to the tongue. The response was amplified (10,000 X), filtered (100 Hz–30 kHz), and averaged with an A–D board (National Instruments, 6052) in a LabVIEW-driven data acquisition system. For CAP, thresholds were measured under computer control in response to 5-ms tone pips (0.5-ms rise–fall with a $\cos^2$ onset envelope, delivered at 20/sec). At each SPL, 32 responses were averaged (with stimulus polarity flipped on half of the presentations to remove microphonic potentials). Threshold was defined as the sound pressure required to produce a peak-to-peak neural response of 3 μV. For CM measurement, continuous tones were presented and the voltage from the round-window electrode digitized, averaged and analyzed as described for DPOAEs.

C. Distortion product otoacoustic emissions (DPOAEs)

The DPOAE at $2f_1$-f_2 was recorded with a custom acoustic assembly consisting of two electrostatic drivers (ES-1; Tucker Davis Technologies) to generate primary tones (f_1 and f_2 with $f_2/f_1=1.2$ and f_2 level 10 dB$<f_1$ level) and a Knowles miniature microphone (EK3103) to record ear-canal sound pressure via a probe tube concentrically situated within the common sound-delivery tube. The sensitivity of the probe-tube microphone (volts/dB SPL at the tip of the probe tube) was measured in a small coupler using a calibrated 1/4″ condenser microphone (Bruel and Kjaer). Stimuli were generated digitally and presented to the drivers via custom-built low-distortion attenuators. Ear-canal sound pressure was amplified and digitally sampled at 4 μs with sampling locked to the digital stimulus generation. FFTs were computed on waveform averages (25 repetitions of a 8192-point buffer) to lower the noise floor, and 5 spectra were averaged, from which the $2f_1$-f_2 DPOAE amplitude and surrounding noise floor were extracted.

III. RESULTS

A. Cochlear threshold sensitivity and loss of gain of the cochlear amplifier

A previous report on cochlear sensitivity in mice with targeted deletion of the prestin gene (Liberman *et al.*, 2002) reported threshold elevation in homozygous mutants ranging from ~40 dB at 5.6 kHz to 60 dB at 22.6 kHz. Thresholds in mice heterozygous for the deletion were surprisingly close to the wildtypes, i.e., within 3–6 dB. These estimates of gain loss due to the elimination of the putative motor protein of the cochlear amplifier were based on measures of auditory brainstem responses (ABRs), the far-field potentials extracted from needle electrodes in the scalp in response to tone pips at different test frequencies.

In the present study, the loss of gain in each mutant animal was estimated by measuring compound action potentials (CAPs) from the round window membrane in response to tone pips. These potentials represent the summed sound-evoked activity of the auditory nerve. As shown in Fig. 1, CAP results are similar to those reported with ABRs. The mean threshold difference between the most sensitive homozygous mutants [black open circles in Fig. 1(a)] and the most sensitive wildtypes [black filled circles in Fig. 1(a)] ranged from ~35 dB at 3.2 kHz to ~55 dB at 16 kHz [Fig. 1(b)]. As reported previously (Liberman *et al.*, 2002), heterozygous prestin mutants showed cochlear sensitivity very close to that of the wildtypes. The small (~3 dB), but statistically significant, decrease in sensitivity for heterozygotes *re* wildtypes reported previously using ABR measures was not replicated in the present study using CAP [although it has recently been verified in a second group of animals studied via ABR (Wu *et al.*, 2004)]. The lack of difference in the

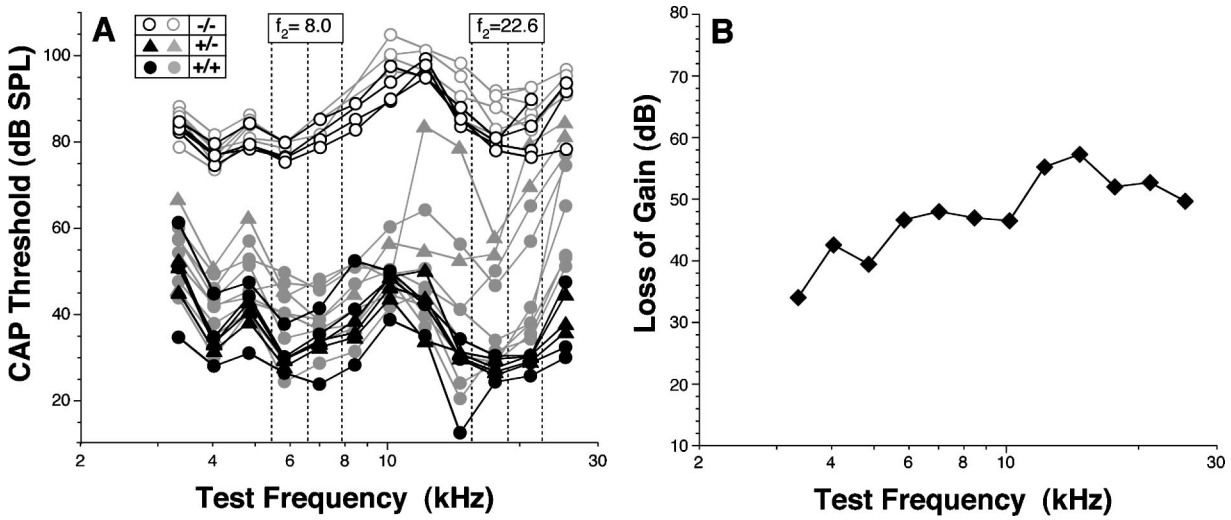


FIG. 1. Loss of cochlear amplifier gain associated with loss of prestin, as seen in cochlear potentials from wildtype and mutant mice. (a) CAP thresholds for all of the wildtype (filled circles), heterozygous (filled triangles) and homozygous (open circles) mutant mice studied. For each genotype, animals are separated into those with better thresholds (black) and worse thresholds (gray) based on the mean values at the three test frequencies closest to those evaluated in the DPOAE testing when $f_2=22.6$ ($f_1=18.8$ and $2f_1-f_2=15.06$). (b) Loss of gain computed by subtracting mean CAP thresholds for the best wildtype and heterozygote ears from the means for the best homozygote ears.

present study may be explained by the smaller sample size and the larger intra-group variance in the present study. The increased variance arises because measurement of CAP involves an invasive surgical procedure which must produce varying degrees of threshold shifts.

B. Effects of gain loss on distortion products: Ear canal vs cochlear microphonic

In the present study, distortion products were measured in every animal in two different ways: (1) as acoustic signals, by an analysis of the ear-canal sound pressure waveform, and (2) as electric signals, by analysis of the cochlear microphonic (CM) potentials recorded from the round window, which represent the summed transduction currents through cochlear hair cells (e.g., Dallos and Cheatham, 1976).

Distortion product otoacoustic emissions (DPOAEs) at $2f_1-f_2$ were measurable in all wildtypes and heterozygous mutants, and DPOAE amplitudes in heterozygotes were similar to those seen in wildtypes. Data are shown for $f_2=22.6$ in Fig. 2(a); similar results were obtained at other frequencies tested (5.6 to 45.2 kHz in half-octave steps).

DPOAEs were measurable in roughly half of the homozygous prestin mutants. The strongest responses were seen for $f_2=22.6$ and 8.0 kHz, frequencies close to the regions of maximum sensitivity in these mutants [Fig. 1(a)]. Those homozygous animals in which DPOAEs were measurable [black open circles in Fig. 2(a)] corresponded to cases with the lowest CAP thresholds [black open circles in Fig. 1(a)]. DPOAE amplitudes in the homozygotes, when present, were clearly lower than those in either wildtypes or heterozygotes. As illustrated in Fig. 2(a), the horizontal shift in the DPOAE growth functions in mutants *re* wildtypes was about 50 dB, i.e., similar to the threshold shift (i.e., ''loss of gain'') measured by the CAP.

Cochlear distortion at $2f_1-f_2$ was also measured in the round-window CM in all animals [Fig. 2(b)]. As seen in DPOAE measurements, distortion product (DP)CM amplitudes were similar in wildtypes and heterozygotes, whereas those in homozygous mutants were greatly depressed, especially at low stimulus levels. Note that homozygotes showing measurable DPOAEs (black unfilled circles) had larger DPCM than those for which the DPOAEs were below the noise floor (gray unfilled circles). Note also that the horizontal shift in DPCM growth functions for homozygotes is smaller than the loss of gain measured via CAP [Fig. 1(a)] or via DPOAEs [Fig. 2(a)].

The primary components (f_1 and f_2) in the CM response were also measured in each animal. As shown in Fig. 2(c), the CM amplitudes were, in general, lower in homozygotes than in wildtypes and heterozygotes. Among homozygotes, amplitudes of the primaries in the CM tended to be larger in the animals with measurable ear canal DPOAEs (black unfilled circles) than in those for which the DPOAEs were in the noise (gray unfilled circles). To assess possible changes in the operating point of mechano-electric transduction in the prestin null mice, CM was also measured in response to a low-frequency tone (2 kHz) at 90 dB SPL and compared to the input waveform as Lissajous patterns: no differences were seen between homozygotes and wildtypes (data not shown).

The decreased CM amplitudes in homozygotes [Figs. 2(b), 2(c)] must arise, in part, because of basal-turn OHC loss in these ears (Liberman *et al.*, 2002), and, in part, because of the loss of cochlear amplification. However, the horizontal shift in DPCM growth functions [Fig. 2(b)] is less than the loss of gain measured via CAP (Fig. 1), because the cochlear amplifier is less important to round-window CM than CAP. CM arises from hair cells distributed throughout the basal turn (Kletsky and Zwislocki, 1980), and, for many of these generators, the primaries are well below their best frequency. In contrast, the CAP threshold is dominated by responses generated near the best frequency and is thus more dependent on the cochlear amplifier.

At high stimulus levels, the DPCM in homozygotes was

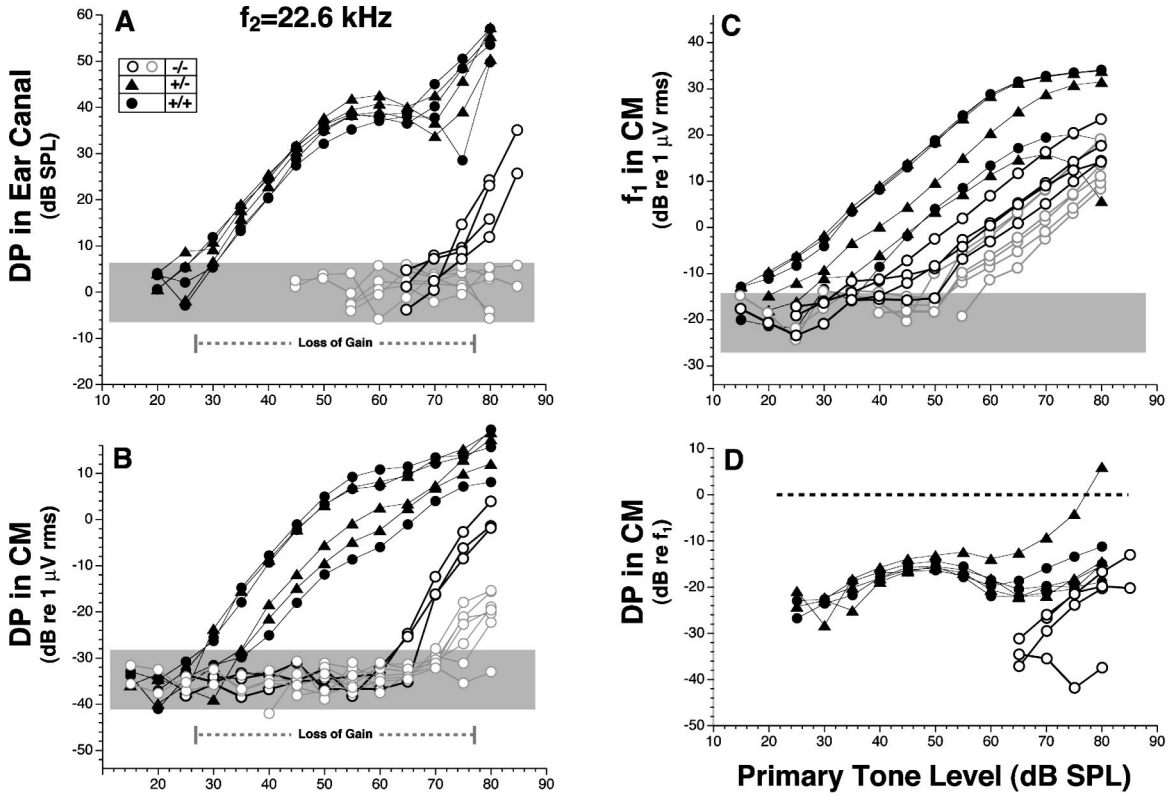


FIG. 2. Distortion products at 2 f_1-f_2. although attenuated, remain in the most sensitive homozygous mutant animals, in both the ear-canal sound pressure [panel (a)] and in the cochlear microphonic (CM) recorded from the round window [panels (b) and (d)]. The key for panel (a) applies to all panels. For all panels, data are shown for f_2=22.6 kHz from the most sensitive wildtypes and heterozygotes (as defined in Fig. 1) as well as for all the homozygotes. For homozygotes, those with the best CAP thresholds (Fig. 1) are shown in black, while those with worse thresholds are shown in gray. The gray band indicates the noise floor for these measurements. Panel (a): Amplitude vs level functions for DPOAE at 2 f_1-f_2 recorded in the ear canal. Panel (b): Amplitude vs level functions for the 2 f_1-f_2 component measured in CM. Panel (c): Amplitude vs level functions for the CM response at the lower of the two primary frequencies (f_1): Panel (d): Normalized amplitude vs level curves for the 2 f_1-f_2 component measured in the CM, referred to the appropriate $f1$ amplitude measured at the same level in the same ear.

almost as large as that in wildtypes, when viewed relative to the CM amplitudes to the primaries [Fig. 2(d)]. Since there is loss of cochlear amplification, and there may be variable degrees of hair-cell loss in the prestin null mouse (Liberman *et al.*, 2002), the absolute values of the CM and the DPCM are difficult to interpret. A more relevant measure is the ratio of DPCM to CM produced by the primaries. As shown in Fig. 2(d), the DPCM amplitudes at high stimulus levels are roughly 20 dB down from the amplitudes of the primary components in the same ears, and there are almost no differences between wildtypes and homozygotes in this measure.

C. Loss of ear-canal DPOAEs post-mortem

The stimulus range over which DPOAE measurements can be made is limited by nonlinearities in the stimulus delivery and measurement systems, which produce distortion components at 2 f_1-f_2 at high stimulus levels. Given the small size and high evoking levels for the DPOAEs measured in the homozygotes, it is critical to demonstrate that they do not arise from system distortion.

The fact that those mutant animals with the highest CAP thresholds showed no measurable ear canal distortions, even at the highest SPLs, suggests that distortion artifacts arising from acoustic delivery or signal measurement are not generating the DPOAEs in the more sensitive mutants. Further evidence for a biological origin for the attenuated DPOAEs in homozygotes is seen in Fig. 3(a), where DPOAE amplitudes evoked by 80 dB SPL primaries are tracked over time, before and after death and/or cochlear partition destruction. In all cases, DPOAE amplitudes fell precipitously within a few minutes after death, clearly demonstrating a biological origin. For those few cases in which a measurable DPOAE lingered post-mortem, subsequent cochlear partition destruction (accomplished by the insertion of a small right-angle hook through the round window without moving the acoustic assembly or disarticulating the stapes) eliminated the responses. Note that the time course for the disappearance of the high-level DPOAEs closely mirrors the time course for post-mortem decay of the cochlear microphonics [Fig. 3(b)].

IV. DISCUSSION

A. Existing ideas about mammalian DPOAE generation: High vs low-level responses

Studies in animal models show the vulnerability of DPOAEs, at low stimulus levels, to manipulations which damage OHCs, e.g., ototoxic drugs (Henley *et al.*, 1996), acoustic overstimulation (Mills, 2003) or decreased endocochlear potential (Avan *et al.*, 2003), and their lack of vulnerability to loss of inner hair cells (e.g., Liberman *et al.*, 1997). The prevailing view of DPOAE generation is that (1) non-

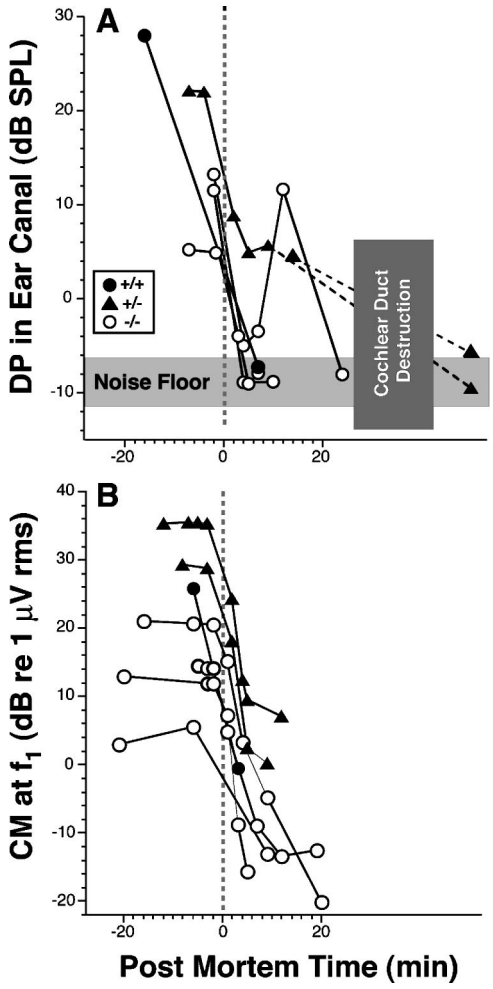


FIG. 3. Vulnerability of the high-level DPOAEs measured in the most sensitive prestin -/- ears demonstrates that these signals are generated by ''active'' processes in the animal, not artifacts of the acoustic system. Panel (a): The amplitude of the DPOAE at $2f_1$-f_2 in response to the primaries with $f_2=8.0$ kHz and 80 dB SPL. Each connected set of symbols shows data from a different case: genotypes are shown in the key. Panel (b): Shows the time course of the post-mortem decrease in CM responses to the primary tone (f_1 for $f_2=8$ kHz) for the same cases illustrated in panel (a). Death was caused by intracardiac anesthetic injection and defined as the time at which respiration ceased.

linearities in mechano-electric transduction in OHCs produce distortion components in the OHC receptor potential when two tones are present; that (2) these electrical distortions drive the electro-mechanical transducer, i.e., prestin-mediated electromotility in OHCs; and that (3) this OHC-based somatic motility at the distortion frequency is coupled to motions of the cochlear partition which propagate from the site of generation (near the cochlear region tuned to f_2) both apically to the cochlear region tuned to $2f_1$-f_2 and basally to stapes, where the energy is transmitted to tympanic membrane motion and sound generation in the ear canal at $2f_1$-f_2 (Zheng *et al.*, 1996; Lukashkin and Russell, 1998; Withnell and Yates, 1998; Kalluri and Shera, 2001; Mom *et al.*, 2001).

It has been suggested that different mechanisms underlie DPOAE generation at low vs high stimulus levels (e.g., Mills and Rubel, 1994). These speculations derived from two observations: (1) that the growth of DPOAE amplitude with primary level can show prominent nonmonotonicities (i.e., dips) associated with phase shifts, at primary SPLs of 60–70 dB SPL; (2) that a number of manipulations, including i.v. diuretics, cochlear ischemia and death, reduce DPOAE amplitudes for low-level primaries (below the dip), with little effect on DPOAE amplitudes for high-level primaries (above the dip). These observations, and the idea that the effects of OHC somatic motility on basilar membrane motion are large only at low SPLs (e.g., Ruggero and Rich, 1991), led Mills and Rubel (1994) to postulate two DPOAE sources: one ''active'' (dominating at low SPLs) and one ''passive'' (dominating at high SPLs), with opposite phase such that they cancel at about 65 dB SPL (and produce the dip in the amplitude vs level function). They speculated that the two processes might arise from different structures and/or different cochlear locations, and that only the ''active'' low-level process required OHC amplification deriving from somatic electromotility.

Recently, this two-source model has been criticized on both theoretical and empirical grounds. Frolenkov *et al.*, (Frolenkov *et al.*, 1998) perfused *in vitro* blockers of OHC electromotility, through an *in vivo* preparation and found that DPOAEs were abolished at both high and low primary levels. Avan and colleagues (Mom *et al.*, 2001) provided indirect evidence for an OHC role at both high and low SPLs, showing that the DPOAEs remaining after cochlear ischemia were highly vulnerable to acoustic overstimulation, and arguing that this temporary acoustic injury targets mainly OHC stereocilia. Russell and colleagues (Lukashkin *et al.*, 2002) showed empirically that furosemide delivery, designed to selectively reduce the low-level DPOAE, shifted the growth-function dip to *higher* input SPLs, i.e., the opposite direction predicted from a simple two-component, anti-phase model. They also showed mathematically that a single nonlinearity, appropriately configured, can reproduce the nonmonotonicities and phase shifts seen in the DPOAE data (Lukashkin and Russell, 1998). Finally, Mills (Mills, 2002) reported that the regularity of the growth function notches revealed by an exhaustive examination of the 4-parameter space defined by the frequency and level of the primaries is more easily explained by a one-component model deriving from a single nonlinearity.

B. DPOAEs without prestin: Can mammalian stereocilia drive the cochlear partition?

The prestin null mouse provides a powerful tool in studying the source(s) of DPOAEs. The prestin gene, expressed only in OHCs, codes for a key component of the molecular motor underlying cochlear amplification (Zheng *et al.*, 2000). The loss of prestin expression leads to loss of *in vitro* electromotility (Liberman *et al.*, 2002) and a 40–60 dB loss of cochlear sensitivity *in vivo* (Fig. 1). Cochlear effects of prestin elimination are arguably simpler to interpret than those of other manipulations used to reduce cochlear amplifier gain, such as furosemide injection, perfusion of electromotility blocking agents, acoustic overstimulation, ototoxic drugs, cochlear ischemia or death. These chemical and mechanical manipulations likely have multiple sites of action, and thus other effects beyond reducing the gain of the cochlear amplifier. Such additional sites of action may explain

why the electromotility blocker studied by Frolenkov *et al.*, (1998) eliminated high-level DPOAEs in their study, whereas prestin deletion in our study did not.

Several lines of evidence suggest that OHC stereocilia transduction is normal in prestin-null ears. The round-window CM data (Fig. 2) show that, in the absence of prestin, nonlinearities in OHC stereocilia transduction are still producing a distortion component at $2f_1$-f_2 in the OHC receptor currents. DPCM at $2f_1$-f_2 was $\sim$20 dB down from the primaries, at high SPLs, in both mutants and wildtypes [Fig. 2(d)]. These data argue that the $2f_1$-f_2 distortion source is normal in the absence of prestin and that the set point of the transduction channels (i.e., the open probability at rest) has not been changed by prestin deletion. The normality of forward transduction in the mutants is further corroborated by (1) the lack of difference in our CM Lissajous patterns for wildtypes vs prestin nulls and (2) CM measurements of Cheatham and colleagues who measured ''pseudo-transducer functions'' (peak positive and negative CM as a function of peak pressure, normalized to the maximum response in each animal) and found no differences between prestin knockouts and wildtypes (Cheatham *et al.*, 2004).

The persistence of high-level DPOAEs in prestin-null mice shows that OHC somatic motility is not required for their production. Furthermore, DPOAEs without prestin cannot arise according to the conventional explanation, i.e., that nonlinearities in forward transduction in OHC stereocilia produce a distortion-frequency component in receptor current which is then reverse-transduced and amplified via OHC somatic motility into distortion-frequency vibrations of the organ of Corti.

Where is the nonlinearity that produces DPOAEs in prestin-null mice? Nonlinearities in prestin-based motility have been eliminated by the targeted deletion (Liberman *et al.*, 2002). Although motion of the basilar membrane is normally nonlinear, it is likely that the basilar membrane, in isolation, is linear and that its motion nonlinearities originate elsewhere in the feedback system of the cochlear amplifier (e.g., Kanis and de Boer, 1993), which has been eliminated in these prestin-null mice. Thus, the only remaining known nonlinearities are in OHC stereocilia. Although these nonlinearities can no longer produce mechanical motion via distortions in receptor currents that are reverse transduced by OHC somatic motility, they could directly generate mechanical distortions, because (1) they are manifest in the mechanical properties of the OHC stereocilia and (2) the OHC stereocilia are mechanically coupled to the organ of Corti. Thus, a simple view of the persistent DPOAEs in the prestin-null mouse is that distortions in the organ of Corti motion, a requisite for DPOAE generation, arise from the direct coupling of the mechanical nonlinearities of OHC stereocilia bundles. Such nonlinearities have been described in hair cells from nonmammalian vertebrates which lack somatic motility (e.g., Ricci *et al.*, 2002). With tone-pair stimuli, such nonlinearities produce $2f_1$-f_2 distortion in hair bundle motion (Jaramillo *et al.*, 1993).

The finding that DPOAEs in prestin-null mice disappear rapidly after death (Fig. 3) suggests that these DPOAEs are produced by ''active'' processes in OHC stereocilia. The mechanical nonlinearity of hair bundles is strongly affected by fast adaptation due to calcium-gated closure of the transduction channel (Hudspeth *et al.*, 2000). Mammalian OHCs also display a fast calcium-dependent adaptation (Kennedy *et al.*, 2003). Thus, in the mammalian ear, bundle stiffness nonlinearities should be highly vulnerable to ischemia/death, because (1) the normal endocochlear potential declines dramatically after death (Sadanaga and Morimitsu, 1995) and (2) this $+80$ mV DC potential is a large part of the driving force for calcium entry through the transduction channel. Indeed, in the prestin null mice, the high-level DPOAEs decayed rapidly after death, with a similar post-mortem time course as the CM (Fig. 3). Thus, the high-level DPOAEs must arise from an ''active'' (i.e., energy dependent) process that depends on endocochlear potential. This strongly suggests that the DPOAE-generating processes is affected by stereocilia fast adaptation, as opposed to the stereocilia nonlinearity that remains without fast adaptation.

If stereocilia mechanical nonlinearities can directly drive cochlear mechanics and produce DPOAEs in the absence of prestin, the contribution of this direct mechanical coupling to normal DPOAE generation must be considered. In other words, has the loss of reverse transduction unmasked a phenomenon which is normally too small to play an important role, or does the direct coupling of stereocilia mechanical nonlinearities normally make a significant (or even dominant) contribution to DPOAE generation? There are several ways of looking at this issue. In prestin-null mice DPOAE ''thresholds'' were shifted by an amount matching the loss of gain extracted from cochlear neural responses [Fig. 2(a)]. At first, this seems consistent with the idea that, at threshold, the same distortion generators are operating in normal and prestin-null mice, and that the lack of a cochlear amplifier has been compensated for by raising sound level. However, the full cochlear-amplifier gain is present only at low levels, near CF; off CF, and at higher levels, the gain is less. Thus, increasing primary levels by an amount equal to the gain reduction returns BM motion to threshold levels near CF; but away from CF, it produces BM motions greater than normal at the DPOAE threshold. Thus, in prestin-null mice at the DPOAE threshold, increased BM motion away from the f_2 place might produce DPOAEs not present at threshold levels in normal mice. Another approach to this question is to compare the decrement in high-level DPOAE amplitudes associated with a lack of prestin [$\sim$35 dB *re* wildtypes: Fig. 2(a)] to that expected by the elimination of the cochlear amplifier at these high stimulus levels. At 70–80 dB SPL, cochlear amplifier gain is reduced to about 1/3 of its value (in dB) at low-levels, according to basilar membrane iso-response contours (Ruggero, 1992). According to our mouse data, normal low-level amplifier gain is 50–55 dB [Fig. 1(b)]. Thus, in prestin-null mice, the estimated loss of gain at high levels (15–20 dB) is smaller than the 35 dB reduction in DPOAE amplitude. Two other factors complicate this comparison, with opposing influences on the size of the discrepancy: (1) at high levels, DPOAE generation sites spread basally from the f_2 place, recruiting cochlear loci stimulated below the best frequency where amplifier gain is less even in normal ears (thereby increasing the size of the discrepancy); how-

ever, (2) there is basal-turn OHC loss in prestin null mice (Liberman *et al.*, 2002), thus the number of generation sites will be smaller than in wildtypes. Uncertainties as to precise values for all contributing factors make a definitive conclusion difficult. Nevertheless, these rough calculations suggest that direct coupling of stereocilia mechanical nonlinearities into motion of the cochlear partition could, with prestin-based mechanical amplification, produce DPOAEs within 10–20 dB of those seen in normal ears.

In summary, the presence of DPOAEs in prestin-null mice indicates that the only remaining nonlinearity, the stereocilia mechanical nonlinearity, can drive the motion of the cochlear partition and that this phenomenon needs to be explicitly considered in models of cochlear mechanics. The results further suggest that, although prestin is required for cochlear amplification in mammals (Liberman *et al.*, 2002), ''active processes'' in the stereocilia bundle could also contribute, as seen for hair cell systems of nonmammalian vertebrates (Jaramillo *et al.*, 1993).

ACKNOWLEDGMENTS

We thank J. Gao for assistance in genotyping. Research was supported by RO1 grants from the NIDCD: DC00188 to MCL; DC00235 to JJG; DC04761, DC05168 and CA21765 to JZ; and by the American Lebanese Syrian Associated Charities (ALSAC).

Avan, P., Bonfils, P., Gilain, L., and Mom, T. (**2003**). ''Physiopathological significance of distortion-product otoacoustic emissions at 2 $f1$-$f2$ produced by high- versus low-level stimuli,'' J. Acoust. Soc. Am. **113**, 430–441.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). ''Evoked mechanical responses of isolated cochlear outer hair cells,'' Science **227**, 194–196.

Cheatham, M. A., Huynh, K., Edge, R., Gao, J., Zuo, J., and Dallos, P. (**2004**). ''Assessment of hair cell transducer function in prestin knockout mice,'' *Abstracts of the XXVIIth ARO Midwinter Meeting*, Vol. 27, p. 270.

Corey, D. P., and Hudspeth, A. J. (**1979**). ''Ionic basis of the receptor potential in a vertebrate hair cell,'' Nature (London) **281**, 675–677.

Dallos, P., and Cheatham, M. A. (**1976**). ''Production of cochlear potentials by inner and outer hair cells,'' J. Acoust. Soc. Am. **60**, 510–512.

Frolenkov, G. I., Belyantseva, I. A., Kurc, M., Mastroianni, M. A., and Kachar, B. (**1998**). ''Cochlear outer hair cell electromotility can provide force for both low and high intensity distortion product otoacoustic emissions,'' Hear. Res. **126**, 67–74.

Goldstein, J. L., and Kiang, N. Y. S. (**1968**). ''Neural correlates of the aural combination tone 2 $f1$-$f2$,'' Proceedings of the IEEE **56**, 981–992.

Henley, C. M., Weatherly, R. A., Martin, G. K., and Lonsbury-Martin, B. (**1996**). ''Sensitive developmental periods for kanamycin ototoxic effects on distortion-product otoacoustic emissions,'' Hear. Res. **98**, 93–103.

Hudspeth, A. J., Choe, Y., Mehta, A. D., and Martin, P. (**2000**). ''Putting ion channels to work: Mechanoelectrical transduction, adaptation, and amplification by hair cells,'' Proc. Natl. Acad. Sci. U.S.A. **97**, 11765–11772.

Jaramillo, F., Markin, V. S., and Hudspeth, A. J. (**1993**). ''Auditory illusions and the single hair cell,'' Nature (London) **364**, 527–529.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kanis, L. J., and de Boer, E. (**1993**). ''Self-suppression in a locally active nonlinear model of the cochlea: a quasilinear approach,'' J. Acoust. Soc. Am. **94**, 3199–3206.

Kennedy, H. J., Evans, M. G., Crawford, A. C., and Fettiplace, R. (**2003**). ''Fast adaptation of mechanoelectrical transducer channels in mammalian cochlear hair cells,'' Nat. Neurosci. **6**, 832–836.

Kletsky, E. J., and Zwislocki, J. J. (**1980**). ''CM tuning can be compatible with sharply tuned receptor potentials,'' Hear. Res. **2**, 549–557.

Liberman, M. C., Chesney, C. P., and Kujawa, S. G. (**1997**). ''Effects of selective inner hair cell loss on DPOAE and CAP in carboplatin-treated chinchillas,'' Aud. Neurosci. **3**, 255–268.

Liberman, M. C., Gao, J., He, D. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **419**, 300–304.

Lonsbury-Martin, B. L., and Martin, G. K. (**1990**). ''The clinical utility of distortion-product otoacoustic emissions,'' Ear Hear. **11**, 144–154.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). ''One source for distortion product otoacoustic emissions generated by low- and high-level primaries,'' J. Acoust. Soc. Am. **111**, 2740–2748.

Lukashkin, A. N., and Russell, I. J. (**1998**). ''A descriptive model of the receptor potential nonlinearities generated by the hair cell mechanoelectrical transducer,'' J. Acoust. Soc. Am. **103**, 973–980.

Mills, D. M. (**2002**). ''Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response,'' J. Acoust. Soc. Am. **112**, 1545–1560.

Mills, D. M. (**2003**). ''Differential responses to acoustic damage and furosemide in auditory brainstem and otoacoustic emission measures,'' J. Acoust. Soc. Am. **113**, 914–924.

Mills, D. M., and Rubel, E. W. (**1994**). ''Variation of distortion product otoacoustic emissions with furosemide injection,'' Hear. Res. **77**, 183–199.

Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). ''Origin of cubic difference tones generated by high-intensity stimuli: effect of ischemia and auditory fatigue on the gerbil cochlea,'' J. Acoust. Soc. Am. **110**, 1477–1488.

Mountain, D. C. (**1980**). ''Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics,'' Science **210**, 71–72.

Ricci, A. J., Crawford, A. C., and Fettiplace, R. (**2002**). ''Mechanisms of active hair bundle motion in auditory hair cells,'' J. Neurosci. **22**, 44–52.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). ''Two-tone distortion on the basilar membrane of the chinchilla cochlea,'' J. Neurophysiol. **77**, 2385–2399.

Ruggero, M. A. (**1992**). ''Responses to sound of the basilar membrane of the mammalian cochlea,'' Curr. Opin. Neurobiol. **2**, 449–456.

Ruggero, M. A., and Rich, N. C. (**1991**). ''Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane,'' J. Neurosci. **11**, 1057–1067.

Sadanaga, M., and Morimitsu, T. (**1995**). ''Development of endocochlear potential and its negative component in mouse cochlea,'' Hear. Res. **89**, 155–161.

Santos-Sacchi, J. (**1991**). ''Reversible inhibition of voltage-dependent outer hair cell motility and capacitance,'' J. Neurosci. **11**, 3096–3110.

Siegel, J. H., Kim, D. O., and Molnar, C. E. (**1982**). ''Effects of altering organ of Corti on cochlear distortion products f_2-f_1 and 2 f_1-f_2,'' J. Neurophysiol. **47**, 303–328.

Withnell, R. H., and Yates, G. K. (**1998**). ''Onset of basilar membrane nonlinearity reflected in cubic distortion tone input–output functions,'' Hear. Res. **123**, 87–96.

Wu, X., Gao, J., Guo, Y. and Zuo, J. (**2004**). ''Hearing threshold elevation precedes hair-cell loss in prestin knockout mice,'' Brain Res. Mol. Brain Res. **126**, 30–37.

Zheng, J., Shen, W., He, D. Z., Long, K. B., Madison, L. D., and Dallos, P. (**2000**). ''Prestin is the motor protein of cochlear outer hair cells,'' Nature (London) **405**, 149–155.

Zheng, X. Y., Wang, J., Salvi, R. J., and Henderson, D. (**1996**). ''Effects of kainic acid on the cochlear potentials and distortion product otoacoustic emissions in chinchilla,'' Hear. Res. **95**, 161–167.

Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration

Yonatan I. Fishman[a)]
Department of Neurology, Albert Einstein College of Medicine, Kennedy Center, Room 322, 1300 Morris Park Avenue, Bronx, New York 10461

Joseph C. Arezzo and Mitchell Steinschneider
Department of Neuroscience, Albert Einstein College of Medicine, Kennedy Center, Room 322, 1300 Morris Park Avenue, Bronx, New York 10461

(Received 4 November 2003; revised 11 June 2004; accepted 11 June 2004)

Auditory stream segregation refers to the organization of sequential sounds into ''perceptual streams'' reflecting individual environmental sound sources. In the present study, sequences of alternating high and low tones, ''...ABAB...,'' similar to those used in psychoacoustic experiments on stream segregation, were presented to awake monkeys while neural activity was recorded in primary auditory cortex (A1). Tone frequency separation (ΔF), tone presentation rate (PR), and tone duration (TD) were systematically varied to examine whether neural responses correlate with effects of these variables on perceptual stream segregation. ''A'' tones were fixed at the best frequency of the recording site, while ''B'' tones were displaced in frequency from ''A'' tones by an amount $=\Delta$F. As PR increased, ''B'' tone responses decreased in amplitude to a greater extent than ''A'' tone responses, yielding neural response patterns dominated by ''A'' tone responses occurring at half the alternation rate. Increasing TD facilitated the differential attenuation of ''B'' tone responses. These findings parallel psychoacoustic data and suggest a physiological model of stream segregation whereby increasing ΔF, PR, or TD enhances spatial differentiation of ''A'' tone and ''B'' tone responses along the tonotopic map in A1. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1778903]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.64.Bt [SEM] Pages: 1656–1670

I. INTRODUCTION

Auditory scene analysis refers to the processes by which components of environmental sounds are perceptually grouped or segregated to form perceptual representations of sound sources (Bregman, 1990). Stream segregation is a fundamental feature of auditory scene analysis and denotes how interleaved sequences of sounds generated by multiple sources are decomposed into distinct perceptual streams corresponding to individual sound sources in the environment. Stream segregation can be experimentally demonstrated by listening to a sequence of high and low frequency tones in an alternating pattern, ''...ABAB....'' When frequency separation (ΔF) between the tones is small (generally $<10\%$) or presentation (alternation) rate (PR) is slow (generally <10 Hz), a single coherent alternating sequence is perceived at a tempo corresponding to the PR. At large ΔFs or rapid PRs, the sequence perceptually segregates into two auditory streams, one corresponding to ''A'' tones and the other corresponding to ''B'' tones, in which the tones are perceived at a tempo corresponding to half the PR [i.e., the ''trill threshold,'' (Miller and Heise, 1950; Bregman and Campbell, 1971; van Noorden, 1975; Anstis and Saida, 1985; Bregman, 1990)]. Stream segregation is facilitated by decreasing the interstimulus interval, i.e., the silent interval between the end of one tone and the start of the next (Bregman *et al.*, 2000). Interstimulus interval can be decreased either by increasing PR, or by increasing tone duration (TD), while holding PR constant (Beauvois, 1998; Bregman *et al.*, 2000). In between the ''coherent'' and ''segregated'' perceptual regions lies an ''ambiguous'' region, wherein alternating sequences can be perceived as either coherent or segregated, depending upon attentional set and prior expectations (McAdams and Bregman, 1979; Carlyon *et al.*, 2001; Bey and McAdams, 2002). It is important to emphasize, however, that while attention and learning are capable of strongly influencing stream segregation, they are not required for generating the basic phenomenon of stream segregation, particularly when ΔF is large and PR is rapid (McAdams and Bregman, 1979; Bregman, 1990). Indeed, when ΔF is larger than a critical value, called the temporal coherence boundary (van Noorden, 1975), two streams are always heard, despite subjects' efforts to maintain a coherent percept. Conversely, when ΔF is below a critical value, called the fission boundary (van Noorden, 1975), a single stream is always heard, regardless of subjects' efforts to hear two streams. A preattentive or ''obligatory'' component of stream segregation is further supported by noninvasive event related potential studies evaluating physiological indices of automatic change detection (the ''mismatch negativity'') for stimuli within perceptual streams (Sussman *et al.*, 1999; Nager *et al.*, 2003).

Several lines of evidence suggest the involvement of auditory cortex in stream segregation. Lesions of auditory

a)Electronic mail: yfishman@aecom.yu.edu

 © 2004 Acoustical Society of America

cortex in animals and humans are associated with deficits in processing auditory sequences (Jerison and Neff, 1953; Cowey and Weiskrantz, 1976; Colombo *et al.*, 1996; Kelly *et al.*, 1996; Liegeois-Chauvel *et al.*, 1998). Physiological studies demonstrate a capacity for auditory cortex to encode the temporal pattern of acoustic events by highly synchronized responses locked to the onset of sound components (Creutzfeldt *et al.*, 1980; Phillips, 1993; Eggermont, 1994; Steinschneider *et al.*, 1994; Phillips, 1995; Steinschneider *et al.*, 1998, 1999; Eggermont, 2000a, b; Heil, 2001; Steinschneider *et al.*, 2003). Potential neural correlates of stream segregation have been identified in primary auditory cortex (A1) of macaque monkeys (Fishman *et al.*, 2001a). These findings have been corroborated by other laboratories using similar stimuli [bat auditory cortex: Kanwal *et al.* (2003); monkey auditory cortex: Micheyl *et al.* (2003)]. In the study by Fishman *et al.* (2001a), monkeys were presented with alternating tone sequences (''..ABAB...''), while neural ensemble activity was recorded from lamina IV and lower lamina III of A1. ''A'' tones were fixed at the best frequency (BF) of the recording site, i.e., the tone frequency eliciting the largest neuronal response, whereas ''B'' tones were separated in frequency from ''A'' tones by an amount $=\Delta F$, which ranged from 10 % to 50 % of the ''A'' tone frequency. At slow PRs (e.g., 5 Hz), ''B'' tones were capable of eliciting a response at the ''A'' tone site. However, at rapid PRs (e.g., 20 and 40 Hz), neural responses to ''B'' tones at the ''A'' tone site were differentially suppressed, resulting in a neural pattern consisting predominantly of ''A'' tone responses occurring at half the sequence PR, analogous to what occurs perceptually when the ''A'' tones form a separate stream from the ''B'' tones. In other words, under stimulus conditions favoring the perception of a single auditory stream (''coherence''), neurons responded to both BF and non-BF tones, whereas under stimulus conditions favoring perception of two separate auditory streams (''segregation''), neurons responded only to BF tones. Fishman *et al.* (2001a) hypothesized that differential suppression of non-BF tone responses by preceding BF tones would increase the spatial differentiation of neural activity evoked by ''A'' tones and ''B'' tones along the tonotopic map in A1, thereby facilitating the segregation of activity into independent frequency channels and subsequent attentional selection of perceptual streams.

There were several limitations of this earlier study (Fishman *et al.*, 2001a). First, only one ΔF was investigated at each recording site. Second, alternating tone sequences were presented in blocks, such that there was an extended period (approximately 500 ms) of silence between each sequence presentation, with each sequence block lasting 490 ms. Continuous, uninterrupted stimulus sequences are preferable, as they are more comparable to stimulus designs used in human psychoacoustic studies. Finally, the earlier study did not examine the effects of tone duration (TD), an important factor influencing stream segregation (Beauvois, 1998; Bregman *et al.*, 2000).

The present study significantly expands upon this earlier work by examining several ΔFs at each recording site and by varying both PR and TD to determine whether neural responses evoked by alternating tone sequences in A1 parallel the effects of ΔF, PR, and TD on perceptual stream segregation in humans. Furthermore, sequences are continuous, such that during the data acquisition period, which begins several seconds after sequence presentation is initiated (see Sec. II), every ''A'' tone is preceded by a ''B'' tone and every ''B'' tone is preceded by an ''A'' tone. There is therefore no ''first tone'' of the sequence with respect to the neural responses analyzed. We define ''neural stream segregation'' as occurring when responses to non-BF ''B'' tones at the BF location of ''A'' tones in A1 are significantly diminished relative to BF ''A'' tone responses, resulting in a neural pattern dominated by BF ''A'' tone responses occurring at half the PR.

If temporal patterns of synchronous ''on'' responses in A1 are indeed potential contributors to stream segregation, we expect to confirm the following hypotheses, consistent with the psychoacoustics of stream segregation: (1) As PR increases, responses to ''B'' tones at the ''A'' tone BF site should diminish in amplitude to a greater extent than responses to ''A'' tones. (2) Increasing TD should facilitate differential attenuation of ''B'' tone responses, such that it occurs at slower PRs and at smaller ΔFs. Therefore, increases in ΔF, PR, or TD should result in a decrease in the amplitude of ''B'' tone responses relative to ''A'' tone responses.

The fact that newborn infants (Demany, 1982; McAdams and Bertoncini, 1997), birds (Hulse *et al.*, 1997; MacDougall-Shackleton *et al.*, 1998), fish (Fay, 1998, 2000), and macaque monkeys (Izumi, 2002) show behavioral evidence of stream segregation, and the fact that stream segregation appears to be, at least in part, an automatic perceptual phenomenon (van Noorden, 1975; McAdams and Bregman, 1979; Bregman, 1990; Sussman *et al.*, 1999; Nager *et al.*, 2003), support the existence of a phylogenetically conserved, possibly hard-wired, physiological mechanism underlying stream segregation, which is modulated by top-down influences such as learning and attention. These considerations justify the use of nonbehaving macaques as animal models for examining early cortical mechanisms involved in stream segregation.

II. METHODS

Two adult male macaque monkeys (*Macaca fascicularis*) were studied using previously described methods (Steinschneider *et al.*, 1992, 1994, 1998). Animals were housed in our AAALAC-accredited Animal Institute under daily supervision of laboratory and veterinary staff. All experiments were conducted in accordance with institutional and federal guidelines governing the experimental use of primates. For several weeks prior to surgery, animals were acclimated to the recording environment, while sitting in custom-fitted primate chairs.

A. Surgical procedure

Under pentobarbital anesthesia and using aseptic techniques, holes were drilled bilaterally into the dorsal skull to accommodate ''honeycomb'' matrices composed of 18-gauge stainless steel tubes glued together in parallel. Tubes served to guide electrodes towards A1 for repeated intracortical recordings performed on separate days (usually 3 days per

week). In general, several different auditory physiological studies were conducted during each recording session. Matrices were stereotaxically positioned to target A1. They were oriented at a 30° anterior-posterior angle and with a slight medial-lateral tilt in order to direct electrode penetrations perpendicular to the superior surface of the superior temporal gyrus, thereby satisfying one of the major technical requirements of one-dimensional current source density (CSD) analysis (Vaughan and Arezzo, 1988). Matrices and Plexiglas bars, used for painless head fixation during the recordings, were embedded in a pedestal of dental acrylic secured to the skull with inverted bone screws. Peri- and post-operative antibiotic and anti-inflammatory medications were administered.

Recordings began after a 2-week post-operative recovery period. Recordings were conducted in an electrically shielded, sound-attenuated chamber, with the animals awake and with arms comfortably restrained. Efforts were made to maintain animals in an alert state throughout the recordings by frequent visits of the experimenters into the recording chamber in between stimulus blocks, and by delivery of fruit juice rewards (which were not contingent on any behavior). The intracortical electroencephalogram (EEG) was continuously monitored on an oscilloscope for slow-wave activity, indicative of drowsiness, and one of the monkeys was also monitored via video camera. Despite efforts to maintain animals in an alert state, the possibility of intermittent drowsiness during the neurophysiological recordings cannot be excluded.

B. Neurophysiological recordings

Intracortical recordings were performed using linear-array multi-contact electrodes containing 14 recording contacts, evenly spaced at 150-μm intervals (Barna *et al.*, 1981). Individual contacts were constructed from 25-μm-diam stainless steel wires, each with an impedance of about 200 kΩ. An epidural stainless steel guide tube positioned over the occipital cortex served as a reference electrode. Field potentials were recorded using unity-gain headstage preamplifiers, and subsequently amplified 5000 times by differential amplifiers (*Grass, Inc.*) with a frequency response down 6 dB at 3 Hz and at 3 kHz. Signals were digitized at 2 kHz and averaged by computer (Neuroscan software and hardware, *Neurosoft, Inc.*) to yield auditory evoked potentials. Multi-unit activity (MUA) based on high-pass filtered data provided an online measure of summed action potential activity of neuronal aggregates within a sphere of about 100 μm in diameter surrounding each recording contact (Vaughan and Arezzo, 1988; see Fishman *et al.* (2001b) for technical details regarding derivation of MUA). One-dimensional CSD analyses characterized the laminar pattern of net current sources and sinks within A1 neural populations generating the AEPs [Freeman and Nicholson, 1975; see Steinschneider *et al.* (1998) and Fishman *et al.* (2001a) for technical details regarding derivation of CSD].

Unprocessed data were simultaneously stored (6-kHz sample rate) on a 16-channel digital tape recorder (DT-1600, *MicroData Instrument, Inc.*) for off-line analysis of multi-unit cluster activity and construction of peri-stimulus time histograms (PSTHs). Data were band-pass filtered (54 dB/oct; RP2 modules, *Tucker Davis Technologies, Inc.*) between 450 and 3000 Hz prior to spike analysis using Brainware software and hardware (*Tucker Davis Technologies, Inc.*; sample rate = 65 kHz, bin width = 1 ms). Spike trigger thresholds were set at ± 2.5 times the peak of the low-amplitude, high-frequency background activity. PSTHs were constructed from multi-unit spike data (30–50 stimulus presentations) obtained from two adjacent electrode channels situated within the thalamorecipient zone, lower lamina III and lamina IV, as indicated by the presence of prominent MUA and initial current sinks (see Steinschneider *et al.*, 1992, 1994; Fishman *et al.*, 2001b). Previous studies have localized the initial current sinks to thalamorecipient zone layers of A1 (Müller-Preuss and Mitzdorf, 1984; Steinschneider *et al.*, 1992; Sukov and Barth, 1998).

Electrodes were manipulated with a microdrive and positioned within A1 using click-evoked responses as a guide. Pure tone stimuli and continuous alternating tone sequences were presented when the electrode channels bracketed the inversion of early auditory evoked potential components and the largest MUA (typically occurring during the first 50 ms) was observed in the middle channels. Evoked responses to 75 presentations of the stimuli were averaged, with epoch durations of 300 ms (including a 25 ms prestimulus baseline interval) for pure tones and 350 ms for alternating tone sequences. Frequency response functions based on responses evoked by isolated 60 dB SPL pure tones were used to characterize the frequency tuning of the cortical sites. The BF of each cortical site was defined as the pure tone frequency eliciting the largest response in lamina IV and lower lamina III within the first 50 ms poststimulus onset. Therefore, the operational BF corresponded to the peak of the frequency response function.

At the end of the recording period, monkeys were deeply anesthetized with sodium pentobarbital and transcardially perfused with 10% buffered formalin. Tissue was sectioned in the coronal plane (80 μm thickness) and stained for acetylcholinesterase (Bakst and Amaral, 1984) and Nissl substance to reconstruct the electrode tracks and to identify A1 according to previously published criteria (Merzenich and Brugge, 1973; Wallace *et al.*, 1991; Morel *et al.*, 1993; Hackett *et al.*, 1998).

C. Stimuli

All stimuli were generated and delivered at a sample rate of 100 kHz by a PC-based system using RP2 modules (*Tucker Davis Technologies, Inc.*). Isolated pure tones used to generate the frequency response functions ranged from 0.2 to 17.0 kHz, were 175 ms in duration (including 10-ms linear rise/fall ramps), and were presented with a stimulus onset asynchrony of 658 ms. All stimuli were monaurally delivered at 60 dB SPL via a dynamic headphone (MDR-7502, *Sony, Inc.*) to the ear contralateral to the recorded hemisphere. Sounds were introduced to the ear through a 3-in.-long, 60-cc plastic tube attached to the headphone. Sound intensity was measured with a *Bruel and Kjaer* sound level meter (*type 2236*) positioned at the opening of the plastic tube. The

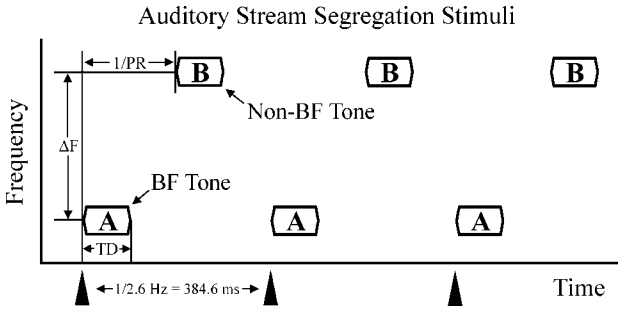


FIG. 1. Schematic representation of the stream segregation stimuli used in the study. Stimuli consist of alternating sequences of BF ''A'' tones and non-BF ''B'' tones. Independent variables include frequency separation (ΔF) between the tones (5, 10, 20, 30, and 40%), presentation rate (PR=5, 10, 20, 40 Hz), and tone duration (TD=25, 50, and 100 ms), which is the same for both ''A'' and ''B'' tones. Triggers sent to the acquisition computer (black triangles) are locked to the onset of ''A'' tones (once every 384.6 ms).

frequency response of the headphone was flattened (± 3 dB) from 0.2 to 17.0 kHz by a graphic equalizer (GE-60, *Rane, Inc.*).

Stream segregation stimuli consisted of continuous sequences of alternating high and low tones, ''...ABAB....'' The range of stimulus parameters corresponded to the range encompassing the transition between coherent and segregated percepts in humans (van Noorden, 1975; McAdams and Bregman, 1979). The frequency of ''A'' tones was fixed at or near the BF of the site, as determined by online measurements of peak MUA, in physiologically determined lamina IV and lower lamina III. The frequency of ''B'' tones was displaced in frequency from that of ''A'' tones by an amount=ΔF, expressed as a percentage of the ''A'' tone frequency. The frequency of ''B'' tones could be either above or below the BF (only one direction was tested per electrode penetration due to time constraints). ΔF, PR, and TD were varied independently. Generally, stimuli were presented in blocks, wherein one parameter (ΔF, PR, or TD) was held constant while the others were varied pseudo-randomly. Since stream segregation in humans can require several seconds to ''build up'' (Anstis and Saida, 1985), we allowed the sequences to run for at least 5 s prior to initiating data acquisition. Therefore, the present study did not examine responses to the initial tones of the sequences, thereby precluding investigation of potential ''build-up'' effects, but rather focused upon ''steady-state'' responses evoked by the continuous alternating sequences. Stimulus configuration is schematized in Fig. 1 (only the upward direction of alternation relative to the BF is shown). ΔF was 5, 10, 20, 30, or 40%; PR was 5.2, 10.4, 20.8, or 41.6 Hz (nominal PR=5, 10, 20, 40 Hz); TD was 25, 50, or 100 ms, including 5-ms linear rise/fall ramps. Although not generally examined in psychoacoustic studies, the PR=40 Hz condition (for tones of 25-ms duration) was included in order to ensure a 0-ms (nominal) interstimulus (silent) interval across all TD conditions. For example, in order for interstimulus interval to equal 0 ms, PR must be 40 Hz when TD is 25 ms, PR must be 20 Hz when TD is 50 ms, and PR must be 10 Hz when TD is 100 ms. Due to time constraints and the large number of conditions, it was not always possible to test all stimulus conditions in every electrode penetration. Duration of ''A'' tones and ''B'' tones was always equal. Since the alternating tone sequences were continuous, with no designated prestimulus ''baseline'' intervals, noninteger PRs were required in order to trigger the acquisition computer on a non-integer multiple of half the period of 60 Hz so as to minimize line noise. Data acquisition triggers coincided with the onset of ''A'' tones once every 384.6 ms (twice the period of 5.2 Hz, the slowest PR tested; see Fig. 1). Therefore, at PR =5.2 Hz, for instance, the data acquisition epoch would include a response to one ''A'' tone and to one ''B'' tone, ''A_B,'' whereas at PR=10.4 Hz, the data acquisition epoch would include a response to two ''A'' tones and to two ''B'' tones, ''A_B_A_B,'' and so on.

In a typical experiment, a given alternating tone sequence was played continuously until 75 epochs of data were collected, averaged, and stored by the acquisition computer. Stimulus parameter values were then changed and a new sequence was presented until 75 epochs of data were collected, averaged, and stored, and so on. Stimulus conditions were designed to minimize temporal overlap between ''A'' and ''B'' tones. For TD=25 ms, sequences were presented at all four PR conditions; for TD=50 ms, PRs were 5.2, 10.4, and 20.8 Hz; for TD=100 ms, PRs were 5.2 and 10.4 Hz. With TD=25 ms and PR=41.6 Hz, tones overlapped for approximately 0.5 ms; with TD=50 ms and PR=20.8 Hz, tones overlapped for approximately 1.0 ms; with TD=100 ms and PR=10.4 Hz, tones overlapped for approximately 2.0 ms. Temporal overlap of tones was uniform, so that the beginning and end of ''A'' tones and ''B'' tones overlapped by equal amounts. Alternating tones also overlapped in time under some conditions in the psychoacoustic study of Bregman *et al.* (2000), with effects on stream segregation apparently indistinguishable from those obtained under conditions where there was no temporal overlap.

D. Data analysis

Amplitudes of transient ''on'' responses evoked by ''A'' and ''B'' tones in alternating tone sequences were examined as a function of ΔF, PR, and TD. Since the alternating tone sequences were continuous, there was no prestimulus baseline to which responses could be normalized. Therefore, measures were based on trough-to-peak amplitudes of responses and hence represented perturbations of ongoing activity evoked by the continuous tone sequences. A response was characterized by an abrupt increase in the slope of the PSTH at a latency (relative to the onset of each tone) of about 10 ms, the mean onset latency of responses. All data are based on the mean amplitude of responses to ''A'' tones and to ''B'' tones within each 350-ms duration epoch. Results were confirmed via spectral analysis of PSTHs using a fast Fourier transform (FFT) algorithm (PsiPlot, *Poly Software International*) by measuring the spectral amplitude at one-half the PR as a function of ΔF, PR, and TD. This measure is based on the rationale that differential reduction in the amplitude of ''B'' tone responses would be reflected by an increase in spectral amplitude at a frequency corresponding to one-half the PR. To allow measurement at 2.6 Hz (one-half the 5.2 Hz PR), prior to computing the FFT, two identical 350-ms epochs of data under each stimulus condition

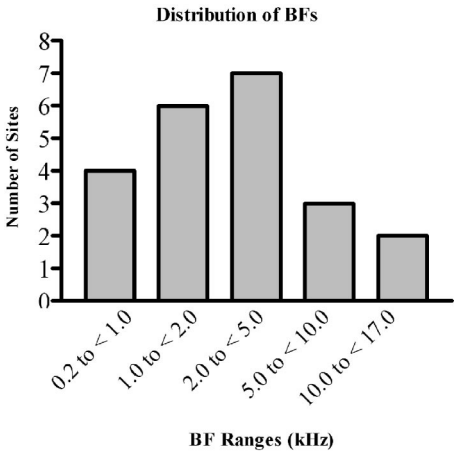


FIG. 2. Distribution of best frequencies (BFs) of the neural populations recorded at the 22 A1 sites examined in the study.

were joined together by a 35 ms block of zeros ($350+35=385$ ms = one period of 2.6 Hz). To improve frequency resolution, the appended PSTHs were padded with zeros to generate a total length of 2048 data points.

III. RESULTS

Results are based on 22 electrode penetrations into A1 of two monkeys. All results presented are based on PSTHs constructed from multi-unit cluster activity recorded in lower lamina III, as physiologically identified by the location of the largest initial current sink in the CSD profile (Steinschneider *et al.*, 1992). Corresponding MUA displayed quantitatively similar results (data not shown; available upon request). BFs of the cortical sites sampled ranged from 700 Hz to 11 kHz and formed the expected anterolateral to posteromedial topographic gradients of low- to high-frequency representation in both animals (Merzenich and Brugge, 1973; Morel *et al.*, 1993; Recanzone *et al.*, 2000). The distribution of BFs is shown in Fig. 2. Non-BF ''B'' tones were *higher* in frequency than BF ''A'' tones in 16 sites and *lower* in frequency in 6 sites. Similar results were obtained for upward and downward directions of tone alternation away from the BF (data not shown; available upon request).

Figure 3 shows PSTHs averaged across electrode penetrations when TD = 25 ms. The border of the gray shaded area above and below the mean PSTH (black line) indicates ± standard error of the mean. Response patterns evoked by alternating tone sequences at different PRs and ΔFs are arranged in separate columns and rows, respectively. Responses evoked by ''A'' tones and by ''B'' tones are labeled accordingly. Across all PR conditions, ''B'' tone responses decrease in amplitude with increasing ΔF from the BF. This result is expected based on the frequency tuning of auditory cortical neurons, according to which responses decrease in amplitude with increasing frequency distance away from the BF.

As PR increases, ''B'' tone responses decrease in amplitude to a greater extent than ''A'' tone responses. This differential attenuation of ''B'' tone responses is greater at larger ΔFs. For instance, when ΔF = 5% and PR = 5 Hz, ''A'' tones and ''B'' tones evoke similar large amplitude responses, which decrease in amplitude with increasing PR to a roughly equal extent. At PR = 40 Hz, responses to both ''A'' tones and ''B'' tones are minimal. In contrast, for larger ΔFs (e.g., 20%), ''B'' tone responses decrease in amplitude to a greater extent than ''A'' tone responses with increasing PR, such that at 20 Hz ''B'' tone responses are markedly diminished relative to ''A'' tone responses, yielding an overall neural pattern consisting predominantly of lower amplitude ''A'' tone responses occurring at half the PR (10 Hz). At PR = 40 Hz, under the ΔF = 20% condition, ''B'' tone responses are virtually absent. Thus, while at slower PRs, neuronal aggregates respond both to BF ''A'' tones and to non-BF ''B'' tones, at faster PRs, neurons respond only to BF ''A'' tones. Increasing either PR or ΔF therefore leads to a decrease in the amplitude of ''B'' tone responses relative to the amplitude of ''A'' tone responses. Dashed boxes enclose PSTHs in which ''B'' tone response amplitudes are less than −9 dB relative to ''A'' tone responses.

Differential attenuation of ''B'' tone responses occurs at slower PRs or at smaller ΔFs when TD is lengthened. This is illustrated in Figs. 4 and 5, where PSTHs averaged across electrode penetrations are shown under the 50 and 100 ms TD conditions, respectively. For instance, when TD = 25 ms (Fig. 3), under the ΔF = 10% and PR = 20 Hz conditions, ''B'' tone responses are greater than half the peak amplitude of ''A'' tone responses. In contrast, when TD = 50 ms, under the same ΔF = 10% and PR = 20 Hz conditions (Fig. 4), ''B'' tone responses are virtually absent, resulting in a neural pattern consisting predominantly of lower amplitude ''A'' tone responses occurring at half the PR (10 Hz). Under the 50 ms TD condition, ''B'' tone responses are less than −9 dB relative to ''A'' tone responses for all PSTHs when PR = 20 Hz.

An even more powerful effect of PR is demonstrated when TD = 100 ms (Fig. 5). Whereas prominent ''B'' tone responses are evoked under the TD = 25 ms condition when PR = 10 Hz and ΔF = 5%, 10%, and 20% (Fig. 3), under the TD = 100 ms condition, large amplitude ''B'' tone responses evoked when PR = 5 Hz are virtually absent when PR = 10 Hz, for all ΔFs tested. Thus, the longer the TD, the slower the PR or smaller the ΔF at which ''B'' tone responses are differentially attenuated.

PSTH results are quantified in Fig. 6. Ratios of ''B'' tone response amplitude to ''A'' tone response amplitude under each stimulus condition were computed at individual electrode penetrations and then averaged across electrode penetrations (error bars indicate ± standard error of the mean). Figure 6(a) shows mean B/A response ratios plotted as a function of PR, and with ΔF as the parameter. Data under the 25, 50, and 100 ms TD conditions are plotted in the left, middle, and right columns, respectively. Increasing PR results in a differential reduction in non-BF ''B'' tone response

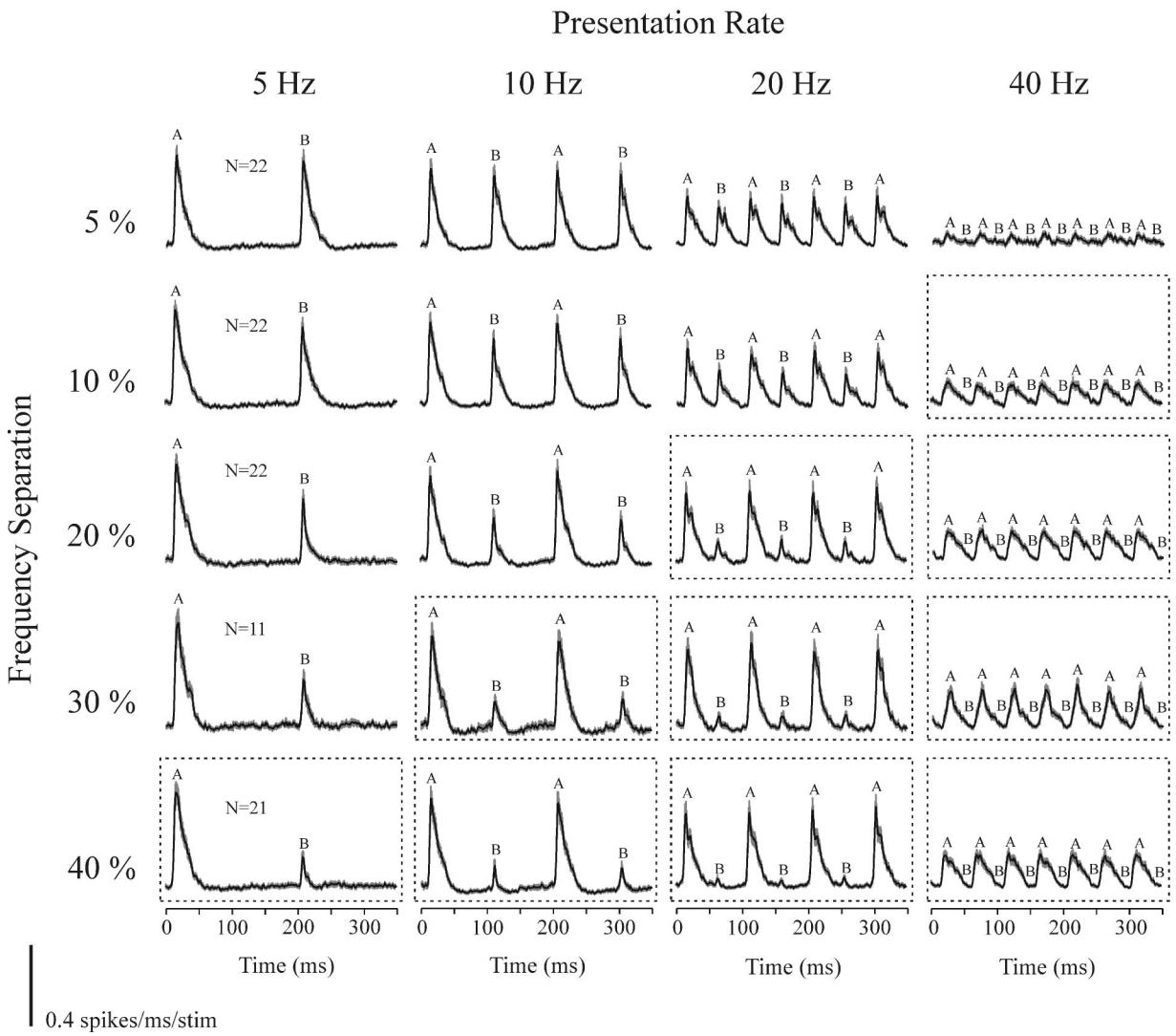


FIG. 3. Peristimulus time histograms (PSTHs) of multi-unit spike activity evoked by alternating tone sequences averaged across A1 electrode penetrations. Tone duration=25 ms. PSTHs under the different ΔF and PR conditions are arranged in rows and columns, as indicated. Responses to "A" tones and to "B" tones are labeled. The boundaries of the gray shaded region above and below each average PSTH (black line) represent ± standard error of the mean. Dashed boxes enclose PSTHs in which "B" tone response amplitudes are less than −9 dB relative to the amplitudes of "A" tone responses. Number of sites (N) contributing to each of the average PSTHs in each row is indicated in the far left column.

amplitude, reflected by a decrease in B/A response amplitude ratio. Furthermore, there is an exaggerated decrease in B/A response ratios at longer TDs.

Differential attenuation of non-BF "B" tone responses leads to an effective increase in frequency selectivity at the "A" tone BF location in A1. This is illustrated in Fig. 6(b), which shows mean B/A response ratios plotted as a function of ΔF, and with PR as the parameter. B/A response ratios are set equal to 1.00 at ΔF=0%, reflecting the equal amplitudes of responses to trains of BF "A" tones. As ΔF increases, B/A response ratios decrease, consistent with increasing frequency distance of "B" tones away from the BF. Importantly, B/A response ratios decrease to a greater extent with increasing PR, thereby reducing the effective width of frequency tuning. For instance, when ΔF=10% and TD =25 ms, mean B/A response ratio at PR=10 Hz is about 0.75. In contrast, mean B/A response ratio is about 0.55 when PR=20 Hz, and less than 0.25 when PR=40 Hz, when ΔF=10%. Moreover, given the same PR, effective frequency tuning is narrower at longer TDs (compare ratios at corresponding PRs under the 25, 50, and 100 ms TD conditions: left, middle, and right columns, respectively).

At all TDs, two-way analysis of variance tests (ANOVA, unpaired, due to different sample sizes across conditions) show statistically significant ($p<0.05$) effects of PR (TD =25 ms: $df=3$, $F=105.7$, $p<0.0001$; TD=50 ms: $df=2$, $F=76.5$, $p<0.0001$; TD=100 ms: $df=1$, $F=181.7$, $p<0.0001$) and ΔF (TD=25 ms: $df=4$, $F=91.1$, $p<0.0001$; TD=50 ms: $df=4$, $F=25.7$, $p<0.0001$; TD =100 ms: $df=4$, $F=11.1$, $p<.0001$). Effects of TD, holding PR constant, were not statistically significant under the PR=5 Hz condition (TD=25 vs. 50 vs. 100 ms: $df=2$, $F=1.36$, $p>0.05$). However, effects of TD were statistically significant under the PR=10 Hz condition (TD=25 vs. 50 vs. 100 ms: $df=2$, $F=88.42$, $p<0.0001$) and under the

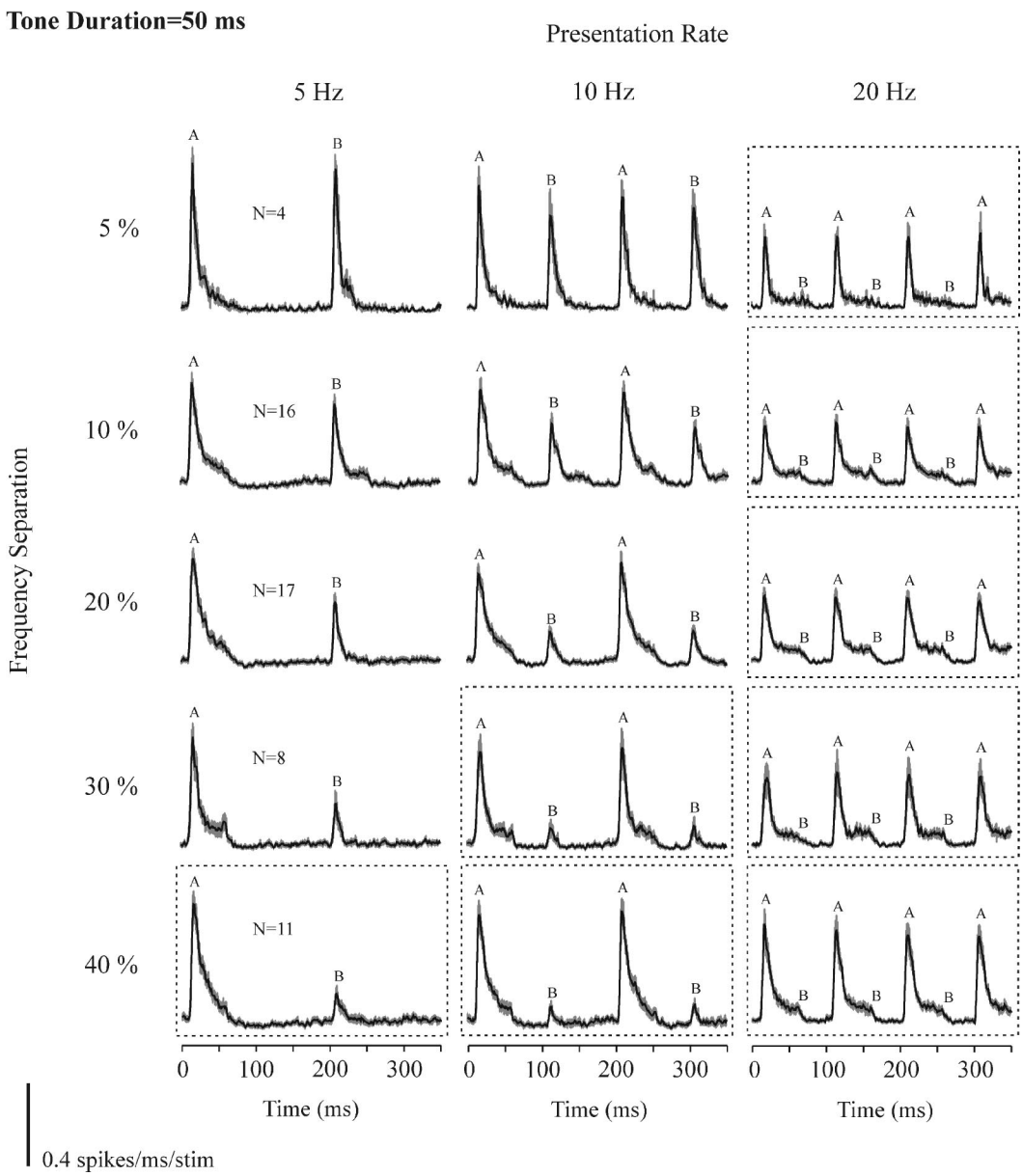


FIG. 4. Same as in Fig. 3, but with tone duration = 50 ms.

PR = 20 Hz condition (TD = 25 vs. 50 ms: $df = 1$, $F = 66.5$, $p < 0.0001$). ΔF, PR, and TD are thus largely interchangeable in their effects on B/A response ratios. Larger ΔFs at slower PRs can yield ratios as low as those obtained for smaller ΔFs at faster PRs, and longer TDs at slower PRs can yield ratios as low as those obtained for shorter TDs at faster PRs. These physiological ''trading relations'' parallel those reported in psychoacoustic studies on perceptual stream segregation (Bregman *et al.*, 2000), and accordingly suggest that the critical variable underlying ''neural stream segregation'' is interstimulus interval (= 1/PR − TD). This is supported in Fig. 6(c), wherein mean B/A response ratios are plotted as a function of interstimulus interval, and with ΔF as the parameter. Shorter interstimulus intervals yield significantly smaller B/A response ratios (unpaired ANOVA: $df = 6$, $F = 28.3$, $p < 0.0001$).

To corroborate the validity of findings based on trough-to-peak measurements of ''A'' tone and ''B'' tone responses, we also computed FFTs of averaged PSTHs (see Sec. II for details) and examined changes in the amplitude spectrum at one-half the PR as a function of PR. This procedure is based on the rationale that differential attenuation of ''B'' tone responses would be reflected by an increase in the response amplitude spectrum at one-half the PR as PR increased. In agreement with trough-to-peak measures, under the 25 ms TD condition, FFTs revealed an increase in the amplitude spectrum at one-half the PR for ΔFs > 5 Hz as PR increased (data not shown). Under the 50- and 100-ms TD conditions, an increase in spectral amplitude at one-half the PR with increasing PR was observed for all ΔFs. Moreover, keeping PR constant, spectral amplitude at one-half the PR increased with increasing TD.

The putative ''neural stream segregation'' demonstrated here occurs within a range of stimulus parameter values similar to the range within which a transition from a coherent to a segregated percept can be demonstrated in human psy-

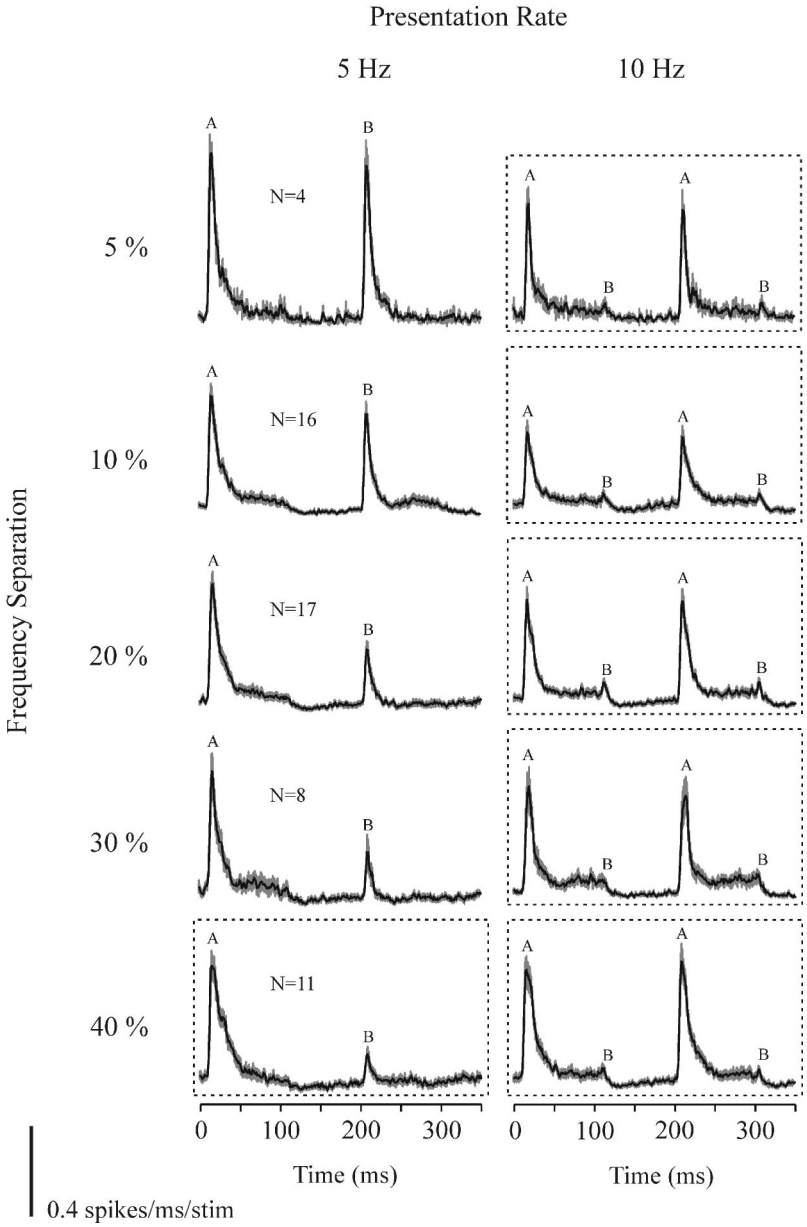


FIG. 5. Same as in Figs. 3 and 4, but with tone duration = 100 ms.

choacoustic experiments. This agreement is illustrated in Fig. 7, where B/A ratio data under the 25 ms (A) and 50 ms (B) TD conditions are superimposed on graphs delineating the perceptual boundaries of stream segregation in human listeners. The boundaries depicted are based upon psychoacoustic results obtained with alternating tones of 40-ms duration, intermediate between the 25- and 50-ms TDs used in the present study (van Noorden, 1975; McAdams and Bregman, 1979). The largest B/A ratios are generally observed under stimulus conditions lying within the ''coherent'' perceptual region, whereas the smallest B/A ratios are generally observed under stimulus conditions lying within the ''segregated'' perceptual region. The apparently discrepant lower B/A ratios lying in the ''coherent'' region of the graph in Fig. 7(b) when PR = 20 Hz, may be due to the fact that perceptual data were obtained with tones of 40-ms duration (McAdams and Bregman, 1979), whereas physiological data are obtained with tones of 50-ms duration. The lower B/A ratios are consistent with the prediction that the longer 50-ms tones would promote greater stream segregation (Bregman *et al.*, 2000).

To quantify the agreement between physiological and perceptual data, stimulus conditions leading to segregated, ambiguous, and coherent percepts in humans were assigned ranks of 1, 2, and 3, respectively, and these ranks were subsequently correlated with B/A response ratios obtained under corresponding stimulus conditions. Mean B/A ratios ($\pm$ standard deviation) under the 25- and 50-ms TD conditions are plotted against perceptual ranks in Figs. 7(c) and (d), respectively. Physiological data and perceptual ranks are significantly correlated (Spearman correlation: $r = 0.75$, $p < 0.00001$ for 25 ms TD; $r = 0.59$, $p < 0.00001$ for 50 ms TD conditions; tests are one-tailed, since the hypothesis is directional).

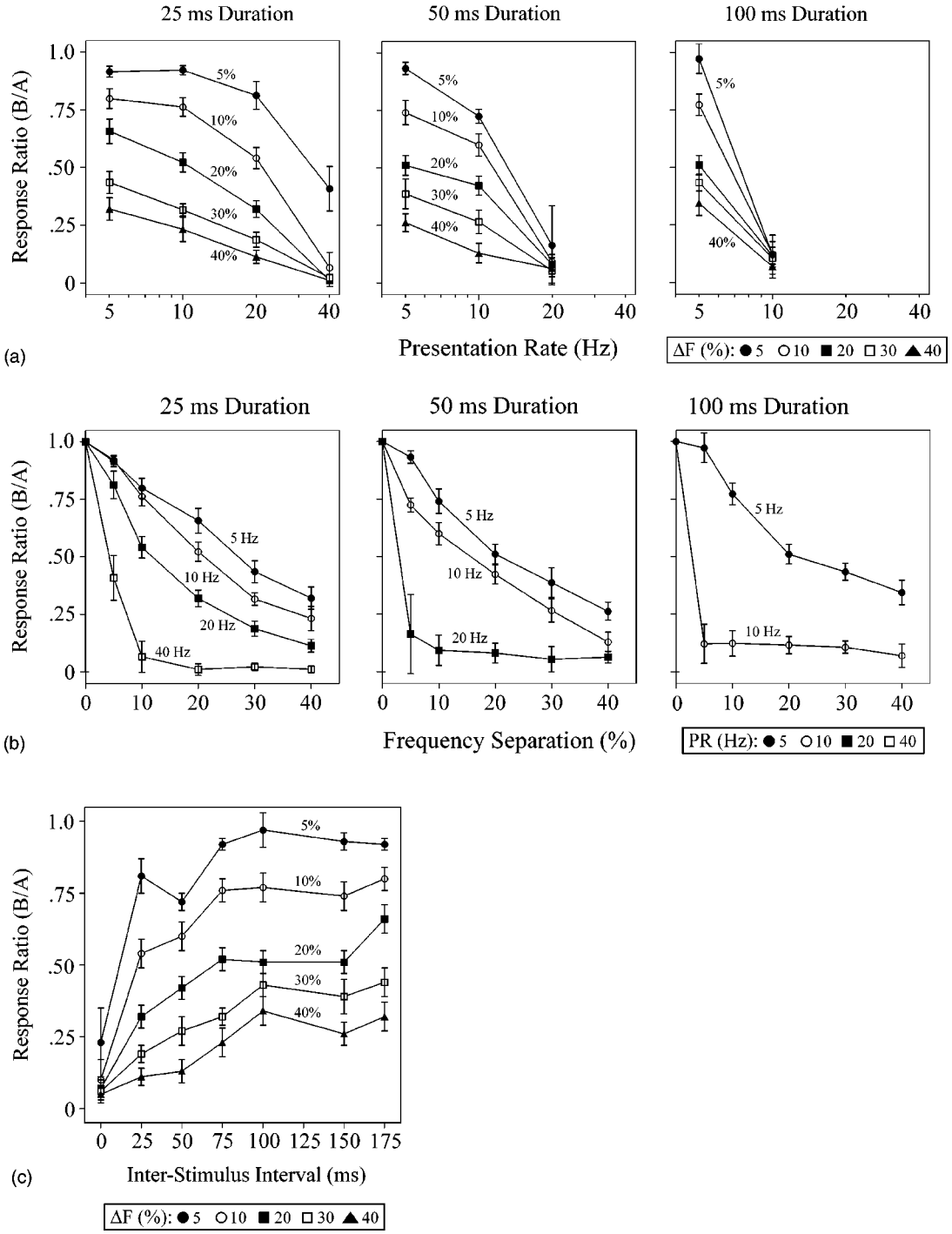


FIG. 6. (a) Mean ratio of ''B'' tone response amplitude to ''A'' tone response amplitude plotted as a function of PR and with ΔF as the parameter (error bars represent ± standard error of the mean; symbols representing different values of the parameter are denoted in the legend). Data under the 25, 50, and 100 ms TD conditions are represented in the graphs in the left, middle, and right columns, respectively. (b) Same data as in (a), but plotted as a function of ΔF and with PR as the parameter. At ΔF = 0%, mean ratios are set equal to 1.0, reflecting equal amplitude of responses to trains of BF ''A'' tones. (c) Same data as in (a) and (b), but plotted as a function of inter-stimulus interval (1/PR-TD) and with ΔF as the parameter.

IV. DISCUSSION

A. Summary of findings and relationship to psychoacoustic studies

The present study extends our earlier findings (Fishman *et al.*, 2001a) by examining the effects of ΔF, PR, and TD, three key variables influencing stream segregation in humans, on neural responses to continuous alternating tone sequences in A1. The results provide additional support for parallels between temporal response patterns of neuronal populations in A1 and perceptual stream segregation. This putative ''neural stream segregation'' is manifested by the differential reduction in the amplitude of responses to non-BF ''B'' tones, leading to a decrease in the B/A response amplitude ratio at the ''A'' tone BF location. The present data suggest that the perception of a single, coherent auditory stream is facilitated when neurons at a specific BF location

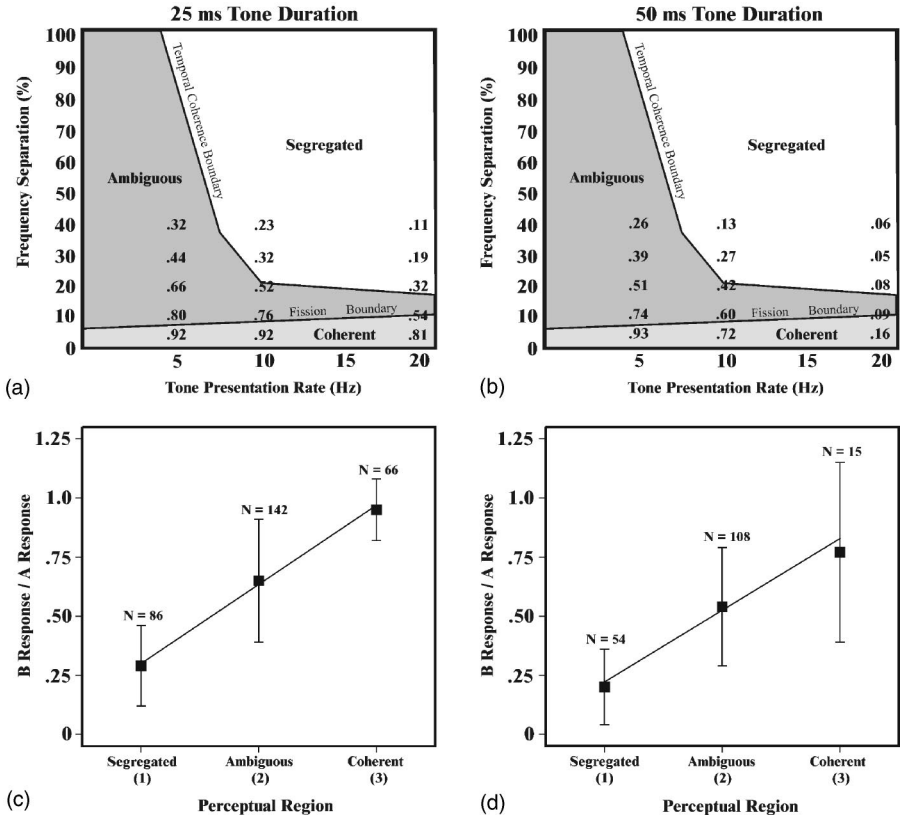


FIG. 7. B/A response amplitude ratio data under the 25 ms (a) and 50 ms (b) TD conditions superimposed on graphs delineating the perceptual boundaries of stream segregation in humans (adapted from McAdams and Bregman, 1979). The smallest B/A response ratios tend to lie in the ''segregated'' perceptual region, whereas the largest B/A response ratios tend to lie in the ''coherent'' perceptual region. Correlation between physiological and perceptual data is evaluated by assigning a rank of 1 to stimulus conditions lying in the ''segregated'' perceptual region, a rank of 2 to stimulus conditions lying in the ''ambiguous'' perceptual region, and a rank of 3 to stimulus conditions lying in the ''coherent'' perceptual region (according to McAdams and Bregman, 1979). Mean (± standard deviation) B/A response ratio under the 25 ms TD condition (c) and under the 50 ms TD condition (d) is plotted against perceptual rank. The sample size (N) contributing to each mean is indicated above the symbols. B/A response ratios are significantly correlated with perceptual ranks. Superimposed linear regression lines emphasize this relationship.

respond strongly to both BF and non-BF tones, whereas the perception of two separate auditory streams is facilitated when non-BF tone responses are markedly diminished relative to BF tone responses, resulting in an overall pattern of neural activity dominated by BF tone responses occurring at half the PR. Differential attenuation of non-BF tone responses is promoted by increases in PR and TD, while increasing ΔF leads to an overall decrease in ''B'' tone response amplitude across all conditions, as expected based upon the frequency tuning of neural populations in A1.

These physiological results are consistent with the psychoacoustic findings of Bregman *et al.* (2000), indicating that the two primary factors controlling stream segregation are ΔF and interstimulus interval (i.e., tone offset-to-onset interval), which depends upon both PR and TD. While interstimulus interval was not explicitly varied in the present study, it was varied indirectly by independently changing PR and TD, since interstimulus interval $= (1/PR) - TD$. In accord with findings of Bregman *et al.* (2000), smaller interstimulus intervals lead to lower B/A response ratios [see Fig. 6(c)], implying that differential attenuation of non-BF tone responses depends critically not on PR or TD *per se*, but rather on interstimulus interval. The present findings suggest therefore that stream segregation may be due to enhancement of frequency selectivity of A1 sites resulting from the effective sharpening of frequency tuning at the BF tonotopic location at brief interstimulus intervals. Future studies should examine effects of keeping interstimulus interval constant, while varying other stimulus variables.

Discrepancies between the present physiological results and those of human psychoacoustic experiments might be due to species differences, restriction of recordings to A1, differences in stimulus parameters, such as rise time, tone duration, and intensity, specific task demands in the human studies (see Moore and Gockel, 2002), and fluctuations in attention, which may modulate results within the ''ambiguous'' perceptual region. In any case, the fact that physiological data are not randomly distributed across perceptual boundaries supports the potential relevance of temporal response patterns of neural populations in A1 to perceptual stream segregation.

B. Relationship to other auditory physiological studies

The differential reduction in the amplitude of non-BF tone responses relative to BF tone responses observed at the higher PRs and with longer TDs may be due to several factors, including differential adaptation or differential suppression of BF and non-BF responses. These mechanisms cannot

be distinguished on the basis of the data obtained in the present study. Suppression may be due to synaptic inhibition or to after-hyperpolarization (e.g., see Eggermont, 2000b). Studies in auditory cortical slices implicate inhibitory synaptic contributions to the hyperpolarization observed following an excitatory afferent input (Cox *et al.*, 1992; Metherate and Ashe, 1994, 1995). Intracellular neurophysiological studies in cat auditory cortex reveal inhibitory postsynaptic potentials (IPSPs) following the initial stimulus-evoked depolarization [de Ribaupierre *et al.* (1972), unanesthetized cat], sustained inhibition lasting the duration of tone stimulation [Volkov and Galazyuk (1992), anesthetized cat], and reduced neuronal membrane input resistance persisting tens to hundreds of milliseconds after termination of auditory stimulation (Volkov and Galazyuk, 1992). Inhibitory processes engaged by tone onset, during tone stimulation, as well as by tone offset may therefore account for the enhanced reduction in the amplitude of ''B'' tone responses at faster PRs and for longer TDs. Increasing PR or TD, which decreases the silent interval between tones, would facilitate suppressive interactions between inhibitory phases of preceding responses and excitatory phases of subsequent responses. Inhibition triggered by tone offset might therefore account for the differential reduction in non-BF tone response amplitude with decreasing interstimulus interval [see e.g., Fig. 6(c)] observed in the present study.

The time-course and frequency dependence of inhibitory temporal interactions shaping the response patterns underlying ''neural stream segregation'' are similar to those governing physiological forward masking in A1, thus suggesting a common mechanism. Calford and Semple (1995) and Brosch and Schreiner (1997) examined the suppressive effects of masker tones of various frequencies on the response to a subsequent probe tone at the characteristic frequency (CF) of neurons in A1 of anesthetized cats. Both studies showed that forward suppression of responses to CF tones became progressively less effective as the masker frequency moved away from the CF of the neuron. Similarly, in the present study, at rapid PRs, BF ''A'' tone responses tend to become progressively less suppressed by preceding non-BF ''B'' tone responses as ΔF increases (e.g., compare mean ''A'' tone response amplitudes at $\Delta F = 10\%$ and at 40% at the fastest PRs in Figs. 3–5). This ''disinhibition'' of ''A'' tone responses occurs over a frequency range roughly corresponding to the width of auditory filters, i.e., $\Delta F = 10\% - 20\%$, in general agreement with the results of physiological and psychoacoustic forward masking studies (Moore, 1978; Calford and Semple, 1995; Brosch and Schreiner, 1997), and with values of ΔF at which alternating tone sequences perceptually segregate [i.e., the ''trill threshold'' (Miller and Heise, 1950)]. The investigations of Calford and Semple (1995) and Brosch and Schreiner (1997) also demonstrated an increase in suppression of probe responses as the onset-to-onset interval between masker and probe tones decreased. Similar findings were reported in an optical study of spatiotemporal inhibition evoked by two-tone sequences in guinea pig auditory cortex (Horikawa *et al.*, 1997). In Brosch and Schreiner's (1997) study, maximal duration of forward inhibition (in terms of stimulus onset asynchrony) ranged from 53 to 430 ms (mean: 143 ms), similar to the range observed in the present study, where reductions in overall response amplitude can be seen at PRs of 10 Hz (stimulus onset asynchrony of 100 ms) and above. By independently varying both stimulus onset asynchrony and the duration of the masker, Brosch and Schreiner (1997) also demonstrated an increase in forward masking with decreases in masker-probe offset-to-onset interval (i.e., interstimulus interval), suggesting, in agreement with our results, the importance of inhibitory processes triggered by sound offsets in addition to onsets. The present findings are also consistent with those of studies examining neural gap detection thresholds in A1 of anesthetized cats (Eggermont, 2000b). Neural responses to a trailing noise burst are attenuated with decreasing duration of the silent gap between leading and trailing noise bursts, demonstrating the importance of onset-to-onset interval. As onset-to-onset interval of leading and trailing noise bursts varied with gap duration, the significance of offset-to-onset interval *per se* was not directly evaluated in Eggermont's (2000b) study. There are important differences, however, between the stimulus conditions of the physiological forward masking studies of Calford and Semple (1995) and of Brosch and Schreiner (1997) and those of the present study. First, in the present study, response patterns reflect not only the effect of non-BF tone responses on subsequent BF tone responses, but also the effect of BF tone responses on subsequent non-BF tone responses, a condition which was not examined in these other investigations. Second, the present study examines ''steady-state'' responses evoked by continuous tone sequences, rather than by isolated tone pairs separated by long periods of silence.

Assuming that forward suppression is largely responsible for the response patterns observed in the present study, differential suppression of non-BF tone responses indicates that BF tone responses are more effective at suppressing subsequent tone responses, and are thus better ''maskers'' than non-BF tone responses. The physiological and anatomical bases of this differential suppression remain unclear. However, a possible explanation is a correlation between the magnitude of the initial excitation and subsequent inhibition, reflecting either after-hyperpolarization (e.g., see Eggermont, 2000b), or perhaps greater activation of local inhibitory interneurons within the same cortical area. Whether this inhibition is generated entirely at the cortical level, or is partly due to subcortical processes, cannot be determined with the present data. Studies have reported suppressive interactions between successive tone responses at the level of the auditory nerve (Smith, 1979; Delgutte, 1990), cochlear nucleus (Boettcher *et al.*, 1990; Shore, 1995), and inferior colliculus (Arehole *et al.*, 1987; Walton *et al.*, 1997), suggesting possible subcortical contributions (see Frisina, 2001, for review). Kanwal *et al.* (2003) designed a physiologically realistic multi-layer neural network model that was able to mimic the pattern of evoked potentials elicited by alternating tone sequences in auditory cortex of mustache bats and to replicate the ''neural stream segregation'' phenomena reported in Fishman *et al.* (2001a). The model incorporated recurrent inhibition proportional to the excitatory response evoked by tone onset, thereby suggesting that inhibition at

the cortical level, at least in principle, is sufficient to account for the results of Fishman *et al.* (2001a) and those of the present study.

C. "Neural stream segregation" in A1

In the present study, "A" tones were fixed at the BF of the recording site and "B" tone responses were differentially diminished relative to "A" tone responses at high PRs or with long TDs. It is crucial to appreciate that a similar, although inverse, response pattern will occur at the tonotopic location in A1 where "B" tones are at the BF and non-BF "A" tones are differentially diminished. At fast PRs or with long TDs, resultant response patterns at the "A" tone site consist of BF "A" tone responses occurring at half the PR, whereas response patterns at the "B" tone site consist of BF "B" tone responses occurring at half the PR. The results of the present study suggest the following account of "neural stream segregation" in A1 similar to that proposed in Fishman *et al.* (2001a): When ΔF is large, there is little spatial overlap in neural activity evoked by "A" tones and "B" tones along the tonotopic map and minimal inhibitory temporal interactions between tone responses. Consequently, alternating sequences will evoke two spatially well-differentiated peaks of activity along the tonotopic map at both slow and fast PRs, a physiological pattern corresponding to perceptual segregation. This is exemplified in Figs. 3–5 by the relatively small "B" tone responses at PR $=5$ Hz when $\Delta F=40\%$. On the other hand, when ΔF is small, there is considerable spatial overlap in neural activity evoked by "A" tones and "B" tones and pronounced mutual inhibitory interactions at fast PRs. At both slow and fast PRs, neither "A" tone responses nor "B" tone responses dominate the overall activity pattern, as both tones are near the BF at both sites and responses inhibit each other to a similar extent, a physiological pattern corresponding to perceptual coherence. This is exemplified in Fig. 3 by the similar-amplitude responses to "A" and "B" tones at all PRs when $\Delta F=5\%$ under the TD$=25$ ms condition.

At intermediate ΔFs, under the slow PR condition, there is again considerable spatial overlap in neural activity evoked by "A" tones and by "B" tones, such that each site responds vigorously to both tones, although with BF responses dominating. At fast PRs, however, responses to non-BF tones are differentially attentuated such that neurons at the "A" tone BF site still respond vigorously to "A" tones, but no longer respond to non-BF "B" tones, whereas neurons at the "B" tone BF site still respond vigorously to "B" tones, but no longer respond to non-BF "A" tones. This differential attenuation results in an effective narrowing of frequency tuning at each of the sites [as exemplified in the graphs of Fig. 6(b)], thereby enhancing the spatial segregation of "A" tone responses and "B" tone responses along the tonotopic map. Thus, we propose that the perception of two separate auditory streams, characteristic of stream segregation for alternating two-tone sequences, correlates with the presence of two spatially differentiated peaks of activity along the tonotopic map in A1, which can result either from differential suppression of non-BF tone responses or from spatial separation of the tone responses due to large ΔF.

A key feature of perceptual stream segregation is a subject's ability to focus attention either on the "A" tone stream or on the "B" tone stream, once the tones have been perceptually segregated by "automatic" physiological mechanisms (Bregman, 1990). Assuming that attention tracks locations of abrupt changes in the spatial distribution of activity along the tonotopic map, enhanced spatial differentiation of "A" tone and "B" tone responses may facilitate the allocation of attention to "A" and "B" frequency channels. Accordingly, once peaks of activity have been spatially segregated by "automatic" neural mechanisms in A1, selective attention directed by top-down mechanisms may subsequently bias processing of activity within the frequency channel corresponding to the "A" tone stream or to the "B" tone stream, either by amplifying the signal in the attended channel or by attenuating the signal in the unattended channel (e.g., Alain and Woods, 1994). General response enhancement effects of attention in human A1 have been reported (Fujiwara *et al.*, 1998; Jancke *et al.*, 1999). Accordingly, psychoacoustic studies have demonstrated modulation of sensory processing in specific frequency channels by selective attention (Dai *et al.*, 1991; Mondor and Bregman, 1994; Botte *et al.*, 1997). The fact that the present investigation was conducted under passive listening conditions suggests that "neural stream segregation" is, at least in part, an automatic or "obligatory" process, a conclusion also supported by noninvasive physiological studies (Sussman *et al.*, 1999; Nager *et al.*, 2003). These results are therefore consistent with Bregman's (1990) characterization of stream segregation as a "primitive" perceptual phenomenon that can be subsequently modulated by attention and by top-down "schema-based" processes, based upon task demands and background knowledge of sound characteristics (see also Bey and McAdams, 2002). Allocation of attention subsequent to or concurrent with automatic processing might also account for the "build-up" of stream segregation over time (Bregman, 1978; Anstis and Saida, 1985; Bregman, 1990; Carlyon *et al.*, 2001), a phenomenon that was not examined in the present study (recall that data acquisition began several seconds after the initiation of sequence presentation, thus precluding analysis of build-up effects). However, preliminary studies by Micheyl *et al.* (2003) in macaque A1 reproduced findings of Fishman *et al.* (2001a) and found that differential suppression of non-BF tone responses indeed required 5–10 s to build up, thereby providing additional evidence for the potential relevance of temporal response patterns in A1 for the phenomenon of auditory stream segregation.

D. Concluding remarks

While the present physiological results parallel psychoacoustic data on stream segregation in humans, several qualifications temper the conclusion that the response patterns described here represent neural substrates of perceptual stream segregation. First, since behavioral experiments were not conducted in parallel with the physiological recordings, it is unclear whether the monkeys participating in this study were able to perceptually segregate the alternating tone sequences similarly to humans. Indeed, since recordings were

conducted under passive listening conditions, changes in alertness and attentional state may account for discrepancies between monkey physiological and human perceptual data. Second, several psychoacoustic studies in humans have shown that stream segregation can occur even in the absence of spectral differences between ''A'' and ''B'' acoustic stimuli (Vliegen and Oxenham, 1999; Grimault *et al.*, 2002; Roberts *et al.*, 2002; for review see Moore and Gockel, 2002). For example, Vliegen and Oxenham (1999) demonstrated that subjects were able to perceptually segregate two alternating harmonic complex tones of different fundamental frequency, despite the fact that both complex tones were high-pass filtered within the same passband so as to minimize spectral cues and the spectral resolvability of their frequency components. Similarly, Grimault *et al.* (2002) showed that stream segregation can occur solely on the basis of differences in the amplitude-modulation rate of alternating broadband noise bursts with a common spectrum, and therefore exclusively on the basis of envelope periodicities. Stream segregation can also occur solely on the basis of intensity differences between ''A'' and ''B'' stimuli (van Noorden, 1975). Therefore, while spectral differences may be sufficient for demonstrating stream segregation, and indeed, appear to be a dominant factor (Singh, 1987; Vliegen *et al.*, 1999), they are not necessary, as listeners can segregate sounds into separate streams when successive sounds differ along either a spectral or a temporal dimension (Singh and Bregman, 1997; Vliegen *et al.*, 1999). Preliminary psychoacoustic studies have shown that forward masking can occur between amplitude-modulated sounds in a modulation-rate specific manner, suggesting the existence of modulation-rate selective neural channels (Wojtczak and Viemeister, 2004). A forward masking mechanism, similar to that proposed here, might therefore also account for stream segregation on the basis of stimulus periodicity. Since the present study used pure tone stimuli, it is unclear whether the present findings can account for stream segregation in the case of sequences comprised of more natural, spectrally complex stimuli such as speech and music. However, one might predict that physiological principles similar to those described here would apply to segregating complex sounds with non-overlapping spectra (Bregman *et al.*, 2001) or with distinct spectral peaks, e.g., formants (Bregman *et al.*, 1990).

The fact that temporal response patterns evoked by alternating tone sequences persist after averaging activity across electrode penetrations (Figs. 3–5) indicates that neural ensemble responses are highly synchronized across cortical sites in A1. Such synchronous bursts of activity would be expected to facilitate transmission of information from primary to secondary auditory cortical areas and beyond for further processing (Lisman, 1997; Usrey and Reid, 1999). Since our results are based on multi-unit cluster activity, it is unclear whether the response patterns observed in the present study would be exhibited by individual auditory cortical neurons. However, Brosch and Schreiner (1997) found nearly identical response patterns for single- and multi-units in their physiological forward masking studies in A1, suggesting that multi-unit data are generally representative of single-unit results, at least for the phenomena described in the present study. Since A1 represents but a single node in a network of both parallel and hierarchically organized auditory fields in the superior temporal gyrus (Rouiller *et al.*, 1991; Morel *et al.*, 1993; Rauschecker *et al.*, 1995; Kaas and Hackett, 1998; Howard *et al.*, 2000), it remains to be determined how activity patterns underlying ''neural stream segregation'' in A1 are transformed at higher levels of the auditory system, and how such information is modulated by attention and by higher-order ''schema-based'' processes in the analysis of auditory scenes containing complex acoustic stimuli such as speech and music.

ACKNOWLEDGMENTS

We thank the Associate Editor and four anonymous reviewers for their helpful comments on an earlier version of the manuscript. Dr. Steven Walkley, May Huang, Linda O'Donnell, and Shirley Seto provided excellent technical, secretarial, and histological assistance. This research was supported by NIH Grant Nos. DC-00657 and NS-07098.

Alain, C., and Woods, D. L. (**1994**). ''Signal clustering modulates auditory cortical activity in humans,'' Percept. Psychophys. **56**, 501–516.

Anstis, S., and Saida, S. (**1985**). ''Adaptation to auditory streaming of frequency-modulated tones,'' J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Arehole, S., Salvi, R. J., Saunders, S. S., and Hamernik, R. P. (**1987**). ''Evoked response 'forward masking' functions in chinchillas,'' Hear. Res. **30**, 23–32.

Bakst, I., and Amaral, D. G. (**1984**). ''The distribution of acetylcholinesterase in the hippocampal formation of the monkey,'' J. Comp. Neurol. **225**, 344–371.

Barna, J., Arezzo, J. C., and Vaughan, Jr., H. G. (**1981**). ''A new multicontact array for the simultaneous recording of field potentials and unit activity,'' Electroencephalogr. Clin. Neurophysiol. **52**, 494–496.

Beauvois, M. W. (**1998**). ''The effect of tone duration on auditory stream formation,'' Percept. Psychophys. **60**, 852–861.

Bey, C., and McAdams, S. (**2002**). ''Schema-based processing in auditory scene analysis,'' Percept. Psychophys. **64**, 844–854.

Boettcher, F. A., Salvi, R. J., and Saunders, S. S. (**1990**). ''Recovery from short-term adaptation in single neurons in the cochlear nucleus,'' Hear. Res. **48**, 125–144.

Botte, M.-C., Drake, C., Brochard, R., and McAdams, S. (**1997**). ''Perceptual attenuation of nonfocused auditory streams,'' Percept. Psychophys. **59**, 419–425.

Bregman, A. S. (**1978**). ''Auditory streaming is cumulative,'' J. Exp. Psychol.: Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA, 1990).

Bregman, A. S., and Campbell, J. (**1971**). ''Primary auditory stream segregation and perception of order in rapid sequences of tones,'' J. Exp. Psychol. **89**, 244–249.

Bregman, A. S., Ahad, P. A., and Van Loon, C. (**2001**). ''Stream segregation of narrow-band noise bursts,'' Percept. Psychophys. **63**, 790–797.

Bregman, A. S., Liao, C., and Levitan, R. (**1990b**). ''Auditory grouping based on fundamental frequency and formant peak frequency,'' Can. J. Psychol. **44**, 400–413.

Bregman, A. S., Ahad, P. A., Crum, P. A. C., and O'Reilly, J. (**2000**). ''Effects of time intervals and tone durations on auditory stream segregation,'' Percept. Psychophys. **62**, 626–636.

Brosch, M., and Schreiner, C. E. (**1997**). ''Time course of forward masking tuning curves in cat primary auditory cortex,'' J. Neurophysiol. **77**, 923–943.

Calford, M. B., and Semple, M. N. (**1995**). ''Monaural inhibition in cat auditory cortex,'' J. Neurophysiol. **73**, 1876–1891.

Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). ''Effects of attention and unilateral neglect on auditory stream segregation,'' J. Exp. Psychol. Hum. Percept. Perform. **27**, 155–127.

Colombo, M., Rodman, H. R., and Gross, C. G. (**1996**). ''The effects of superior temporal cortex lesions on the processing and retention of auditory information in monkeys (*Cebus apella*),'' J. Neurosci. **16**, 4501–4517.
Cowey, A., and Weiskrantz, L. (**1976**). ''Auditory sequence discrimination in *Macaca mulatta*: the role of the superior temporal cortex,'' Neuropsychologia **14**, 1–10.
Cox, C. L., Metherate, R., Weinberger, N. M., and Ashe, J. H. (**1992**). ''Synaptic potentials and effects of amino acid antagonists in the auditory cortex,'' Brain Res. Bull. **28**, 401–410.
Creutzfeldt, O., Hellweg, F. C., and Schreiner, C. (**1980**). ''Thalamocortical transformation of responses to complex auditory stimuli,'' Exp. Brain Res. **39**, 87–104.
Dai, H., Scharf, B., and Buus, S. (**1991**). ''Effective attenuation of signals in noise under focused attention,'' J. Acoust. Soc. Am. **89**, 2837–2842.
de Ribaupierre, F., Goldstein, Jr., M. H., and Yeni-Komshian, G. (**1972**). ''Cortical coding of repetitive acoustic pulses,'' Brain Res. **48**, 205–225.
Delgutte, B. (**1990**). ''Physiological mechanisms of psychophysical masking: observations from auditory-nerve fibers,'' J. Acoust. Soc. Am. **87**, 791–809.
Demany, L. (**1982**). ''Auditory stream segregation in infancy,'' Inf. Behav. Dev. **5**, 261–276.
Eggermont, J. J. (**1994**). ''Neural interaction in cat primary auditory cortex II. Effects of sound stimulation,'' J. Neurophysiol. **71**, 246–270.
Eggermont, J. J. (**2000a**). ''Sound-induced synchronization of neural activity between and within three auditory cortical areas,'' J. Neurophysiol. **83**, 2708–2722.
Eggermont, J. J. (**2000b**). ''Neural responses in primary auditory cortex mimic psychophysical, across-frequency-channel, gap-detection thresholds,'' J. Neurophysiol. **84**, 1453–1463.
Fay, R. R. (**1998**). ''Auditory stream segregation in goldfish (*Carassius auratus*),'' Hear. Res. **120**, 69–76.
Fay, R. R. (**2000**). ''Spectral contrasts underlying auditory stream segregation in goldfish (*Carassius auratus*),'' J. Assoc. Res. Otolaryngol. **1**, 120–128.
Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2001a**). ''Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey,'' Hear. Res. **151**, 167–187.
Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., and Steinschneider, M. (**2001b**). ''Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans,'' J. Neurophysiol. **86**, 2761–2788.
Freeman, J. A., and Nicholson, C. (**1975**). ''Experimental optimization of current source density techniques for anuran cerebellum,'' J. Neurophysiol. **38**, 369–382.
Frisina, R. D. (**2001**). ''Subcortical neural coding mechanisms for auditory temporal processing,'' Hear. Res. **158**, 1–27.
Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., and Shibasaki, H. (**1998**). ''Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer,'' Brain Res. Cognit. Brain Res. **7**, 99–109.
Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). ''Auditory stream segregation on the basis of amplitude-modulation rate,'' J. Acoust. Soc. Am. **111**, 1340–1348.
Hackett, T. A., Stepniewska, I., and Kaas, J. H. (**1998**). ''Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys,'' J. Comp. Neurol. **394**, 475–495.
Heil, P. (**2001**). ''Representation of sound onsets in the auditory system,'' Audiol. Neuro-Otol. **6**, 167–172.
Horikawa, J., Hosokawa, Y., Nasu, M., and Taniguchi, I. (**1997**). ''Optical study of spatiotemporal inhibition evoked by two-tone sequences in the guinea pig auditory cortex,'' J. Comp. Physiol. [A] **181**, 677–684.
Howard, M. A., Volkov, I. O., Mirsky, R., Garell, P. C., Noh, M. D., Granner, M., Damasio, H., Steinschneider, M., Reale, R. A., Hind, J. E., and Brugge, J. F. (**2000**). ''Auditory cortex on the human posterior superior temporal gyrus,'' J. Comp. Neurol. **416**, 79–92.
Hulse, S. H., MacDougall-Shackleton, S. A., and Wisniewski, A. B. (**1997**). ''Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (Sturnus vulgaris),'' J. Comp. Psychol. **111**, 3–13.
Izumi, A. (**2002**). ''Auditory stream segregation in Japanese monkeys,'' Cognition **82**, B113–B122.
Jancke, L., Mirzazade, S., and Shah, N. J. (**1999**). ''Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects,'' Neurosci. Lett. **266**, 125–128.
Jerison, H. J., and Neff, W. D. (**1953**). ''Effect of cortical ablation in the monkey on discrimination of auditory patterns,'' Fed. Proc. **12**, 73–74.
Kaas, J. H., and Hackett, T. A. (**1998**). ''Subdivisions of auditory cortex and levels of processing in primates,'' Audiol. Neuro-Otol. **3**, 73–85.
Kanwal, J. S., Medvedev, A. V., and Micheyl, C. (**2003**). ''Neurodynamics for auditory stream segregation: tracking sounds in the mustached bat's natural environment,'' Network **14**, 413–35.
Kelly, J. B., Rooney, B. J., and Phillips, D. P. (**1996**). ''Effects of bilateral auditory cortical lesions on gap-detection thresholds in the ferret (*Mustela putorius*),'' Behav. Neurosci. **110**, 542–550.
Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., and Chauvel, P. (**1998**). ''Contribution of different cortical areas in the temporal lobes to music processing,'' Brain **121**, 1853–1867.
Lisman, J. E. (**1997**). ''Bursts as a unit of neural information: making unreliable synapses reliable,'' Trends Neurosci. **20**, 38–43.
MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., and White, W. (**1998**). ''Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences,'' J. Acoust. Soc. Am. **103**, 3581–3587.
McAdams, S., and Bertoncini, J. (**1997**). ''Organization and discrimination of repeating sound sequences in newborn infants,'' J. Acoust. Soc. Am. **102**, 2945–2953.
McAdams, S., and Bregman, A. (**1979**). ''Hearing musical streams,'' Comput. Music J. **3**, 26–63.
Merzenich, M. M., and Brugge, J. F. (**1973**). ''Representation of the cochlear partition on the superior temporal plane of the macaque monkey,'' Brain Res. **50**, 275–296.
Metherate, R., and Ashe, J. H. (**1994**). ''Facilitation of and NMDA receptor-mediated EPSP by paired-pulse stimulation in rat neocortex via depression of GABAergic IPSPs,'' J. Physiol. **481**, 331–348.
Metherate, R., and Ashe, J. H. (**1995**). ''Synaptic interactions involving acetylcholine, glutamate, and GABA in rat auditory cortex,'' Exp. Brain Res. **107**, 59–72.
Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (**2003**). ''The neural basis of stream segregation in the auditory cortex,'' Abstract, Association for Research in Otolaryngology, 919.
Miller, G. A., and Heise, G. A. (**1950**). ''The trill threshold,'' J. Acoust. Soc. Am. **22**, 637–638.
Mondor, T. A., and Bregman, A. S. (**1994**). ''Allocating attention to frequency regions,'' Percept. Psychophys. **56**, 268–276.
Moore, B. C. (**1978**). ''Psychophysical tuning curves measured in simultaneous and forward masking,'' J. Acoust. Soc. Am. **63**, 524–532.
Moore, B. C. J., and Gockel, H. (**2002**). ''Factors influencing sequential stream segregation,'' Acust. Acta Acust. **88**, 320–332.
Morel, A., Garraghty, P. E., and Kaas, J. H. (**1993**). ''Tonotopic organization, architechtonic fields and connections of auditory cortex in macaque monkeys,'' J. Comp. Neurol. **355**, 437–459.
Müller-Preuss, P., and Mitzdorf, U. (**1984**). ''Functional anatomy of the inferior colliculus and the auditory cortex: current source density analyses of click-evoked potentials,'' Hear. Res. **16**, 133–142.
Nager, W., Teder-Salejarvi, W., Kunze, S., and Munte, T. F. (**2003**). ''Preattentive evaluation of multiple perceptual streams in human audition,'' NeuroReport **14**, 871–874.
Phillips, D. P. (**1993**). ''Representation of acoustic events in the primary auditory cortex,'' J. Exp. Psychol. Hum. Percept. Perform. **19**, 203–216.
Phillips, D. P. (**1995**). ''Central auditory processing: a view from auditory neuroscience,'' Am. J. Otol. **3**, 338–352.
Rauschecker, J. P., Tian, B., and Hauser, M. (**1995**). ''Processing of complex sounds in the macaque nonprimary auditory cortex,'' Science **268**, 111–114.
Recanzone, G. H., Guard, D. C., and Phan, M. L. (**2000**). ''Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey,'' J. Neurophysiol. **83**, 2315–2331.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). ''Primitive stream segregation of tone sequences without differences in fundamental frequency or passband,'' J. Acoust. Soc. Am. **112**, 2074–2085.
Rouiller, E. M., Simm, G. M., Villa, A. E. P., de Ribaupierre, Y., and de Ribaupierre, F. (**1991**). ''Auditory cortico-cortical interconnections in the cat: evidence for parallel and hierarchical arrangement of the auditory cortical areas,'' Exp. Brain Res. **86**, 483–505.
Shore, S. E. (**1995**). ''Recovery of forward-masked responses in ventral cochlear nucleus neurons,'' Hear. Res. **82**, 31–43.

Singh, P. G. (**1987**). ''Perceptual organization of complex-tone sequences: a tradeoff between pitch and timbre?'' J. Acoust. Soc. Am. **82**, 886–899.
Singh, P. G., and Bregman, A. S. (**1997**). ''The influence of different timbre attributes on the perceptual segregation of complex-tone sequences,'' J. Acoust. Soc. Am. **102**, 1943–1952.
Smith, R. L. (**1979**). ''Adaptation, saturation, and physiological masking in single auditory-nerve fibers,'' J. Acoust. Soc. Am. **65**, 166–178.
Steinschneider, M., Fishman, Y. I., and Arezzo, J. C. (**2003**). ''Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey,'' J. Acoust. Soc. Am. **114**, 307–321.
Steinschneider, M., Reser, D. H., Fishman, Y. I., and Arezzo, J. C. (**1998**). ''Click train encoding in primary auditory cortex of the awake monkey: evidence for two different mechanisms subserving pitch perception,'' J. Acoust. Soc. Am. **104**, 2935–2955.
Steinschneider, M., Schroeder, C. E., Arezzo, J. C., and Vaughan, Jr., H. G. (**1994**). ''Speech-evoked activity in primary auditory cortex: effects of voice onset time,'' Electroencephalogr. Clin. Neurophysiol. **92**, 30–43.
Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (**1999**). ''Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex,'' J. Neurophysiol. **82**, 2346–2357.
Steinschneider, M., Tenke, C. E., Schroeder, C. E., Javitt, D. C., Simpson, G. V., Arezzo, J. C., and Vaughan, Jr., H. G. (**1992**). ''Cellular generators of the cortical auditory evoked potential initial component,'' Electroencephalogr. Clin. Neurophysiol. **84**, 196–200.
Sukov, W., and Barth, D. S. (**1998**). ''Three-dimensional analysis of spontaneous and thalamically evoked gamma oscillations in auditory cortex,'' J. Neurophysiol. **79**, 2875–2884.
Sussman, E., Ritter, W., and Vaughan, Jr., H. G. (**1999**). ''An investigation of the auditory streaming effect using event-related potentials,'' Psychophysiology **36**, 22–34.
Usrey, W. M., and Reid, R. C. (**1999**). ''Synchronous activity in the visual system,'' Annu. Rev. Physiol. **61**, 435–456.
van Noorden, L. P. A. S. (**1975**). ''Temporal Coherence in the Perception of Tone Sequences,'' Ph.D. thesis, The Netherlands, Tech. Hogeschool, Eindhoven, Institute of Perception Research.
Vaughan, H. G., Jr., and Arezzo, J. C. (**1988**). ''The neural basis of event-related potentials,'' in *Human-Event Related Potentials, EEG Handbook (Revised Series)*, edited by T. W. Picton (Elsevier, New York, 1988), Vol. 3, pp. 45–96.
Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). ''The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task,'' J. Acoust. Soc. Am. **106**, 938–945.
Volkov, I. O., and Galazyuk, A. V. (**1992**). ''Peculiarities of inhibition in cat auditory cortex neurons evoked by tonal stimuli of various durations,'' Exp. Brain Res. **91**, 115–120.
Wallace, M. N., Kitzes, L. M., and Jones, E. G. (**1991**). ''Chemoarchitectonic organization of the cat primary auditory cortex,'' Exp. Brain Res. **86**, 518–526.
Walton, J. P., Frisina, R. D., Ison, J. R., and O'Neill, W. E. (**1997**). ''Neural correlates of behavioral gap detection in the inferior colliculus of the young CBA mouse,'' J. Comp. Physiol. [A] **181**, 161–176.
Wojtczak, M., and Viemeister, N. F. (**2004**). ''Forward masking of amplitude modulation,'' Abstract, Association for Research in Otolaryngology, 838.

Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise

Frédéric Apoux[a] and Sid P. Bacon
Psychoacoustics Laboratory, Department of Speech and Hearing Science, Arizona State University, PO Box 870102, Tempe, Arizona 85287-0102

(Received 14 October 2003; revised 18 June 2004; accepted 21 June 2004)

The relative importance of temporal information in broad spectral regions for consonant identification was assessed in normal-hearing listeners. For the purpose of forcing listeners to use primarily temporal-envelope cues, speech sounds were spectrally degraded using four-noise-band vocoder processing. Frequency-weighting functions were determined using two methods. The first method consisted of measuring the intelligibility of speech with a hole in the spectrum either in quiet or in noise. The second method consisted of correlating performance with the randomly and independently varied signal-to-noise ratio within each band. Results demonstrated that all bands contributed equally to consonant identification when presented in quiet. In noise, however, both methods indicated that listeners consistently placed relatively more weight upon the highest frequency band. It is proposed that the explanation for the difference in results between quiet and noise relates to the shape of the modulation spectra in adjacent frequency bands. Overall, the results suggest that normal-hearing listeners use a common listening strategy in a given condition. However, this strategy may be influenced by the competing sounds, and thus may vary according to the context. Some implications of the results for cochlear implantees and hearing-impaired listeners are discussed. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781329]

PACS numbers: 43.66.Ba, 43.66.Mk, 43.71.Es [NFV] Pages: 1671–1680

I. INTRODUCTION

In addition to the traditional Fourier-based spectral approach to speech understanding, in which speech information is conveyed by the distribution of energy along the audio-frequency axis, there is a complementary approach that is based on the temporal structure of speech (Plomp, 1983; Rosen, 1992). A number of behavioral studies have shown that usable speech information is conveyed by the low-frequency amplitude fluctuations (i.e., the temporal envelope) of speech, and that normal-hearing (NH) (van Tasell *et al.*, 1987, 1992; Shannon *et al.*, 1995; Lorenzi *et al.*, 1999; Smith *et al.*, 2002) and hearing-impaired listeners (Turner *et al.*, 1995; Apoux *et al.*, 2001; Healy and Bacon, 2002) can use this information to recognize speech. Recently, Crouzet and Ainsworth (2001) showed that temporal envelopes extracted from distant frequency regions are only partially correlated, indicating that the shape of the modulation spectrum may differ across frequency. Consistent with previous reports (Plomp, 1983; Houtgast and Steeneken, 1985), their results indicate that for envelope modulations $\leqslant$4 Hz (reflecting the average syllable rate in speech), the correlation of envelope information between frequency bands is very high ($>$0.8). For envelope modulations $>$8 Hz, the correlation remains significant only for adjacent spectral bands, suggesting that temporal envelopes extracted from remote frequency regions do not necessarily convey the same information.

So far, various methods have been used to estimate the relative importance of discrete frequency regions. One method is based on the use of a correlational procedure (Doherty and Turner, 1996). To determine the weights applied to various frequency regions of speech, the information in each speech band is independently and randomly degraded by a given amount on each trial by the addition of noise. The weight the subject places upon each band is obtained by calculating the correlation between the amount of degradation in each band (specified as the signal-to-noise ratio in the band) and the correctness of the responses. A second method (the ''hole'' method) consists of creating holes or gaps in the spectrum of speech (Shannon *et al.*, 2001; Kasturi *et al.*, 2002). The weight of a given frequency region is assessed from the drop in performance observed when this particular frequency band is omitted.

To our knowledge, no one has evaluated the relative weight of temporal information in different frequency regions. Thus, the main goal of the present experiments is to address the relative importance of various frequency regions for speech intelligibility, and to investigate more closely the differential contribution of temporal-envelope cues across frequency. Analysis of the speech signal reveals that high-frequency bands contain a significantly greater amount of energy in the mid-frequency modulation spectrum (10–25 Hz) than the lower portion of the frequency spectrum (Greenberg *et al.*, 1998). Accordingly, we reasoned that if modulation frequencies above 8 Hz are only partly correlated over frequency and that amplitude fluctuations up to at least 16 Hz play a significant role in speech identification (Drullman *et al.*, 1994; van der Horst *et al.*, 1999), spectral regions whose modulation spectrum exhibits a significant amount of energy in the 8–16-Hz range (i.e., the high-

[a]Electronic mail: frederic.apoux@asu.edu

frequency region) should be more relevant than others in terms of temporal-envelope cues. On the other hand, spectral regions whose modulation spectrum does not exhibit a great amount of energy above 8 Hz should not play a significant role in the intelligibility of spectrally degraded speech as the same information is available in the other speech bands. A subsidiary purpose of the present study is to compare frequency-weighting functions obtained in different background conditions. Whereas results differ somewhat across studies, one consistent finding is that the measured functions are remarkably similar across listeners tested in a given study. This suggests that NH listeners employ a very similar strategy in a given listening condition but that the strategy varies as a function of the listening context. To examine this issue, consonant identification was measured in both quiet and noisy conditions when using the hole method and only in noisy conditions when using the correlational method.

II. EXPERIMENT

A. Methods

1. Subjects

Six NH listeners (four females, two males) ranging in age from 20 to 29 years participated. One (S3) was author AF. Normal hearing was defined as having pure-tone air-conduction thresholds $\leqslant$20 dB HL (ANSI, 1996) at octave frequencies from 250 to 8000 Hz in both ears. All participants were native American English speakers, except AF who alone had extensive previous experience identifying spectrally degraded speech sounds.

2. Stimuli

The speech stimuli consisted of lists of consonant-vowel (CV) and vowel-consonant (VC) syllables from the UCLA version of the Nonsense Syllable Test, for which the frequency-importance functions have already been determined with both AI and correlational methods. The vowels /a, i or u/ are paired with either 21 (VC) or 22 (CV) different consonants. Each list is spoken by a female and a male talker for a total of 258 items split into 12 lists.

The stimuli were spectrally reduced by processing them through a noise-band processor as used by Shannon *et al.* (1995). Each stimulus was split into four broad frequency bands (third-order Butterworth filter). In each band, the temporal envelope was extracted by full-wave rectification and low-pass filtering (sixth-order Butterworth filter) at 500 Hz so as to provide both envelope and periodicity information (Rosen, 1992). The resulting envelope was then used to modulate a white noise. Finally, the modulated noise was frequency-limited by filtering with the same band-pass filters used in the original analysis band and the obtained modulated bands of noise were combined and low-pass filtered at 5 or 10 kHz according to the particular condition. Two cutoff frequency conditions were tested to allow direct comparisons with previous work. In one condition (TB) the four bands were determined as in the Turner *et al.* (1998) study (1120 Hz and below, 1120–2250 Hz, 2250–3500 Hz, and 3500–10 000 Hz). In the other condition (LB), bands were determined according to the Lorenzi *et al.* (1999) study (800 Hz and below, 800–1500 Hz, 1500–2500 Hz, and 2500–5000 Hz). Low-frequency bands will be referred to throughout this paper as band 1, mid-low frequency bands as band 2, mid-high frequency bands as band 3, and high-frequency bands as band 4.

The speech-modulated noise was calibrated so that its overall level was 70 dB SPL when presented alone. When needed, it was mixed with simplified speech spectrum-shaped noise (constant spectrum level below 500 Hz and 6 dB/oct roll-off above 500 Hz) created at each trial. The noise was bandpass filtered to correspond to the speech bands. For each trial, the level of the band of noise added to a given speech band was computed independently in each of the four bands, so that the difference (in dB) between the rms levels of a speech band and the corresponding band of noise was identical in each band.

3. Procedure

Listeners were tested individually in a double-walled, sound-attenuated booth using a single-interval, 21- or 22-alternative procedure. The experiments were performed on a PC equipped with a soundcard. Stimuli were played to the listeners binaurally through Sennheiser HD 250 Linear II circumaural headphones. Listeners were instructed to ignore the vowel identity and to report only the perceived consonant. No feedback was given during the practice or the experimental sessions.

Each subject initially completed recognition of all 12 lists (each list was presented three times) presented in random order for each cutoff frequency condition. On any given trial, a background noise was added at various signal-to-noise (S/N) ratios randomly selected from 30 ± 18 dB in 3-dB steps. On the basis of this first session, the S/N ratio that the experimenter estimated would yield an approximate one-third decrease in recognition score from the best one obtained was determined for each cutoff frequency condition. This S/N ratio will be referred to throughout this paper as the S/N midpoint. Then, each subject performed the same task again at various S/N ratios randomly selected from the S/N midpoint ± 18 dB in 3-dB steps. In these two practice sessions, the four bands were presented together and the S/N ratio was the same in each band but was computed independently in each band (as described in Sec. II A 2). At the end of the second practice session, the S/N midpoint was readjusted as necessary according to the new data. Notice that for all subjects, the S/N midpoint was determined separately for the two cutoff frequency conditions. However, performance was very similar in both conditions with only a 2% difference. Thus, the same S/N midpoint was used in both conditions.

Two methods were used to assess the perceptual importance of temporal information. The first method consisted of presenting only three out of the four bands in quiet or in noise at the S/N midpoint determined in the practice sessions. To remove a given band, stimuli were synthesized using only three bands and when speech stimuli were presented in a background noise, no noise was added in the spectral region where the speech band was omitted. Analyzed data consisted of the results taken from the 258-item speech ma-

terials consisting of the 12 lists presented once in each band condition for a total of 1032 trials in each cutoff frequency condition. To obtain the perceptual importance of each band we followed the least-squares approach initially proposed by Ahumada and Lovell (1970) and described in a way that was easier to implement in Kasturi *et al.* (2002).[1] The second method, the correlational method, consisted of presenting the four speech bands together in a background noise. The S/N ratio in each band was randomly chosen (in 3-dB steps) from a rectangular distribution that was 36-dB wide, the midpoint of which was determined from the practice sessions and corresponded to the S/N midpoint. On each trial, the S/N ratio chosen for a given band was independent of the S/N ratio in each of the other three bands. The weight placed upon a frequency band was estimated by a point biserial correlation between the S/N ratio in a band and the response (whether correct or incorrect). The four raw correlations for each subject were transformed to relative weights by summing their values and expressing each band's weight as the raw correlation divided by this sum. Therefore, the sum of the relative weights of the four bands equals 1.0. Each list was presented ten times for a total of 2580 trials in each cutoff frequency condition.

The two methods were administered in the following way. First, consonant recognition in quiet was measured for stimuli having one of the four bands removed for the two cutoff frequency conditions. The order in which bands were removed was counterbalanced to avoid order effects. However, this order was the same for both cutoff frequency conditions for a given subject. Second, consonant recognition in noise was measured using the hole and correlational methods, which were counterbalanced among subjects. Three subjects first performed the task corresponding to the correlational method and the remaining three subjects first performed the task corresponding to the hole method.

B. Results and discussion

From the data of the two practice sessions, in which each listener was tested with the four bands presented together, a S/N midpoint was determined separately for each listener. During the first practice session, listeners were able to identify up to 60% of the consonants. Performance obtained in both cutoff frequency conditions was relatively similar, with mean scores averaged across S/N ratios from 12 to 36 dB of 29% and 33% for the 5000- (LB) and 10000-Hz (TB) conditions, respectively. Mean identification scores measured in the second practice session improved only slightly compared to those measured in the first session. This improvement in consonant identification scores was the same in both cutoff frequency conditions (4%). The S/N midpoint used for each subject is shown in Table I. Also listed in Table I are the mean identification scores obtained for both cutoff frequency conditions in the second practice session at S/N ratios ranging from 12 to 36 dB in 3-dB steps.

1. Hole method in quiet

The pattern of results was approximately the same for the six listeners, thus individual data are not reported. The mean percent-correct scores are shown in Fig. 1 for the two cutoff frequency conditions as a function of the removed band. Compared to results obtained in the practice sessions with all four bands, the drop in performance was minimal. This suggests that none of the four bands was critical for consonant identification in quiet and, in a way, is consistent with previous studies showing that only a slight improvement in identification of spectrally reduced consonants is observed when the number of bands is increased from three to four (Shannon *et al.*, 1995; Dorman *et al.*, 1997). Visual inspection of Fig. 1 also suggests that identification scores were only slightly affected by the location of the removed band. Finally, performance appeared to be independent of the cutoff frequency condition. Statistical analysis (ANOVA) confirmed that the main effect of cutoff frequency was not significant ($p=0.49$). The main effect of the removed band was, however, significant [$F(3,15)=6.7$, $p<0.005$]. Post-hoc tests (Tukey, alpha$=0.05$) showed that the scores obtained when the fourth band was removed were significantly lower than the scores obtained when the first or the second band was removed. Finally, the interaction of cutoff frequency and removed band was also significant [$F(3,15)=3.8$, $p<0.05$]. This reflects the finding that the difference in performance between removed-band conditions was greater for the LB condition than for the TB condition.

The averaged consonant confusion matrices were analyzed in terms of percent information transmitted (Miller and

TABLE I. The midpoint S/N ratio employed for final data collection for each subject. Also listed are each subject's overall percent correct scores in the two cut-off frequency conditions (TB and LB) for the second practice session at S/N ratios ranging from 12 to 36 dB in 3-dB steps.

Subject	S/N midpoint (dB)	TB mean (%)	LB mean (%)
S1	21	36	36
S2	21	39	38
S3	28	39	40
S4	24	25	29
S5	27	29	32
S6	30	20	24
Mean	23.5	31	33

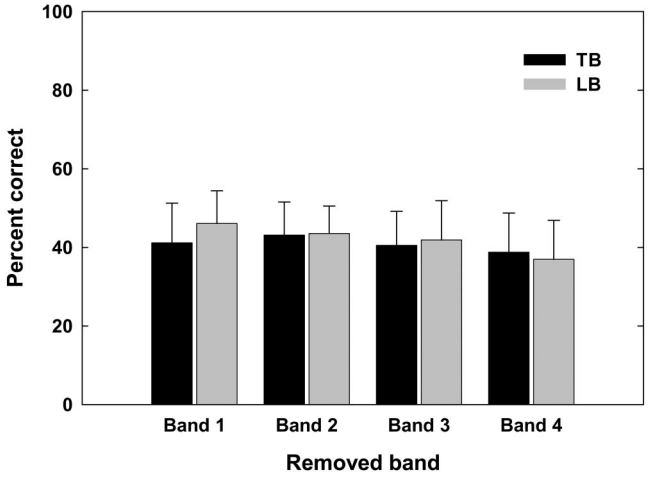


FIG. 1. Mean percent scores for consonant identification in quiet as a function of the band removed. The dark and the gray bars give the scores obtained in the TB and the LB conditions, respectively. Errors bars indicate one standard deviation.

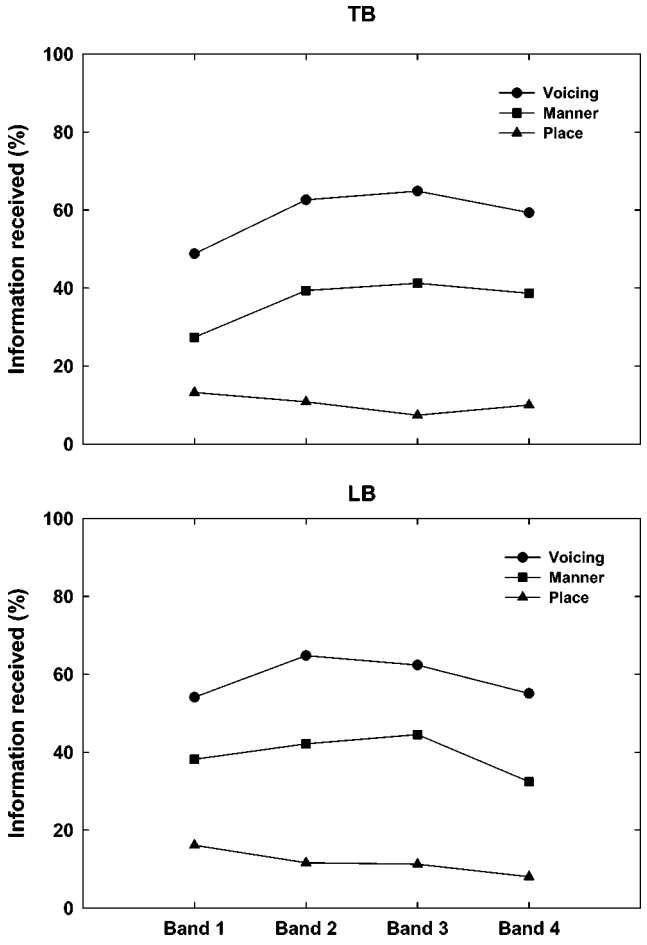


FIG. 2. Percent information transmitted for the features place, manner, and voicing as a function of the band removed. The upper and the lower panels give the scores computed for the TB and the LB conditions, respectively.

Nicely, 1955). The specific reception of three speech features (voicing, manner and place of articulation) was evaluated for each experimental condition. Results of the feature analysis are shown in Fig. 2. In both cutoff frequency conditions, information received regarding place of articulation was particularly poor, suggesting that errors in the overall scores were mainly associated with errors in identifying the place feature. More information was transmitted for voicing and manner. However, transmission of voicing was consistently better than transmission of manner. These results are in agreement with previous studies showing that reception of both voicing and manner information is less affected by degradation of spectral cues than reception of place of articulation information (Shannon *et al.*, 1995; Dorman *et al.*, 1997). The substantial reduction in the transmission of place feature can be considered as an indication that spectral cues were successfully reduced. Consequently, the frequency-weighting functions estimated in the present experiment should effectively reflect the importance of various frequency regions in terms of temporal information. As can be seen in Fig. 2, both voicing and manner information were most affected by a hole in the low-frequency (band 1) or in the high-frequency (band 4) region. These data were used to derive the perceptual weights for the four bands, as shown in Fig. 3. The weighting functions were similar for the two cutoff frequency conditions, and each band was approximately equal in its importance for identifying these consonants, although the weight for band 4 was somewhat greater than that for other bands.

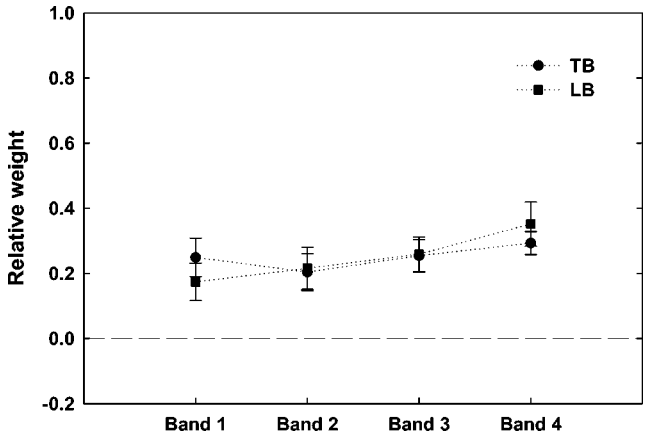


FIG. 3. Frequency-weighting functions derived from consonant identification scores measured in quiet in the TB (circles) and the LB (squares) conditions. Errors bars indicate ±1 standard deviation.

2. Hole method in noise

Again, the pattern of results was very similar among the six listeners and individual data are not presented. The averaged consonant identification scores as a function of the removed band are shown as histogram bars in Fig. 4 for both cutoff frequencies (the circles are the results in quiet, replotted from Fig. 1). Overall, the identification scores were lower than the corresponding condition in quiet; the average decrease in performance was slightly greater in the LB condition (10%) than in the TB condition (6%). Moreover, bands 2 and 4 were more affected by the addition of a background noise than were bands 1 and 3, except that in the LB condition there was a 13% decrease in performance when band 1 was removed. Finally, a clear decrease in consonant identification relative to that obtained in quiet was observed in both

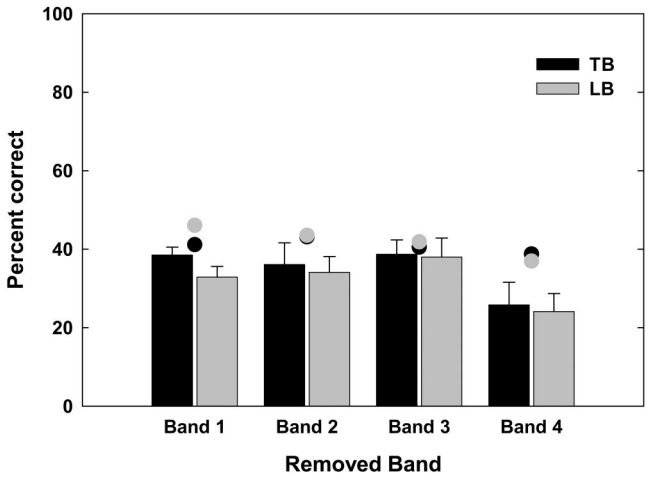


FIG. 4. Mean percent scores for consonant identification in noise as a function of the band removed. The dark and the gray bars give the scores obtained in the TB and the LB conditions, respectively. Errors bars indicate one standard deviation. For reference, the quiet data from Fig. 1 are also shown (filled circles).

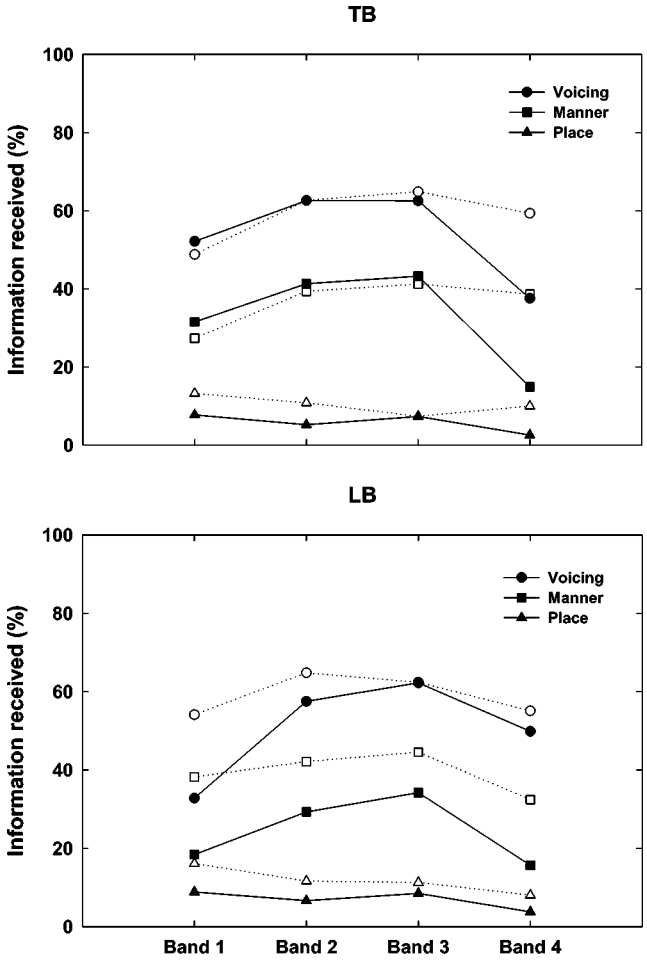


FIG. 5. Percent information transmitted for the features place, manner, and voicing as a function of the band removed. The upper and the lower panels give the scores computed for the TB and the LB conditions, respectively. For reference, data from Fig. 2 are replotted (open symbols).

cutoff frequency conditions when the fourth band was removed. This decrease ranged from 9% to 14%.

A two-way ANOVA with repeated measures showed a highly significant main effect of removed band [$F(3,15) = 35.2$, $p < 0.001$] and a mildly significant main effect of cutoff frequency [$F(1,15) = 6.9$, $p < 0.05$]; the interaction between the two was not significant ($p = 0.17$). Posthoc (Tukey, alpha = 0.05) comparisons confirmed that the scores obtained when the fourth band was removed were significantly ($p < 0.001$) lower than the scores obtained when the first, second, or third band was removed. The scores obtained when bands 1, 2, or 3 were removed were not statistically different from each other.

The consonant features voicing, manner of articulation, and place of articulation were computed and scored in terms of percent information transferred. The results are shown in Fig. 5. Reception of information tended to be poorer than in quiet (whose results are replotted as open symbols from Fig. 2). Nevertheless reception of place information was only slightly degraded by the addition of a background noise, probably because transmitted information for this feature was already poor in the quiet condition. Overall, the reception of voicing was largely unaffected by the addition of background noise, although certain individual bands were

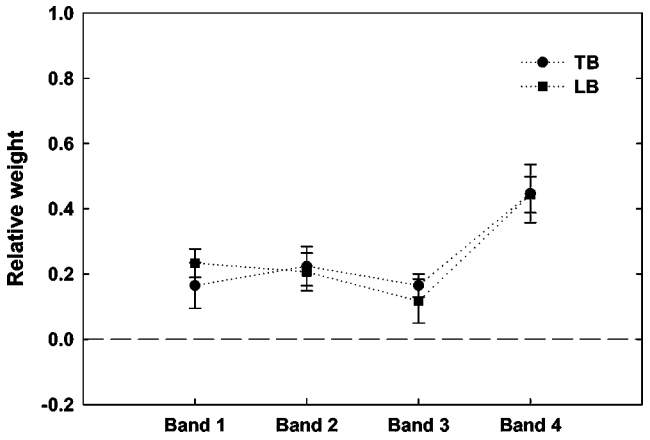


FIG. 6. Frequency-weighting functions derived from consonant identification scores measured in noise in the TB (circles) and the LB (squares) conditions with the hole procedure. Errors bars indicate ±1 standard deviation.

affected. In the TB condition (top panel), background noise mainly reduced reception of voicing when band 4 was omitted, whereas in the LB condition, reception of voicing was clearly degraded only when band 1 was removed. Consequently, scores for voicing were poorer when bands 1 and 4 were omitted. Finally, two different patterns were also observed for reception of manner information. In the TB condition, reception of manner of articulation was only affected by the addition of noise when the fourth band was removed, whereas in the LB condition, it was affected by the addition of noise when each band was removed. Overall, reception of manner was lower in conditions where bands 1 or 4 were omitted relative to conditions where bands 2 or 3 were omitted.

The relative weights of the various bands are shown in Fig. 6 for the TB and LB cutoff frequency conditions. The shapes of the two weighting functions were similar, with a relatively equal weighting across bands 1 to 3, suggesting that these three bands equally contributed to understanding these consonants. Band 4 was found to make the main contribution to identifying the consonants, with a mean weight of 0.45 in both cutoff frequency conditions.

3. Correlational method

Figure 7 displays the mean relative weights as a function of the removed band for the two cutoff frequency conditions. For each condition, the relative weights for each listener were based upon a total of 2580 trials across all 12 lists. Again, the averaged weighting functions obtained with the two cutoff frequency conditions were remarkably similar. Relative importance for bands 1, 2, and 3 was very similar, with weights ranging from 0.08 to 0.18. Consistent with weighting functions derived from the hole method in noise, the highest relative weight was placed upon band 4 in both TB and LB conditions (0.51 and 0.58, respectively). The interlistener variability was found to be greater here than with the hole method. This was mainly due to listener S6 who placed more weight upon band 1 than the others listeners (see Fig. 8, which is discussed in the next section). Another factor is that only one S/N ratio was tested with the

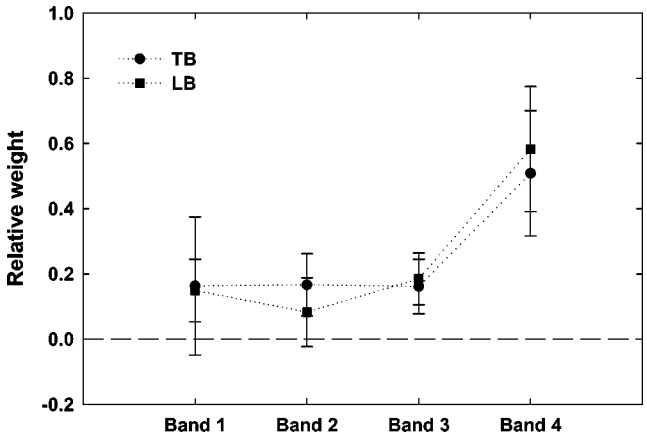


FIG. 7. Frequency-weighting functions derived from consonant identification scores measured in the TB (circles) and the LB (squares) conditions with the correlational procedure. Errors bars indicate ±1 standard deviation.

hole method whereas there were 28 561 possible combinations of S/N ratio for a given trial when using the correlational method. Knowing that each listener's frequency-weighting function was derived from a total of 2580 trials and that there were 28 561 possible combinations of S/N ratio, it is likely that a given combination of S/N ratios occurred only once or twice during the entire experiment. Consequently, one could consider that each subject was tested in a particular condition and the larger variability observed when using the correlational method probably arose from the large combination of S/N ratios.

It is interesting that the hole and correlational methods yield similar weighting functions, despite the fact that the listeners were provided with only three out of the four bands in the hole method. Based on a comparison of their own results with those of Mehr *et al.* (2001), Kasturi *et al.* (2002) concluded previously that these two methods yield similar results. The present study, however, is a stronger test of that, given that the same set of materials and subjects were used for both methods. Indeed, many differences (speech material, speech processing, absence versus presence of a background noise, bands cut-off frequencies) were apparent between the two aforementioned studies.

4. Individual data

The individual relative weights of the four bands obtained in noise with each method are plotted in Fig. 8 for each cutoff frequency condition. Overall, the pattern of individual frequency-weighting functions was roughly similar for each listener, with bands 1, 2, and 3 making a relatively small contribution to consonant identification compared to band 4. The weight of the fourth band ranged from 0.36 to 0.81 across the six listeners used in this experiment (except

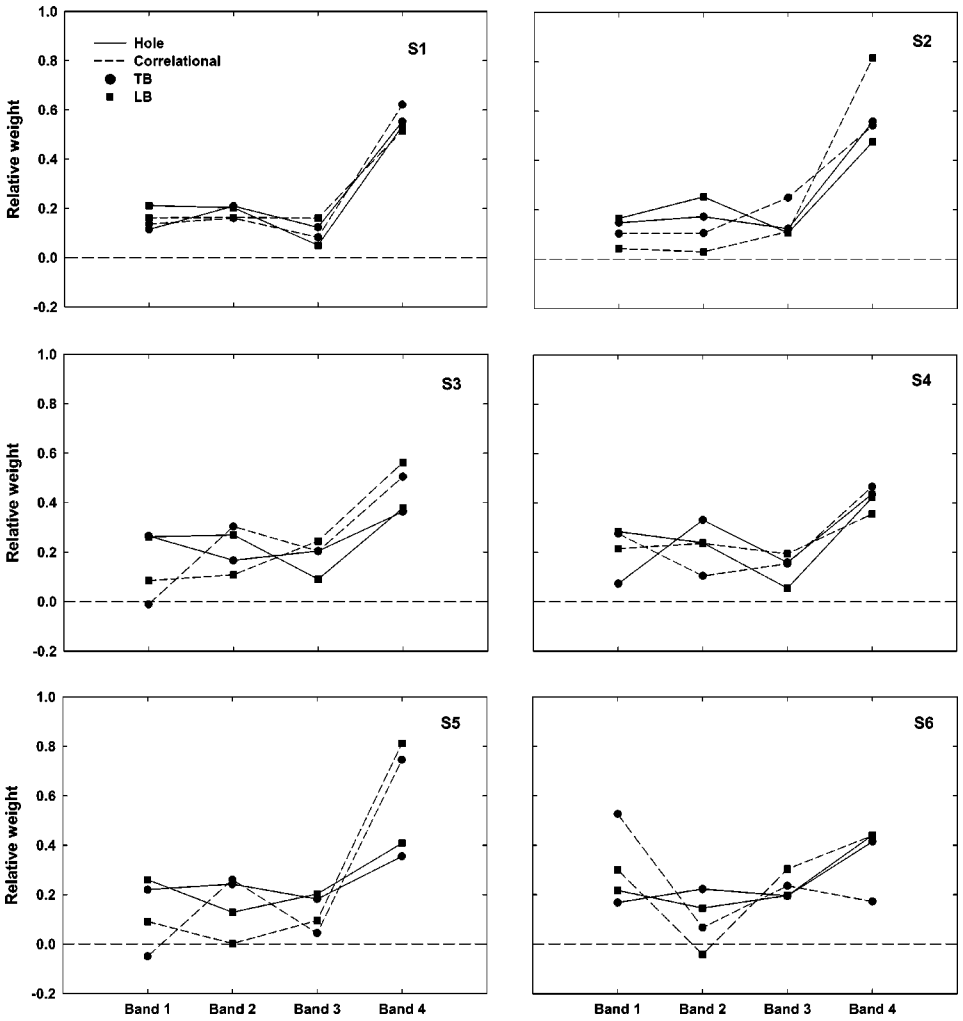


FIG. 8. Individual listener's frequency-weighting functions derived from consonant identification scores measured in noise in the TB (circles) and the LB (squares) conditions. The solid and the dashed lines give the frequency-weighting functions derived from the hole and the correlational methods, respectively.

0.17 for listener S6 in the TB condition with the correlational method). In two listeners, S1 and S2, patterns were very consistent across the conditions. These two listeners were also found to place a higher weight on band 4 than the other listeners. In listeners S3 and S4, the relative importance of the fourth band was still high, but less so than for listeners S1 and S2. The pattern for listener S5 was in between the two patterns described above. When measured using the correlational method, weighting functions were similar to those obtained in listeners S1 and S2, with substantially more weight placed on band 4. When measured using the hole method, weighting functions were similar to those obtained in listeners S3 and S4, with a smaller but still relatively greater weight placed on band 4. Finally, listener S6's weighting functions were also quite similar to those measured in listeners S3, S4, and S5 with the hole method. On the other hand, data derived from the correlational method indicate that this listener placed a higher weight on band 1 and that band 2 contributed little, if any, to consonant identification.

Despite quantitative differences in the weighting functions of the six listeners participating in this study, the same general strategy seems to be used by NH listeners in this speech-recognition task. Previous studies measuring frequency-weighting functions with the full spectrum of information available also reported similar patterns across NH listeners, suggesting that the similarity across listeners observed here was not due to our method of speech processing. Therefore, although frequency-weighting functions may differ as a function of the background condition, they remain roughly the same for all listeners. As suggested by Turner *et al.* (1998), this consistency across NH listeners implies that a valid normal weighting function could be defined, but the results of the present study suggest that the listening strategy implied by such a weighting function could be defined for a given listening condition only.

III. GENERAL DISCUSSION

The primary purpose of the present experiments was to determine the relative importance of temporal-envelope cues in distinct frequency regions for consonant identification. When speech stimuli were presented in quiet, results indicated that each band was approximately equal in its importance for consonant identification. When speech stimuli were presented against a background noise, band 4 was found to receive a substantially higher weight in both cutoff frequency conditions with the two methods. Overall, frequency-weighting functions were relatively independent of the cutoff frequency condition, indicating that the location of the cutoff frequencies defining the bands is not critical. This result is consistent with previous results. For example, Shannon *et al.* (1998) also showed that the exact crossover frequencies in a four-band representation are relatively minor factors in consonant identification when a variety of talkers are used. Consequently, the issue of cutoff frequency will not be considered further.

The higher weight of band 4 is consistent with results obtained by Lorenzi *et al.* (1999). They measured consonant identification under conditions of greatly reduced spectral information similar to the conditions tested in the present experiments. In addition, temporal modulations in the envelope of the speech stimuli were preserved, degraded, or expanded. Their results showed that expanding temporal fluctuations conveyed by band 1, 2, or 3 does not affect consonant identification in noise. However, their results showed a small but significant improvement in intelligibility when the temporal envelope in the fourth band was expanded. This improvement was found to correspond to an increase in the transmission of manner information. Silipo *et al.* (1999) also suggested that the contribution of amplitude fluctuations in the higher spectral bands is more important for intelligibility of spectrally reduced speech. In their experiment speech stimuli (sentences) were partitioned into 14 $\frac{1}{3}$-oct-wide bands. However, subjects were only presented with one to four of the 14 bands. The authors also manipulated the bandwidth of the speech modulation spectrum. In this manner, the modulation spectrum of one or two of the bands was low-pass filtered in a systematic fashion, allowing the authors to assess the importance of low-, mid-, and high-frequency modulations along the audio-frequency axis. Their results indicated that (*i*) only low-frequency modulations ($\leqslant$12 Hz) play a significant role in the low audio-frequency region and (*ii*) mid-frequency modulations (10–25 Hz) are of particular importance for encoding speech information in the spectral region above 1.5 kHz. Because low-rate amplitude variations are common across large regions of the speech spectrum and therefore may not be affected by degradation of temporal information in a given frequency region, it is not surprising that, in the present experiments, performance was not affected when temporal information located in the low-frequency region of the speech spectrum was removed or degraded. In contrast, removing or degrading mid-frequency modulations in the high audio-frequency region lead to a drop in performance because these modulations are specific to this region.

Considering that the frequency-weighting functions obtained in noise were more consistent with the particular importance of mid-frequency modulations in the spectral region above 1.5 kHz reported in previous studies, we propose two related ways to account for the difference in frequency-weighting functions obtained in quiet and in presence of a background noise. Both explanations are based on the possibility that listeners used information present in band 3 to compensate for the omission of band 4 when the stimuli were presented in quiet. A first possibility is that listeners used cues provided by the adjacent transition band. A number of studies have shown that narrow speech bands may transmit a great proportion of the speech spectrum outside their nominal filter passband, especially during time intervals when the speech signal contains little energy in the passband region (Warren and Bashford, 1999; Warren *et al.*, 2000). However, the addition of a background noise would likely have masked most—if not all—of the information provided by the transition bands. To assess the role of transition-band information, two out of our six listeners (S1 and S3) were retested in the hole condition in quiet. In this control experiment, the band analysis was performed using 1000-order FIR filters in order to reduce the availability of transition-band

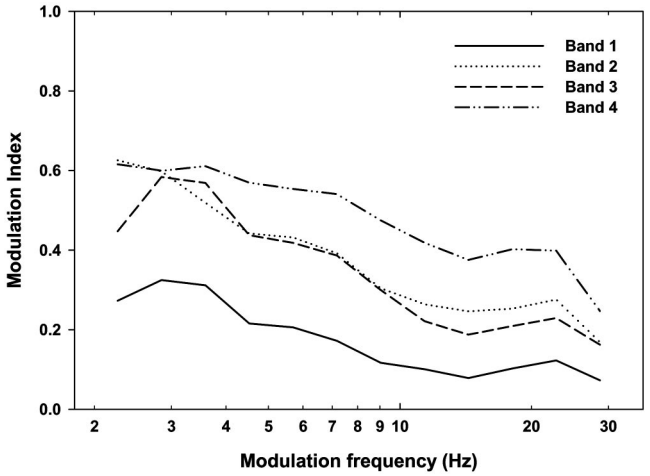


FIG. 9. Modulation spectra of the speech material used in the present experiment. For each band (TB condition), the modulation spectrum was computed from the entire set of speech stimuli.

information. Results showed that intelligibility scores and frequency-weighting functions derived from these new data were identical to those obtained in the main experiment (where third-order Butterworth filters were used). Therefore, the difference in frequency-weighting functions obtained in quiet and in the presence of a background noise almost certainly was not due to additional information provided by the transition bands in the quiet condition. A second possibility is that listeners used cues provided by the adjacent passband. Although we have established that correlation of envelope information remains significant for adjacent spectral bands (irrespective of modulation frequency), it can be seen in Fig. 9 that modulation depth for mid-frequency modulations is significantly lower in band 3 when compared to band 4. Accordingly, detection of fast modulation rates should be more affected in band 3 than in band 4 by the addition of a steady noise because of their inherently lower depth in the undisturbed speech signal. Thus, mid-frequency modulations in the third band were probably ''blurred'' when stimuli were presented against the background noise. As a result, listeners were not able to monitor these fast modulations.

Given the similarities in speech tokens and cutoff frequencies in our study and that of Turner *et al.* (1998), it is possible to assess the effect of spectral reduction on frequency-weighting functions. Comparison of the importance functions obtained in quiet and in noise already showed that listening strategies may vary with the context. The comparison between the results from Turner *et al.* (1998) and the present study suggests that relevant cues for consonant recognition are distributed differently across frequency regions when listeners are forced to rely more on temporal cues. In other words, listeners do not monitor the same regions of the spectrum when speech stimuli are spectrally reduced.

These findings may have clinical implications for both cochlear implantees and hearing-impaired listeners. First, they may be considered as relevant for implant users as the noise-band vocoder has been used as a model for the study of speech recognition in listeners with cochlear implants (Shannon *et al.*, 1995). However, the effects we observed are in the opposite direction to that found by Shannon *et al.* (2001), who showed that holes in the apical (i.e., low-frequency) region are more damaging than holes in the basal or middle regions. This difference probably reflects the fact that the implant users had six to ten electrodes stimulated simultaneously in Shannon's study, allowing a better spectral representation. Accordingly, Shannon *et al.* suggested that the primary cause of the effect was the loss of spectral information. Thus, the present results could be limited to situations where implant users have only few active electrodes available. However, it appears that most implant users are unable to fully utilize the spectral information provided by multielectrode devices. Evidence to support this comes from studies showing that an asymptote in performance is usually observed at four to seven channels in cochlear implantees while in NH listeners, performance continues to improve up to about 20 channels (Fishman *et al.*, 1997; Friesen *et al.*, 2001). Consequently, frequency-weighting functions observed in the present study may apply to implant users who cannot fully utilize all the active electrodes available and they indicate that basal electrodes should contribute most to the understanding of speech in these listeners.

Second, the particular distribution of temporal cues for consonant identification may also have implications for individuals with high-frequency hearing loss. People with cochlear hearing loss often have difficulty in understanding speech when background noise is present (see, e.g., Peters *et al.*, 1998). In persons with mild to moderate loss, the lack of audibility is often sufficient to account for speech identification abilities (Humes and Roberts, 1990). However, audibility cannot adequately predict the reduced speech identification in listeners with a severe hearing loss, suggesting that other factors affect the listeners' use of speech information in the audible signal. One factor is reduced frequency selectivity (see Moore, 1995, 1998 for reviews). Given that frequency resolution ability usually reduces with decreasing auditory sensitivity (Lutman *et al.*, 1991), people with moderate to severe cochlear hearing loss generally exhibit more severely reduced frequency selectivity. Based on the comparison between our results and those of Turner *et al.* (1998) showing that a severe reduction in frequency selectivity (as crudely simulated by the four-channel noise-band processor) results in an increased weight of the high-frequency region, it seems reasonable to speculate that a factor that could contribute to the difficulties of listeners with severe cochlear hearing loss in understanding speech is that the relevant information for consonant identification corresponds to a region where the hearing loss is more severe in most hearing-impaired listeners.

IV. SUMMARY AND CONCLUSION

The purpose of the present study was to determine frequency-importance functions for temporal information used to identify consonants in quiet and in noise. Two methods were used to derive the weight of four broad spectrally degraded speech bands in NH listeners. One method consisted of presenting three of the four bands. The other

method consisted of degrading each frequency band by a random amount of noise. The results can be summarized as follow:

(1) Consistent with previous studies (Lorenzi *et al.*, 1999; Silipo *et al.*, 1999), the present results suggest that temporal cues contained in frequency regions above 3.5 kHz would be relatively more important for consonant identification because those contained in the lower frequency regions are common across large portions of the speech spectrum. Listeners, however, might be able to partially compensate for the loss of temporal information in the high-frequency region in quiet conditions by using correlated information contained in the adjacent (lower) region.
(2) Given that low-frequency modulations are highly correlated across frequency, temporal-envelope cues specific to the high-frequency part of the speech spectrum would be conveyed by mid-and/or high-frequency modulations ($>$10 Hz).
(3) Taken together, the outcomes of the present experiments and of a previous study (Turner *et al.*, 1998) in which the speech stimuli and the band divisions were identical suggest that the distribution of the most critical information for speech intelligibility may vary with the type of corruption imposed on the speech signal (e.g., noise, spectral reduction).
(4) The frequency-weighting functions, however, are very similar across NH listeners, whether or not speech stimuli are spectrally reduced.
(5) Finally, the results of the present study suggest that listeners with limited access to the spectral details of speech may encounter greater difficulties in understanding speech when they simultaneously present some difficulties to access speech information from the high-frequency region.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Deafness and Other Communication Disorders (NIDCD Grant No. DC01376). We thank Takayuki Arai and Steven Greenberg for their generous help with computing modulation spectra. We also thank the two reviewers, Dianne van Tasell and Chris Turner, for their helpful comments on a previous version of this manuscript.

[1]The weight w_i of each frequency band was calculated by predicting the response of the listener as a linear combination of the strength of each band, i.e.,

$$R_k=\sum_{i=1}^{4} w_i E_{ik} \quad (1)$$

where R_k is the mean percent-correct score for condition k and E_{ik} is the strength of the ith band corresponding to condition k. The strength of each band is a binary value that can be either 0 or 1 depending on whether the band is off or on, respectively. The value of k ranges from 1 to 4 spanning all band combinations. Forming the prediction error e_k as

$$e_k=R_k-\sum_{i=1}^{4} w_i E_{ik} \quad (2)$$

we can estimate the band weights by minimizing the sum of all the squared errors with respect to w_i. Alternatively, Eq. (1) can be written in matrix form as

$$R=EW, \quad (3)$$

where R is a four-dimensional vector containing the mean percent-correct scores for conditions 1 to 4, E is the data matrix (4×4) consisting of the strengths of each band, and $W=[w_1,w_2,w_3,w_4]$ is a four-dimensional vector consisting of the desired band weights.

We calculated the weights W by solving the matrix equation given by Eq. (3) using a least-squares approach:

$$W=(E^TE)^{-1}E^TR. \quad (4)$$

After obtaining the solution from Eq. (4), we normalized the weights so that the sum of all weights was equal to 1.

Ahumada, A., and Lovell, J. (**1970**). ''Stimulus features in signal detection,'' J. Acoust. Soc. Am. **49**, 1751–1756.
ANSI (**1996**). *ANSI S3.6-1996, Specifications for Audiometers* (American National Standards Institute, New York).
Apoux, F., Crouzet, O., and Lorenzi, C. (**2001**). ''Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: Effects on identification performance and response times,'' Hear. Res. **153**, 123–131.
Crouzet, O., and Ainsworth W. A. (**2001**). ''On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation,'' Workshop on Consistent and Reliable Acoustic Cues for Sound Analysis, Aalborg, Denmark, September 2001.
Doherty, K. A., and Turner, C. W. (**1996**). ''Use of the correlational method to estimate a listener's weighting function of speech,'' J. Acoust. Soc. Am. **100**, 3769–3773.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). ''Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs,'' J. Acoust. Soc. Am. **102**, 2403–2411.
Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.
Fishman, K., Shannon, R. V., and Slattery, W. H. (**1997**). ''Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor,'' J. Speech Lang. Hear. Res. **40**, 1201–1215.
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). ''Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,'' J. Acoust. Soc. Am. **110**, 1150–1163.
Greenberg, S., Arai, T., and Silipo, R. (**1998**). ''Speech intelligibility derived from exceedingly sparse spectral information,'' Int. Conf. Spoken Lang. Proc., Sydney, Australia, December 1998.
Healy, E., and Bacon, S. P. (**2002**). ''Across frequency comparison of temporal speech information by listeners with normal and impaired hearing,'' J. Speech Lang. Hear. Res. **45**, 1262–1275.
Houtgast, T., and Steeneken, H. J. M. (**1985**). ''A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,'' J. Acoust. Soc. Am. **77**, 1069–1077.
Humes, L. E., and Roberts, L. (**1990**). ''Speech-recognition difficulties of the hearing-impaired elderly: contributions of audibility,'' J. Speech Hear. Res. **33**, 726–735.
Kasturi, K., Loizou, P. C., Dorman, M., and Spahr, T. (**2002**). ''The intelligibility of speech with ''holes'' in the spectrum,'' J. Acoust. Soc. Am. **112**, 1102–1111.
Lorenzi, C., Berthommier, F., Apoux, F., and Bacri, N. (**1999**). ''Effects of envelope expansion on speech recognition,'' Hear. Res. **136**, 131–138.
Lutman, M. E., Gatehouse, S., and Worthington, A. G. (**1991**). ''Frequency resolution as a function of hearing threshold level and age,'' J. Acoust. Soc. Am. **89**, 320–328.
Mehr, M. A., Turner, C. W., and Parkinson, A. (**2001**). ''Channel weights for speech recognition in cochlear implant users,'' J. Acoust. Soc. Am. **109**, 359–366.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford U. P., Oxford).
Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).
Peters, R. W., Moore, B. C. J., 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. (**1983**). ''The role of modulation in hearing,'' in *Hearing-Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, New York).
Rosen, S. (**1992**). ''Temporal information in speech: Acoustic, Auditory and Linguistic aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Silipo, R., Greenberg, S., and Arai, T. (**1999**). ''Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representation,'' in Proc. Eurospeech, Budapest, Hungary, September 1999.
Shannon, R. V., Galvin, J. J., and Baskent, D. (**2001**). ''Holes in hearing,'' J. Assoc. Res. Otolaryngol. **3**, 185–199.
Shannon, R. V., Zeng, F., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). ''Use of temporal envelope cues in speech recognition and hearing-impaired listeners,'' J. Acoust. Soc. Am. **97**, 2568–2576.
Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). ''Frequency-weighting functions for broadband speech as estimated by a correlational method,'' J. Acoust. Soc. Am. **104**, 1580–1585.
van der Horst, R., Rens Leeuw, A., and Dreschler, W. A. (**1999**). ''Importance of temporal-envelope cues in consonant recognition,'' J. Acoust. Soc. Am. **105**, 1801–1809.
van Tasell, D. J., Greenfield, D. G., Logemann, J. J., and Nelson, D. A. (**1992**). ''Temporal cues for consonant recognition: Training, talker generalization, and use in evaluation of cochlear implants,'' J. Acoust. Soc. Am. **92**, 1247–1257.
van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widen, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
Warren, R. M., and Bashford, J. A. (**1999**). ''Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband,'' J. Acoust. Soc. Am. **106**, 47–52.
Warren, R. M., Bashford, J. A., and Lenz, P. W. (**2000**). ''Intelligibility of bandpass speech: Effects of truncation or removal of transition band,'' J. Acoust. Soc. Am. **108**, 1264–1268.

Spectral loudness summation and simple reaction time[a)]

Eva Wagner and Mary Florentine[b)]
Institute for Hearing, Speech & Language and Department of Speech-Language Pathology & Audiology (106A FR), Northeastern University, Boston, Massachusetts 02115

Søren Buus and Joseph McCormack
Institute for Hearing, Speech & Language and Communications & Digital Signal Processing Center, Electrical & Computer Engineering Department (440 DA), Northeastern University, Boston, Massachusetts 02115

(Received 17 November 2003; revised 17 June 2004; accepted 18 June 2004)

The present study evaluates the relation between loudness and simple reaction time (RT). Loudness matches between a narrowband noise (125 Hz wide) and a broadband noise (1500 Hz) were made at levels from near threshold to near 100 dB SPL. Over a similarly wide range of levels, RT to each of the noise bands was also measured. As reported often in previous loudness-matching studies, loudness summation depended strongly on level. With increasing SPL, the level difference between the noises needed to keep them equally loud first increased, to around 10 dB at moderate levels, and then decreased. Except for one listener, the RT data show the same pattern. The level difference needed to keep RT to the two noises the same first increased and then decreased. These results show that RT is closely related to loudness, but not to sensation level. If RT depended on sensation level, the level difference between the two noises needed to achieve equal RT would not change with SPL because the difference in sensation level between two sounds is a constant. Overall, the average results provide strong support for the contention that simple RT and loudness are closely related. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1780573]

PACS numbers: 43.66.Cb, 43.66.Yw [GK] Pages: 1681–1686

I. INTRODUCTION

Early experiments by Wundt (1874), Piéron (1920), and Chocholle (1940) showed that simple reaction time (RT) decreases monotonically with increases in sound intensity. Chocholle (1940) was the first to investigate the relation between RT and sound intensity over a wide range of levels and frequencies. He found that RT decreases systematically as signal intensity increases and that equally loud tones produce equal RTs. Subsequent investigations have supported Chocholle's conclusion (e.g., Pfingst *et al.*, 1975; Scharf, 1978; Marshall and Brandt, 1980; Humes and Ahlstrom, 1984; Florentine *et al.*, 2004).

Almost all the measured RTs were to pure tones, and equally loud tones were often at the same sensation level (SL) (e.g., Pfingst *et al.*, 1975), leaving open the possibility that SL rather than loudness governs RT. Chocholle's (1940) data indicate that RT depends on loudness and not on SL. However, Chocholle himself noted that some of his listeners heard clicks at the signal onset at high intensities, which could have influenced his data. Kohfeld *et al.* (1981) in their partial replication of Chocholle's experiments found a less straightforward relationship between RT and loudness than did Chocholle. Their results suggested that over a wide range of frequencies, equally loud tones produced equal RTs at high levels (60 and 80 phons) but not at low levels (20 and 40 phons). That outcome is subject to question, because the equal-loudness contours reported by Kohfeld *et al.* (1981) differ substantially from those obtained in most other studies. Their deviant RTs at the low SLs probably reflected the fact that the tones were not equally loud, a result of methodological biases in their loudness matches. In a later study, Buus *et al.* (1982) obtained the same unusual loudness matches when they replicated the Kohfeld *et al.* method, but obtained ''standard'' loudness matches with two more conventional methods. Tones matched in loudness by these conventional methods yielded nearly identical RTs.

Stimuli whose loudness changes differently as a function of level can be used to examine the relationship between RT and loudness. Such a condition is readily achieved with noise bands of different bandwidths. Many studies have shown that the loudness of a complex sound varies with bandwidth (e.g., Fletcher and Munson, 1933; Zwicker and Feldtkeller, 1955; Zwicker *et al.*, 1957; Scharf, 1970; Florentine *et al.*, 1978). This effect, known as (spectral) loudness summation, depends strongly on level; the amount of loudness summation is greatest at moderate levels. Accordingly, the loudness functions for narrowband and broadband noises have quite different shapes and allow for a new way of evaluating the relation between RT and loudness as well as distinguishing between the effect of SL and loudness on RT.

II. METHOD

A. Listeners

Six listeners participated in this experiment. None had a history of hearing difficulties, and all had pure-tone audio-

[a)]Portions of this work were presented at the *27th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology in February 2004* (Wagner *et al.*, Abstracts of the 27th Annual Mid-Winter Meeting Assoc. Res. Otolaryngol, 2004).

[b)]Author to whom correspondence should be addressed. Electronic mail: florentin@neu.edu

metric thresholds at or below 15 dB HL (ANSI, 1996) at octave frequencies between 0.25 and 8 kHz. Their ages ranged between 19 and 31 years. Of the six listeners, all but L3 and L6 had previously participated in psychoacoustic experiments including L2, who is the first author. However, only L1 and L5 were experienced in RT tasks. Listeners listened with their preferred ear.

B. Stimuli

Stimuli, similar to those used by Florentine *et al.* (1978), were two bands of noise with an equivalent rectangular duration of 200 ms and linear rise and fall times of 6.67 ms. The stimulus bandwidths were 124 Hz (940–1064 Hz) and 1500 Hz (500–2000 Hz). The noises were generated in the frequency domain by setting the real and imaginary parts of all spectral lines within the pass band to random values chosen from a Gaussian distribution. The filter had a trapezoidal shape when plotted on a logarithmic scale with slopes of 70 dB/octave. New exemplars were generated for each presentation.

C. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli through a 16-bit D/A converter (TDT DD1) at a 41.67 kHz sample rate, recorded the listeners' responses, and executed the experimental procedure. The output of the D/A converter was attenuated (TDT PA4), lowpass filtered (TDT FT5, $f_c = 20\,\text{kHz}$, 135 dB/octave), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of the Sony MDR-V6 headset. The listeners were seated in a sound-attenuating booth with a response box, telegraph key, and headphones; all other equipment was positioned outside the booth.

For routine calibration, the output of the headphone amplifier was led to a 16-bit A/D converter (TDT DD1), such that the computer could sample the waveform, calculate its spectrum and rms voltage, and display the results before each block of trials. The SPLs reported below are based on the assumption of a frequency-independent output of 116 dB SPL for an input of 1 V rms.

D. Procedure

The procedure consisted of three parts. At the beginning, absolute thresholds were measured for each of the noises to obtain a reference for setting the sensation levels in the subsequent experiments. Next, several blocks of RT measurements were obtained. Finally, loudness matches were made between the two noises. The entire experiment was completed in two to four sessions depending on listener availability.

1. Absolute thresholds

Absolute thresholds were obtained with an adaptive procedure in a two-interval, two-alternative forced choice (2I, 2AFC) paradigm. Each trial contained two observation intervals, which were marked by lights and separated by 500 ms. The noise signal was presented in either the first or the second observation interval with equal *a priori* probability. The listener's task was to indicate which interval contained the signal by pressing the corresponding key on a response terminal. Two hundred milliseconds after the listener responded, the correct answer was indicated by a 200-ms light. Following the feedback, the next trial began after a 500-ms delay.

A single threshold measurement was based on three interleaved adaptive tracks. On each trial, one track was selected at random. For each track, the level of the signal was initially set approximately 15 dB above the listener's expected threshold, i.e., the initial presentation was always audible. It decreased after three consecutive correct responses and increased after one incorrect response. This procedure converged on the signal level yielding 79.4% correct responses (Levitt, 1971). The step size was 5 dB until the second reversal, after which it was 2 dB. Reversals occurred when successive signal levels changed direction from increasing to decreasing, or *vice versa*. Each track ended after five reversals. The threshold for one track was calculated as the average of the signal levels at the fourth and fifth reversals. One threshold measurement was taken as the average threshold across the three tracks.

Two such threshold measurements (for a total of six tracks) were obtained for each listener and each noise. Threshold measurements in which the standard error across the three tracks exceeded 2 dB were discarded and replaced. If acceptable measurements differed less than 2 dB, the average of the two thresholds was used as a reference to set the SL. Otherwise, a third threshold measurement was obtained and the average of all three thresholds was used.

2. Reaction-time measurements

Listeners pressed a telegraph key with their preferred hand as soon as they heard a noise stimulus. The RT was defined as the time elapsed between the onset of the stimulus and the response. The listener started each run by pressing a key on the response terminal. A trial started with a flashing LED, which marked the beginning of a random-duration foreperiod that was inserted between the response and the presentation of the next tone. The duration of the foreperiod was the sum of a fixed 1-s interval plus an exponentially distributed random variable with a mean of one second. When the listener pressed the telegraph key, lights flashed on the response terminal to indicate that the response had been recorded and that a new trial was starting. RTs were retained for analysis only if they were between 125 ms and 4 s after the onset of the stimulus. After the response or after 4 s, whichever came first, the trial was terminated. Responses occurring during the foreperiod or the first 125 ms of the stimulus were counted as false positives. If no response occurred within 4 s, the trial was counted as a miss, and the LED was flashed once to indicate to the listener that the stimulus was missed and that a new trial was starting. Missed trials were repeated twice at most. Each repeat presentation was preceded by a cue trial at the next higher noise level in an attempt to cue the listener to the signal.

Within each block of trials, 124 and 1500-Hz wide noise bands were presented five times at each level and bandwidth

in quasirandom order, not counting cue trials and repeats of missed presentations. The stimuli were presented at 0 and 3 dB SL and then in steps of 5 dB up to a maximum level of 100 dB SPL for the narrowband noise and a maximum level of 90 dB SPL for the broadband noise. Depending on threshold, 16 to 20 levels were tested. On each trial, the level and bandwidth were selected at random with the restriction that the SL could not change more than 30 dB from the previous trial if the bandwidth did not change and 25 dB if the bandwidth did change. Dummy trials at intermediate levels were added as necessary to ensure completion of all planned trials. Four blocks of trials were obtained for each listener, such that RTs generally were based on 20 responses for each bandwidth and sensation level.

3. Loudness matches

Loudness matches between the two noises were obtained in a two-interval, two-alternative forced choice (2I, 2AFC) paradigm. On each trial, the listener heard two stimuli separated by 600 ms. The fixed-level stimulus followed the variable-level stimulus or *vice versa* with equal *a priori* probability. The listener's task was to indicate which sound was louder by pressing a key on the response box. The response initiated the next trial after a 1-s delay. No feedback was provided. In a single block, both the narrowband and the broadband noise were presented each at five fixed levels. This resulted in ten interleaved tracks; in each track, one of the two noises was presented 20 times at a given fixed level. The track for each trial was selected at random from all those tracks that had not yet ended. Fixed levels for the narrowband stimuli were 5, 32, 57, 73, and 85 dB SL, and fixed levels for the broadband stimuli were 5, 25, 45, 65, and 85 dB SL. The first block served as training and was excluded from data analysis. Following training, three blocks of trials were obtained.

The level of the variable stimulus was determined by a Method of Maximum Likelihood (MML; for a general introduction to the MML procedure, see Formby *et al.*, 1996; for a historical review, see Florentine *et al.*, 2001). After each trial, the computer used a number of candidate psychometric functions to calculate the probability of obtaining the listener's responses to all of the stimuli that had been presented so far, given each psychometric function. Candidate functions were cumulative Gaussian functions with midpoints ranging in 2-dB steps from 30 dB below to 30 dB above the expected matching point. Standard deviations ranged from 2 to 32 dB separated by a factor $\sqrt{2}$. On each trial, the variable level was set to the midpoint plus or minus one standard deviation according to the most probable psychometric function. To avoid difficulties arising from the highly uncertain estimate of the standard deviation in the first few trials, the magnitude of the random derivation from the midpoint was fixed at 8 dB for the first four trials. This calculation was also performed at the end of each track and the midpoint of the most probable psychometric function was taken as the level at which the loudness of the variable stimulus equaled that of the fixed stimulus. At the end of one block, the midpoint and standard deviation were reestimated using MML with a set of more finely spaced candidate psychometric functions whose midpoints were spaced 0.2 dB apart and ranged from 3 dB below to 3 dB above the coarse estimate obtained during the matching procedure. Standard deviations ranged from 0.5 to 2 times the standard deviations estimated after the last trial with the original set of candidate functions in steps of a factor $\sqrt[4]{2}$.

E. Data analysis

1. Reaction-time experiments

Because RTs have highly skewed distributions, especially as they become slower near the threshold, the median RTs were used to represent the central tendency for each listener and stimulus. Data for 0 dB SL were excluded because too few responses were obtained to give reliable RTs. To examine the statistical significance of the effects of bandwidth and SL a three-way ANOVA for repeated measures with SL (17, continuous variable), bandwidth (2), and listeners (5) was performed. The log of the median RTs for individual listeners was used as the dependent variable. Only data for levels above 10 dB SL were included in this analysis to avoid undue influence from the sharply increasing RT as the level approaches threshold. To investigate whether learning might influence the effects of level and bandwidth, a three-way ANOVA [SL×presentation number (continuous variable)×listener] for repeated measures was performed on the logarithms of RTs obtained in individual repetitions. To avoid undue influence from the large variability at low SLs, the analysis was limited to data for SLs greater than 25 dB SL.

To summarize the effects of SL and bandwidth, data for each listener and bandwidth were fitted by a smooth function that described RT as a function of SL. Initially, Piéron's function (Piéron, 1914) was fit to the data,

$$\mathrm{RT} = \beta I^{-\alpha} + t_0. \tag{1}$$

In this function t_0 is an irreducible RT, which is reached at the highest levels, β is a free parameter, I is the intensity of the stimulus, and α is the exponent. However, the Piéron function fit the data poorly; it failed to capture the continued decline of RTs with increasing level well above the threshold. Adding a decreasing linear component to the function turned out to provide a better fit to the results:

$$\mathrm{RT} = \beta I^{-\alpha} + t_0 - aI. \tag{2}$$

In this modified Piéron function, a is the slope of the linear component. Finally, interpolations were made with the best-fitting functions to determine the level differences necessary to produce equal RTs for the two types of noises.

2. Loudness matches

For each listener and condition, the three equal-loudness levels were averaged, resulting in two sets of data: the averages of the tracks with the narrowband stimulus fixed and the averages of the tracks with the broadband stimulus fixed. Polynomial fits were then made to the combined data. Visual inspection showed that a third-order polynomial provided the best fit to the data.

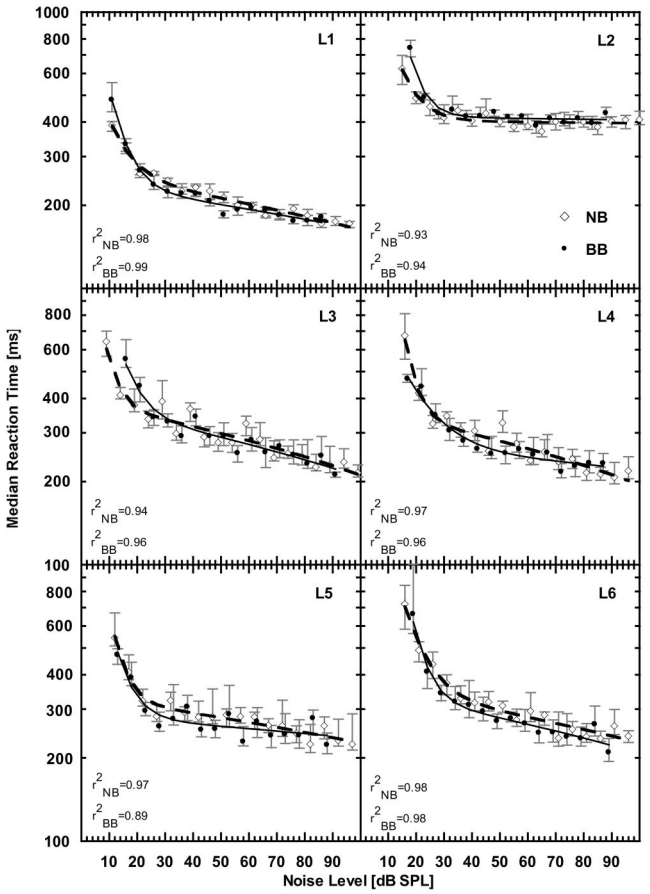


FIG. 1. Individual data from six listeners: Median RTs as a function of level for the two types of noise. The vertical bars show the semi-interquartile ranges. Modified Piéron functions are fitted to the data. R^2 values are shown as a measure of the goodness of the fit.

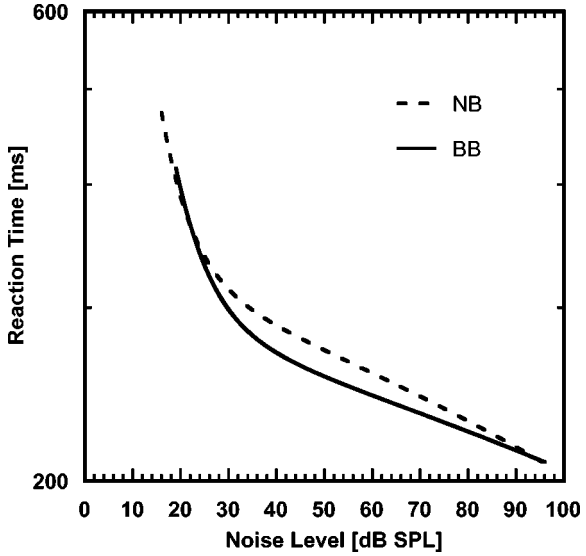


FIG. 2. Geometric averages of the functions fitted to the individual RTs as a function of level for five of the six listeners shown in Fig. 1.

III. RESULTS AND DISCUSSION

Figure 1 shows individual RT data for six listeners. In general, RTs decrease with increasing SPL. Although there are individual differences, five of the six listeners show a similar trend: At moderate levels, RTs tend to be shorter for the broadband noise than for the narrowband noise. For L3 this trend is less distinct than for the four other listeners due to a higher variability in the data, but a difference in RTs, although small, is shown by the curve fits to the data. One listener (L2) shows no consistent difference except at low levels and around 50 dB SPL, where the RTs are shorter for the narrowband noise than for the broadband noise. A replication of the RT measurements with this listener produced similar aberrant results. Since L2 is obviously different from the other listeners, these data are not included in further analysis.

Figure 2 shows the geometric means of the fitted functions for the other five listeners. The same pattern as in the individual data is apparent: At moderate levels, RTs for the broadband noise are shorter than for the narrowband noise. The ANOVA (SL $\times$ bandwidth $\times$ listener) showed statistically significant effects for SL ($F=82.3$, $p=0.0008$), bandwidth ($F=55.81$, $p=0.0017$), and their interaction ($F=12.85$, $p=0.0231$). Differences among listeners are highly significant ($F=27.3$, $p<0.0001$), as was the interaction between SL and listener ($F=66.22$, $p=0.0001$). The interaction between bandwidth and listener was not statistically significant. Presentation number did not have a significant effect nor did it interact significantly with the two stimulus variables (SL and bandwidth). This indicates that practice did not appreciably influence the effects of bandwidth and SL on RT.

We turn now to the results of the loudness matches. Figure 3 shows individual data for six listeners. The level difference between the narrowband noise and the broadband noise at which they were judged equal in loudness is plotted as a function of the level of the broadband noise. The filled circles show data obtained with the level of the narrowband noise varied, the unfilled circles show data obtained with the level of the broadband noise varied, and the solid line is a polynomial fit to the data points. Owing to the way in which loudness summation depends on level, the level difference is greatest at moderate levels and smallest at low and high levels. There are individual differences, e.g., in the level at which the maximum amount of loudness summation occurs. The amount of variability is not uncommon for loudness measurements. The mean measure helps reduce fluctuations that may be due to individual differences in loudness matches. Figure 4 shows the average of five listeners. These results agree well with those in the literature on spectral loudness summation (e.g., Fletcher and Munson, 1933; Zwicker and Feldtkeller, 1955; Zwicker *et al.*, 1957; Scharf, 1970; Florentine *et al.*, 1978).

To compare the RT data with the loudness data, we need to determine the level difference between the two noises that yields the same RT. This difference is readily ascertained by taking the horizontal differences between the two RT functions in Fig. 2. The resulting differences are plotted as a function of the SPL of the broadband noise and are represented by a dashed line in Fig. 5. Also plotted in Fig. 5, as the solid line (taken from Fig. 4), is the difference in level required to yield equal loudness between the two noises. The good agreement between the two curves reflects the close relationship between loudness and RT; the same difference in

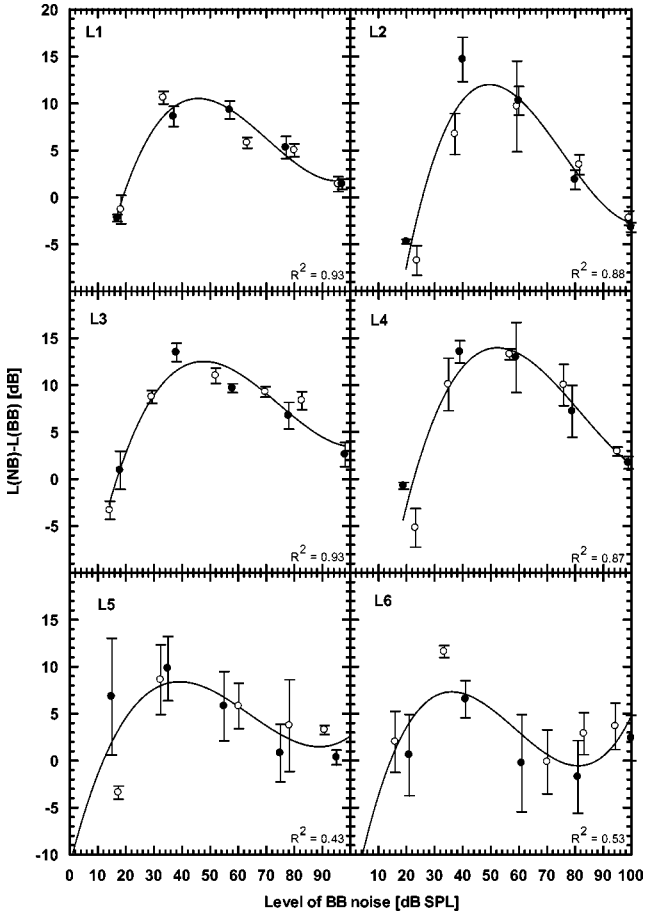


FIG. 3. Individual data from six listeners: Level difference between the narrowband noise and the broadband noise at which they were judged equal in loudness plotted as a function of the level of the broadband noise. The filled circles indicate that the level of the narrowband noise was varied, the unfilled circles indicate that the level of the broadband noise was varied. The vertical bars show the standard errors of the mean. The solid line is a polynomial fit to the data. R^2 values are shown as a measure of the goodness of the fit.

level between the two noises that yields equal loudness also yields equal RT. Especially compelling is the fact that the level difference varies substantially with SPL in the same way for loudness and for RT. However, there is a discrepancy between the two curves at low levels. RT measurements are more uncertain at low levels and the variability between listeners is larger. Therefore, it is difficult to assess if a real difference exists based on five listeners. The dotted line in Fig. 5 shows the level difference (average of 2 dB for the five listeners) between the two noises if presented at the same SL. It indicates that if RT depended on SL, the level difference required to achieve equal RT would not change with SPL because the difference between the SL of two sounds is a constant.

Good correspondences between equal-loudness and equal-latency data have been also found in studies using pure-tone stimuli (Chocholle, 1940; Pfingst *et al.*, 1975; Marshall and Brandt, 1980). One study that used noise stimuli (Humes and Ahlstrom, 1984) measured loudness (by magnitude estimation) and RT as a function of level for one-third octave bands of noise centered at 1, 2, and 4 kHz. They found that listeners with the steepest RT functions had the steepest loudness functions, leading them to conclude that RT provides an indirect estimate of loudness. Arieh and Marks (2003) demonstrated the relation between loudness and RT in a study of induced loudness reduction (ILR), i.e., the decrease of the loudness of a tone when preceded by a louder tone of the same frequency. The amount of loudness reduction was the same whether measured by magnitude estimation or by RT.

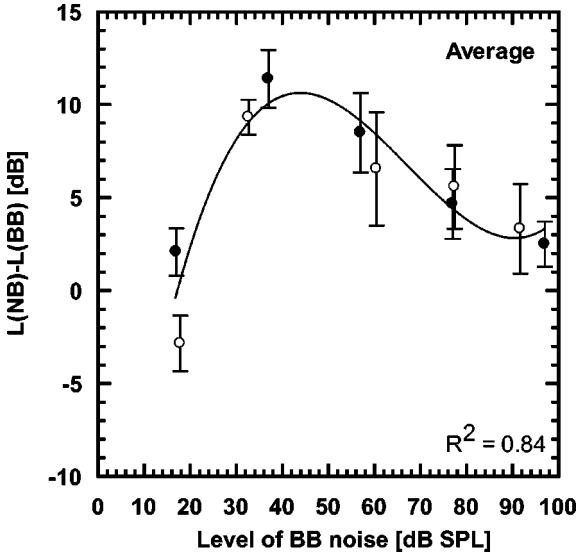


FIG. 4. Averaged data of five of the six listeners shown in Fig. 3.

Altogether, the present data support the contention that, at least for young normal listeners, loudness and RT are closely related. Moreover, the data allow a clear dissociation between the influence of loudness and SL on RT. Although it remains to be seen how well the relation between loudness and RT generalizes to populations other than young normal listeners, the RT paradigm can be of advantage compared to other methods of measuring loudness. It is a simple task that

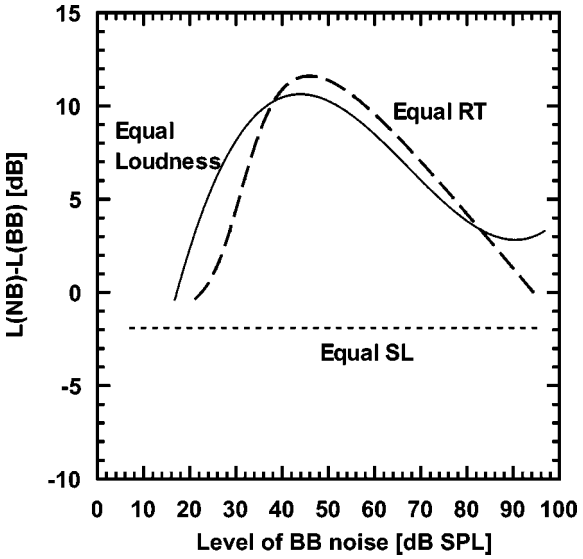


FIG. 5. Averaged data for five listeners. The level difference between the two noises that yields the same RTs (dashed line) or equal loudness (solid line) is plotted as a function of the SPL of the broadband noise. The dotted line indicates the level difference between the two noises if presented at the same SL.

does not require complicated responses or decisions. Hence, it can and has been used to investigate loudness perception in normal-hearing infants (Leibold and Werner, 2002) and in animals (e.g., Stebbins, 1966; Pfingst *et al.*, 1975). Animal models can be developed, which facilitate comparisons among species and between psychophysical and physiological data. Such comparisons should provide a deeper insight into the basic mechanisms underlying loudness perception.

ACKNOWLEDGMENTS

The authors thank Bertram Scharf, Brian Moore, and an anonymous reviewer for helpful comments. Thanks are also due to Michael Stahl for critically reading a draft of the paper. Søren Buus contributed substantially to this work, unfortunately he passed away prior to the submission of the revision. This research was supported by NIH/NIDCD Grant No. R01 DC 02241.

Arieh, Y., and Marks, L. E. (**2003**). ''Recalibrating the auditory system: a speed-accuracy analysis of intensity perception,'' J. Exp. Psychol. Hum. Percept. Perform. **29**, 523–536.

ANSI (**1996**). ANSI S3.6-1996, *Specification for Audiometers* (American National Standards Institute, New York).

Buus, S., Greenbaum, H., and Scharf, B. (**1982**). ''Measurements of equal loudness and reaction times,'' J. Acoust. Soc. Am. **72**, Suppl. 1, 594.

Chocholle, R. (**1940**). ''Variation des temps de réaction auditif en fonction de l'intensité à diverses fréquences,'' (''Variation in auditory reaction time as a function of intensity at various frequencies''), Annee Psychol. **41**, 65–124.

Fletcher, H. F., and Munson, W. A. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Florentine, M., Buus, S., and Bonding, P. (**1978**). ''Loudness of complex sounds as a function of the standard stimulus and the number of components,'' J. Acoust. Soc. Am. **64**, 1036–1040.

Florentine, M., Marvit, P., and Buus, S. (**2001**). ''Maximum-likelihood yes–no procedure for gap detection: effect of track length,'' J. Am. Acad. Audiol **12**, 113–120.

Florentine, M., Buus, S., and Rosenberg, M. (**2004**). ''Reaction-time data support the existence of softness imperception in cochlear hearing loss,'' in Auditory Signal Processing: Physiology, Psychoacoustics, and Models, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer-Verlag, New York).

Formby, C., Sherlock, L. P., and Green, D. M. (**1996**). ''Evaluation of a maximum-likelihood procedure for measuring pure-tone thresholds under computer control,'' J. Am. Acad. Audiol **7**, 125–129.

Humes, L. E., and Ahlstrom, J. B. (**1984**). ''Relation between reaction time and loudness,'' J. Speech Hear. Res. **27**, 306–310.

Kohfeld, D. L., Santee, J. L., and Wallace, N. D. (**1981**). ''Loudness and reaction time: I,'' Percept. Psychophys. **29**, 535–549.

Leibold, L. J., and Werner, L. A. (**2002**). ''Relationship between intensity and reaction time in normal-hearing infants and adults,'' Ear Hear. **23**, 92–97.

Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustics,'' J. Acoust. Soc. Am. **49**, 467–477.

Marshall, L., and Brandt, J. F. (**1980**). ''The relationship between loudness and reaction time in normal hearing listeners,'' Acta Oto-Laryngol. **90**, 244–249.

Pfingst, B. E., Hienz, R., Kimm, J., and Miller, J. (**1975**). ''Reaction-time procedure for measurement of hearing. I. Suprathreshold functions,'' J. Acoust. Soc. Am. **57**, 421–430.

Piéron, H. (**1914**). ''Recherches sur les lois de variation des temps de latence sensorielle en fonction des intensités excitatrices,'' Annee Psychol. **20**, 17–96.

Piéron, H. (**1920**). ''Nouvelles recherches sur l'analyse du temps de latence sensorielle en fonction des intensités excitatrices,'' Annee Psychol. **22**, 58–142.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundation of Modern Auditory Theory. 1*, edited by J. V. Tobias (Academic, New York).

Scharf, B. (**1978**). ''Loudness,'' in *Handbook of Perception. 4*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Stebbins, W. C. (**1966**). ''Auditory reaction time and the derivation of equal loudness contours for the monkey,'' J. Exp. Anal Behav. **9**, 135–142.

Wundt, W. (**1874**). *Grundzüge der Physiologischen Psychologie*, 2nd ed. (Engelman, Leipzig).

Zwicker, E., and Feldtkeller, R. (**1955**). ''Über die Lautstärke von gleichförmigen Geräuschen,'' Acustica **5**, 303–316.

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical band width in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

Masked auditory thresholds in sciaenid fishes: A comparative study

John Ramcharitar[a] and Arthur N. Popper
Department of Biology and Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland 20742

(Received 2 December 2003; revised 19 May 2004; accepted 20 May 2004)

Western Atlantic sciaenids comprise a taxonomically diverse teleost family with significant variations in the relationship between the swim bladder and the otic capsule. In this study, the auditory brainstem response (ABR) was used to test the hypothesis that fishes with different peripheral auditory structures (black drum, *Pogonias chromis* and Atlantic croaker, *Micropogonias undulatus*) show differences in frequency selectivity. In a black drum the swim bladder is relatively distant from the otic capsule while the swim bladder in Atlantic croaker possesses anteriorly-directed diverticulas that terminate relatively near the otic capsule. Signals were pure tones in the frequency range, 100 Hz to 1.5 kHz, and thresholds were determined both with and without the presence of simultaneous white noise at two intensity levels (124 dB and 136 dB, *re*: 1 μPa). At the 124 dB level of white noise background, both the black drum and Atlantic croaker showed similar changes in auditory sensitivity. However, in the presence of the 136 dB white noise masker, black drum showed significantly greater shifts in auditory thresholds between 300 and 600 Hz. The results indicate that the two species differ in frequency selectivity since the Atlantic croaker was less susceptible to auditory threshold shifts, particularly at the higher level of masking. This difference may be linked to peripheral auditory mechanisms. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1771614]

PACS numbers: 43.66.Dc, 43.64.Qh, 43.64.Wn Pages: 1687–1691

I. INTRODUCTION

Auditory masking studies have revealed that teleost fishes process acoustic stimuli in ways similar to those of other vertebrate groups including birds and mammals (reviewed in Fay, 1988). These similarities include the following: sound detection by ''auditory filters,'' masking release when the interfering sound is separated from the signal either spatially or in frequency, the ability to perform frequency and intensity discrimination, and the presence of critical bands (Buerkle, 1969; Chapman and Hawkins, 1973; Tavolga, 1974; Fay *et al.*, 1978; Coombs and Popper, 1979; Fay, 1988). As such, fish models are potentially useful for understanding the relationship between form and function in the vertebrate inner ear (Fay and Popper, 2000; Lanford *et al.*, 2000, Ramcharitar *et al.*, 2002, 2004).

Frequency selectivity (differential detection of signals of various frequencies) can be determined by investigating the effects of maskers on auditory thresholds across an array of test frequencies (McGee *et al.*, 1976; Wightman *et al.*, 1977; Pickles, 1979). Psychophysical tuning curves (PTCs) may then be constructed. This approach has been successfully applied to fishes (Fay *et al.*, 1978; Popper and Coombs, 1982) where comparative PTC studies have demonstrated interspecific differences in frequency selectivity among several teleost species (Popper and Coombs, 1982; McCormick and Popper, 1984). Two fishes, goldfish (*Carassius auratus*) and clown knifefish (*Notopterus*), both of which have connections between the swim bladder and the inner ear, show PTCs (for a 500 Hz signal) that are sharply tuned and have multiple-peaks. On the other hand, a species with no such swim bladder-inner ear coupling, the squirrelfish *Adioryx*, has a broadly tuned, unipeaked PTC for the same test signal (Popper and Coombs, 1982). Specializations of the peripheral auditory system may therefore confer frequency selectivity. In order to test this hypothesis, masking in members of the teleost family sciaenidae that have known differences in the swim bladder–inner ear relationship was examined.

Western Atlantic sciaenids are part of a taxonomically diverse and economically important teleost family. These are largely vocal species that show substantial interspecific differences in the structure and function of their auditory systems (Chao, 1978; Davorec, 1983; Ramcharitar *et al.*, 2002). All sciaenids exhibit enlargement of at least one otolith (relative to those for the vast majority of teleost species for which data are available) and there is significant variation in the relationship between the swim bladder, and the inner ear in different species (Chao, 1978; Ramcharitar, 2002).

In this study the auditory brainstem response (ABR) was used to investigate the effects of white noise on hearing sensitivity in two species of sciaenids which have different swim bladder–inner ear relationships, the black drum (*Pogonias chromis*) and the Atlantic croaker (*Micropogonias undulatus*). In the black drum, the swim bladder is relatively distant from the otic capsule, while in the Atlantic croaker the swim bladder possesses anteriorly-directed projections that terminate near the inner ear (Chao, 1978; Davorec, 1983; Ramcharitar *et al.*, 2002).

Based upon the PTC data from fishes with different

[a]Current address: Department of Psychological and Brain Sciences, Johns Hopkins University, Ames Hall, 3400 N. Charles Street, Baltimore, Maryland, MD 21218. Electronic mail: jramch@jhu.edu

swim bladder–inner ear configurations, it was hypothesized that the Atlantic croaker will be better able to detect signals in the presence of a background of white noise when compared to the black drum.

II. MATERIALS AND METHODS

Animals used in this study were supplied by the Public Service Enterprise Group, NJ. Fishes were kept in captivity in large cylindrical tanks (889.5 liters, diameter = 121.9 cm). Water was filtered and aerated, and fishes were fed daily. Fifteen fish per species were used in this study, which was approved by the University of Maryland Institutional Animal Care and Use Committee. All fishes used were 11.5 ± 2.5 cm in total length (TL).

A. Gross morphology

Five fishes per species were euthanized using buffered Ethyl 3-aminobenzoate methanesulfonate (MS-222, Sigma) and ventral dissections were performed to expose the swim bladder and otic capsule. Digital images of the inner ear, otoliths, and swim bladder, as well as of the relationship between the swim bladder and otic capsule were then taken using a Nikon Coolpix 4500 camera.

B. Auditory brainstem response

Prior to starting experiments, the position of the brainstem was determined using gross dissections. Three fish were euthanized and dorsal dissections were performed to expose the brain and rostral region of the spinal cord. Once the brainstem was mapped, landmarks on the external surface of the fishes were identified in order to make electrode placement consistent between ABR experiments.

C. Tank set-up

Fish were suspended using a mesh sling into a 19 L plastic bucket filled with water. They were positioned about 35 cm above an underwater speaker (UW-30, Underwater Sound Inc., Oklahoma City, OK) and fully submerged so that the top of the head was 5 cm below the water surface. A reference electrode was placed on the dorsal surface of the fish's head, medial to, and just behind, the eyes. A recording electrode was placed on the dorsal surface of the fish just over the brainstem. Both of these electrodes were inserted just beneath the skin to improve the signal-to-noise ratio. A ground electrode was placed in the water near the body of the fish. The electrodes used were stainless steel (Rochester Electro-Medical Inc., Tampa, FL, USA) and insulated with epoxy, except for the very tip (1 mm).

D. Stimulus generation and signal acquisition

Sounds were played and responses collected using a Tucker-Davis Technologies Inc. (TDT; Gainesville, FL) physiological apparatus using SigGen and BioSig software (TDT). Stimuli were generated by a computer with the TDT software through a power amplifier connected to a UW-30 underwater speaker. Calibration of the acoustic output to the test tank was performed by analyzing the speaker output at the position of the fish at each frequency using a LC-10 hydrophone (calibration sensitivity of −208.6 dB *re*: 1 V/μPa; ±3 dB, 0.1-180 kHz, omni-directional), attached to a digital oscilloscope. Additionally, a calibrated hydrophone (InterOcean Model 902 Acoustic Listening and Calibration System) was used to monitor speaker output.

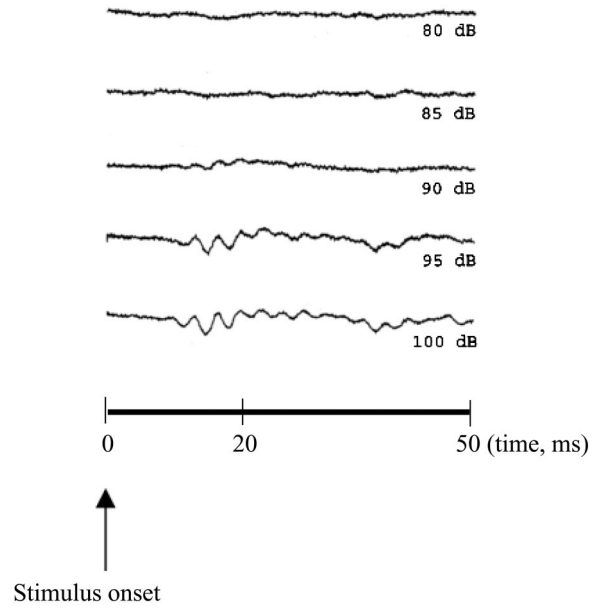


FIG. 1. Typical ABR traces. These are averaged traces for the black drum at a test frequency of 100 Hz in the absence of white noise. The auditory threshold in this case is judged to be 90 dB *re*: 1 μPa (as determined by the emergence of negative downward peaks within the first 20 seconds of stimulus onset). The sound level is specified for each trace (80–100 dB *re*: 1 μPa).

Test signals were pure tones in the frequency range, 100–1500 Hz. Broadband white noise at two test levels were used simultaneously with test tones in order to determine masked thresholds. The masker levels were 124.3 dB and 136.2 dB (*re*: 1 μPa), and were generated by a TDT RP2.1 unit. These maskers were selected as they represent the limits of white noise levels that could be generated in our experimental set-up. Tones were 10 ms in duration with a 2 ms rise and fall time. Data were acquired over a 50 ms period starting with test tone onset. Sound intensity levels at each frequency were increased in 5 dB increments from 60 dB to 125 dB, until a typical ABR waveform (series of negative peaks within the first 20 ms of stimulus onset, Fig. 1) was observed, or no detectable response. A total of 400 responses were averaged for each signal tested.

E. Controls

Throughout the experimental testing the ABR system was calibrated using goldfish, a species which has been studied extensively and for which there are established audiograms (reviewed in Fay, 1988). In each test case, expected audiograms were achieved for goldfish, thereby confirming that our ABR methods produced data that were consistent with other experimental data. Experimental controls were established using euthanized fish or no fish in the setup in order to establish that the ABR responses were not artifacts. These tests also confirmed that our data were reliable.

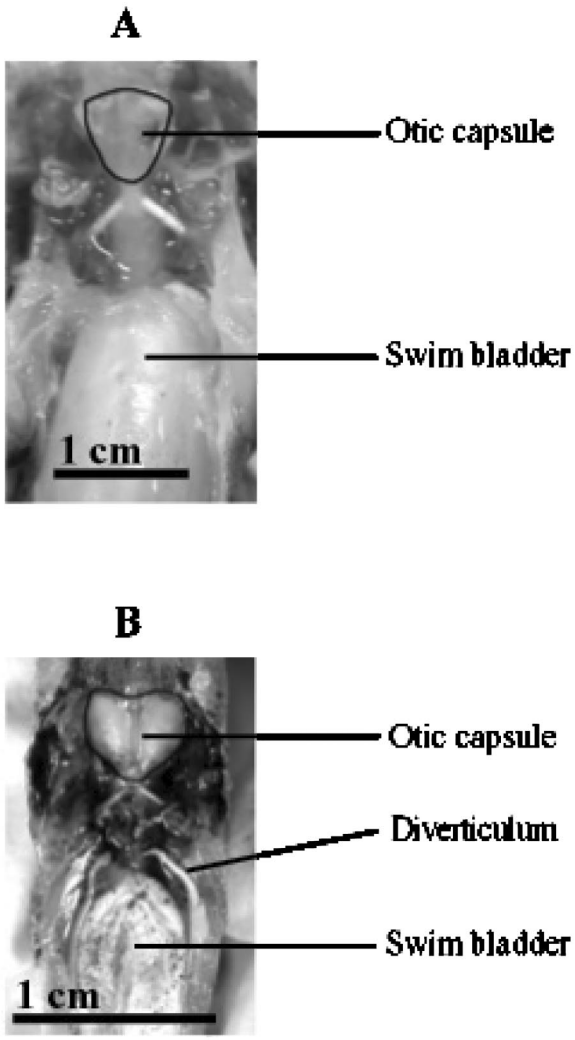


FIG. 2. Swim bladder–otic capsule relationship in the black drum (A) and the Atlantic croaker (B). These are digital pictures of ventral dissections with the otic capsules outlined in black (dorsal to top of page). The swim bladder of the black drum has no projections and it is relatively distant from the otic capsule, while that of the Atlantic croaker has extensions (diverticulas) that terminate relatively close to the otic capsule. The black drum was 16 cm (TL) and the Atlantic croaker was 13 cm (TL).

F. Statistical analysis

All statistical analyses were done as analyses of variance (ANOVA) with the Tukey posthoc test when significant differences were found (Zar, 1984). For all tests, $\alpha = 0.05$ was used as the significance level.

III. RESULTS

A. Swim bladder–inner ear relationship

Figure 2 shows the relationship between the swim bladder and otic capsule in an Atlantic croaker and a black drum. An Atlantic croaker has two lateral elongate diverticulas that project anteriorly and end in close proximity (3.81 ± 0.21 mm, mean TL $\pm$ standard deviation in the specimens of the size studied) to the posterior edge of the otic capsule. The swim bladder of the black drum does not possess these morphological specializations and the minimum distance between the swim bladder and caudal edge of the otic capsule was 7.42 ± 0.86 mm.

B. Auditory thresholds

The unmasked audiograms of the Atlantic croaker and the black drum are presented in Fig. 3. The Atlantic croaker detected frequencies from 100 Hz to 1000 Hz, while black drum responded to sounds from 100 Hz–800 Hz. There were no significant differences ($p > 0.05$) in hearing sensitivity between the species from 100 Hz to 800 Hz. While the best frequency for the Atlantic croaker was 300 Hz, black drum was most sensitive across the range, 100–500 Hz.

C. Auditory threshold shifts

There were statistically significant inter-specific differences in the effects of simultaneous white noise on frequency range of detection, as well as on auditory thresholds [Figs. 4(a), 4(b)]. In the presence of the white noise maskers, the black drum was no longer able to detect signals at the high frequency end of its detection range. At the lower masker level (124 dB *re*: 1 μ Pa), there were no differences in audi-

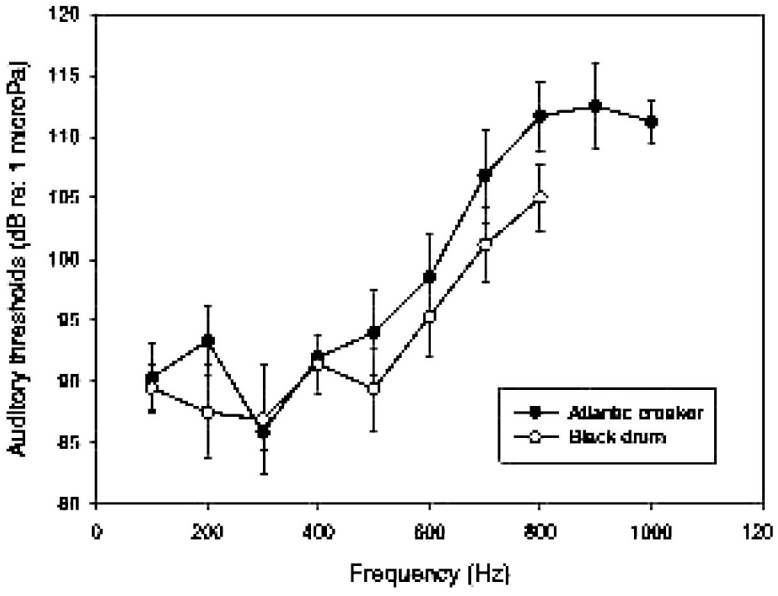


FIG. 3. Audiograms (unmasked) for the black drum and the Atlantic croaker. Auditory thresholds expressed as "mean $\pm$ standard deviation."

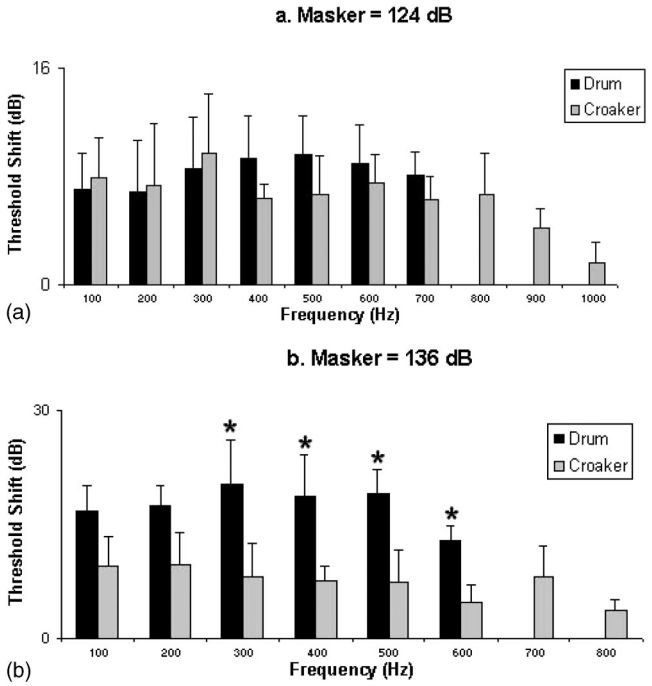


FIG. 4. Auditory threshold shifts. (a). The relationship between the auditory threshold shift and frequency for the black drum and the Atlantic croaker at the 124 dB (*re*: 1 μPa) masker level. There are no significant differences for test frequencies 100–700 Hz. (b). The relationship between the auditory threshold shift and frequency at the 136 dB (*re*: 1 μPa) masker level. The black drum showed significantly higher shifts for the test frequencies 300–600 Hz. Auditory thresholds are expressed as ''mean ± standard deviation.''

tory threshold shifts between the two species [Fig. 4(a)]. At 136 dB (*re*: 1 μ Pa), the black drum showed significantly greater shifts in auditory thresholds compared to the Atlantic croaker, particularly in the frequency range 300–600 Hz [Fig. 4(b)].

IV. DISCUSSION

Statistically significant inter-specific differences in masked auditory thresholds were demonstrated for two sciaenid species with different swim bladder–inner ear relationships. The black drum (the swim bladder distant from the ear) was generally more susceptible to the effects of masking than the Atlantic croaker (the swim bladder extensions terminate near the ear), as it showed a reduction in the frequency detection range (with respect to the speaker output limit) and a relatively uniform loss of sensitivity across test signals. These results support the hypothesis that fishes with different peripheral auditory systems also tend to show variations in frequency selectivity.

Psychophysical tuning curves (PTCs) for several teleosts provide evidence for interspecific differences in frequency selectivity that are correlated with the morphology of the peripheral auditory system (Popper and Coombs, 1982; McCormick and Popper, 1984). Fishes can be classified as hearing ''generalists'' or ''specialists'' based on the presence (or absence) of accessory auditory organs (e.g., swim bladders and auditory bullas) that enhance audition (Popper and Fay, 1999). Three species that possess swim bladders that are connected with the inner ear (*Carassius*, *Notopterus*, and *Gnathonemus*) have PTCs that are sharply tuned and multi-peaked, while the PTC for a species lacking any such connection (*Adioryx*) is broadly tuned and unipeaked. Sharp tuning of the PTCs of the former species, together with their multiple peaks, indicate that these fishes may be detecting frequencies differentially across the range of test signals. Conversely, the unipeaked, broadly tuned PTC of *Adioryx* suggests a relatively low degree of frequency selectivity. The masking data presented for sciaenids in this study also suggest that hearing specialists may tend to be frequency selective when compared to generalists.

The mechanisms underlying frequency selectivity in fishes are largely unresolved and may include both peripheral and central components (Fay *et al.*, 1978; McKibben and Bass, 1999; Weeg *et al.*, 2002). Even given the interspecific variation in shape and size of sagittas, the audiograms of these two species show that in the absence of a white noise background there are no significant differences in hearing sensitivity up to 800 Hz. According to the models proposed by Lychakov and Rebane (1993, 2000), sensitivity differences resulting from otolith morphology should be evident in the low frequency range. Also, studies on the acoustic properties of swim bladders have revealed that they are heavily damped (low Q) with relatively broad responses over the range to which fishes are responsive (Popper, 1974; Sand and Hawkins, 1973). Assuming that this is true for the Atlantic croaker, its swim bladder morphology may not then account for trends observed in its masked thresholds. There is however some evidence to suggest that swim bladder morphology affects its acoustic properties (Schellart and Popper, 1992), and given that the Atlantic croaker possesses a swim bladder with elongated lateral diverticulas (Chao, 1978; Davorec, 1983), further analysis (e.g., using laser vibrometry) is needed to determine whether or not this organ may be involved in frequency selectivity.

Differences in masking in the two sciaenids may be due to differences in responses to the two components of acoustic stimuli, particle motion and pressure. Fish otolithic end organs are responsive to particle motion up to several hundred Hertz (Fay, 1984; Fay and Simmons, 1999), while the detection of higher frequency pressure signals may be mediated via the swim bladder and other gas bladders (van Bergeijk, 1967). In a pressure field, a gas bladder can re-radiate acoustic energy as particle displacements which can then potentially stimulate the inner ear if it is in close proximity to the otic capsule (Blaxter *et al.*, 1981). As such, the Atlantic croaker may be sensitive to both acoustic particle motion and pressure on account of its relatively close swim bladder–inner ear relationship, while the black drum may be responsive only to particle displacement fields (as its swim bladder is distant from the otic capsule). The Atlantic croaker was relatively less susceptible to masking, particularly at the higher white noise level and at frequencies greater than 200 Hz. It is therefore possible that acoustic pressure detection in the Atlantic croaker correlates with its masking properties, and a study of particle motion versus pressure detection can help to determine if this is the case. While it may be argued that the differences in masking in the two species are a function of the acoustic set-up in our experiments, both species were tested under identical acoustic parameters, and so dif-

ferences between the species can only be explained by the biology of the auditory apparatus and not by tank acoustics.

The Atlantic croaker demonstrated a relatively strong ability to detect signals between 300 and 600 Hz. The biological significance of this may be related to the detection of conspecific vocalizations, given high levels of ambient noise. A spectral analysis of vocalizations, coupled with physiological investigations of neuronal response properties, could lead to an explanation of the pattern of frequency selectivity observed in an Atlantic croaker. Feature detecting auditory neurons have been demonstrated for several other vocal teleosts (e.g., Crawford, 1993; McKibben and Bass, 1999). An analysis of ambient noise levels in the habitats of the Atlantic croaker and the black drum may also helpful in understanding the behavioral relevance of interspecific differences observed in their masked auditory thresholds. Differences in frequency selectivity may also be linked to development (e.g., reproductive condition, ontogeny of peripheral auditory structures).

This study demonstrated a correlation between morphology of the peripheral auditory system and frequency selectivity. A further comparative study of masking in sciaenids is needed to derive potential structural correlates of frequency selectivity in this taxonomically diverse family. The mechanisms that account for differential frequency detection abilities can be further explored by PTC analysis (Fay *et al.*, 1978; Popper and Coombs, 1982; McCormick and Popper, 1984), as well as with the use of band narrowing maskers which have revealed the existence of critical bands for several teleost species (Buerkle, 1969; Tavolga, 1974).

ACKNOWLEDGMENTS

This work was supported by the Public Service Enterprise Group, a New Jersey-based diversified energy company. Words of appreciation go to Dr. R. R. Fay for his comments and advice. Special thanks to members of the Popper Lab (UMD) and to Ferial S. Khan for their critical reviews and assistance.

Blaxter, J. H. S., Denton, E. J., and Gray, J. A. B. (**1981**). ''Acousticolateralis system in cluepid fishes,'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 39–53.

Buerkle, U. (**1969**). ''Auditory masking and the critical band in Atlantic Cod (*Gadus morhua*),'' J. Fish Res. Bd. Canada **26**, 1113–1119.

Chao, L. N. (**1978**). ''A basis for classifying western Atlantic Sciaenidae,'' NOAA Tech. Rep. Cir. 415.

Chapman, C. J., and Hawkins, A. D. (**1973**). ''A field study of hearing in the Cod (*Gadus morhua*), L,'' J. Comp. Physiol. **85**, 147–167.

Coombs, S., and Popper, A. N. (**1979**). ''Hearing differences among Hawaiian squirrelfish related to differences in the peripheral auditory system,'' J. Comp. Physiol. [A] **132**, 203–207.

Crawford, J. D. (**1993**). ''Central auditory neurophysiology of sound producing mormyrid fish: The mesencephalon of *Pollimyrus isidori*,'' J. Comp. Physiol. A **172**, 1–14.

Davorec, J. E. (**1983**). ''Sciaenid fishes of Western Peninsula Florida,'' Fl Dept. Nat. Res. MRL, St. Petersburg, Florida.

Fay, R. R. (**1984**). ''The Goldfish ear codes the axis of acoustic particle motion detection in three dimensions,'' Science **225**, 951–954.

Fay, R. R. (**1988**). ''Hearing in vertebrates: A psychophysics databook,'' Hill-Fay Associates, Winnetka, IL, pp. 26–156.

Fay, R. R., Ahroon, W., and Orawski, A. (**1978**). ''Auditory masking patterns in the goldfish (*Carassius auratus*): Psychophysical tuning curves,'' J. Exp. Biol. **74**, 83–100.

Fay, R. R., and Popper, A. N. (**2000**). ''Evolution of hearing in vertebrates: The inner ears and processing,'' Hear. Res. **149**, 1–10.

Fay, R. R., and Simmons, M. (**1999**). ''The sense of hearing in fishes and amphibians,'' in *Comparative Hearing: Fish and Amphibians*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 269–318.

Lanford, P. J., Platt, C., and Popper, A. N. (**2000**). ''Structure and function in the saccule of the goldfish (*Carassius auratus*): A model of diversity in the non-amniote ear,'' Hear. Res. **143**, 1–13.

Lychakov, D. V., and Rebane, Y. T. (**1993**). ''Effect of otolith shape on directional sound perception in fish,'' J. Evol. Biol. **28**, 531–536.

Lychakov, D. V., and Rebane, Y. T. (**2000**). ''Otolith regularities,'' Hear. Res. **143**, 83–102.

McCormick, C. A., and Popper, A. N. (**1984**). ''Auditory sensitivity and psychophysical tuning curves in the elephant nose fish (*Gnathonemus petersii*), J. Comp. Physiol. A **155**, 753-761.

McGee, T., Ryan, A., and Dallos, P. (**1976**). ''Psychophysical tuning curves of chinchillas,'' J. Acoust. Soc. Am. **60**, 1146–1150.

McKibben, J. R., and Bass, A. H. (**1999**). ''Peripheral encoding of behaviorially relevant acoustic signals in a vocal fish: Single tones,'' JCPA **184**, 563–576.

Pickles, J. O. (**1979**). ''Psychophysical frequency resolution in the cat as determined by simultaneous masking and its relation to auditory nerve resolution,'' J. Acoust. Soc. Am. **66**, 1725–1732.

Popper, A. N. (**1974**). ''The response of the swim bladder of the goldfish (*Carassius auratus*), J. Exp. Biol. **60**, 295–304.

Popper, A. N., and Coombs, S. (**1982**). ''The morphology and evolution of the ear in actinopterygian fishes,'' Am. Zool. **22**, 311–328.

Popper, A. N., and Fay, R. R. (**1999**). ''The auditory periphery in fishes,'' in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 44–100.

Ramcharitar, J. (**2002**). ''Unique ear Structure of silver perch *Bairdiella chrysoura*,'' Bioacoustics **12**, 122–124.

Ramcharitar, J. U., Deng, X., Ketten, D., and Popper, A. N. (**2004**). ''Form and Function in the Unique Inner Ear of a Teleost: The Silver Perch (*Bairdiella chrysoura*), J. Comp. Neurol. **475**, 531–539.

Ramcharitar, J., Higgs, D., and Popper, A. N. (**2002**). ''Sciaenid inner ears: a study in diversity,'' Brain Behav. Evol. **58**, 152–162.

Sand, O., and Hawkins, A. D. (**1973**). ''Acoustic properties of the cod swim bladder,'' J. Exp. Biol. **58**, 797–820.

Schellart, N. A. M., and Popper, A. N. (**1992**). ''Functional aspects of the evolution of the auditory system of actinopterygian fish,'' in *The Evolutionary Biology of Hearing*, edited by R. R. Fay, A. N. Popper, and D. B. Webster (Springer-Verlag, New York), pp. 295–322.

Tavolga, W. N. (**1974**). ''Signal/noise ratio and the critical band in fishes,'' J. Acoust. Soc. Am. **55**, 1323–1333.

van Bergeijk, W. A. (**1967**). ''The evolution of vertebrate hearing,'' Contrib. Sens. Physiol. **2**, 1–49.

Weeg, M. S., Fay, R. R., and Bass, A. H. (**2002**). ''Directionality and frequency tuning of primary saccular afferents of a vocal fish, the plainfin midshipman (*Porichthys notatus*),'' J. Comp. Physiol. [A] **188**, 631–641.

Wightman, F. L., McGee, T., and Kramer, M. B. (**1977**). ''Factors influencing frequency selectivity in normal and hearing-impaired listeners,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York).

Zar, J. H. (**1984**). *Biostatistical Analysis*, 2nd ed. (Prentice–Hall, Englewood Cliffs, NJ).

The discrimination of baboon grunt calls and human vowel sounds by baboons

Robert D. Hienz,[a] April M. Jones, and Elise M. Weerts
Division of Behavioral Biology, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine/Bayview Campus, Baltimore, Maryland 21224-6823

(Received 3 February 2004; revised 11 June 2004; accepted 14 June 2004)

The ability of baboons to discriminate changes in the formant structures of a synthetic baboon grunt call and an acoustically similar human vowel (/ɛ/) was examined to determine how comparable baboons are to humans in discriminating small changes in vowel sounds, and whether or not any species-specific advantage in discriminability might exist when baboons discriminate their own vocalizations. Baboons were trained to press and hold down a lever to produce a pulsed train of a standard sound (e.g., /ɛ/ or a baboon *grunt* call), and to release the lever only when a variant of the sound occurred. Synthetic variants of each sound had the same first and third through fifth formants (F_1 and F_{3-5}), but varied in the location of the second formant (F_2). Thresholds for F_2 frequency changes were 55 and 67 Hz for the grunt and vowel stimuli, respectively, and were not statistically different from one another. Baboons discriminated changes in vowel formant structures comparable to those discriminated by humans. No distinct advantages in discrimination performances were observed when the baboons discriminated these synthetic *grunt* vocalizations. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1778902]

PACS numbers: 43.66.Gf, 43.80.Lb [WWA] Pages: 1692–1697

I. INTRODUCTION

Humans and other animals differ in their relative abilities to discriminate changes in pure tones versus changes in formant structures of vowel-like sounds. On the one hand, humans appear to be able to make much finer pure-tone discriminations than other animals. For example, humans detect changes in pure tones as small as 1.1 and 3.2 Hz at 600 and 2000 Hz (Wier *et al.*, 1977), which corresponds to a 0.16%–0.18% change from the base testing frequency. Birds, monkeys, and cats, however, exhibit much higher discrimination thresholds for pure tones; birds detect a 2% change in tone frequency (Sinnott *et al.*, 1980); monkeys detect a 1%–3% change in tone frequency (Sinnott *et al.*, 1987), and cats detect about a 6%–7% change in tone frequency (Hienz *et al.*, 1993). On the other hand, the discrimination of changes in formant structures of vowel-like sounds appears to be much more similar among humans and other animals. Average human thresholds for second formant changes in vowel sounds, for example, are about 1.5% of the center frequency (Kewley-Port and Watson, 1994). Similarly, monkeys can detect a change of 1.6% for single-formant vowels (Sommers *et al.*, 1992), and a change of 2.4%–3.2% for normal human vowels (Sinnott and Kreiter, 1991); cats can detect a change of 2.3% for the synthetic human vowel /ɛ/ (Hienz *et al.*, 1996). Thus, human, monkey, and cat discrimination performances are much more similar for vowel-like sounds than they are for pure tones, with this basic difference being predominantly due to the much greater tone discrimination abilities of humans.

It has been demonstrated that baboons can readily learn to discriminate between isolated human vowel sounds and that these discriminations are made on the basis of spectral differences among the vowels (Hienz and Brady, 1988). Additionally, baboons easily learn to discriminate their own ''grunt'' vocalizations in the laboratory, and can correctly recognize the sexual identity of the caller (Rendall *et al.*, 2004). The present report describes the ability of male baboons to discriminate changes in the second formant locations of both baboon ''grunt'' vocalizations and the human vowel /ɛ/.

II. METHOD

A. Subjects

Six adult male baboons (*Papio Anubis*) were housed in individual cages, with water provided *ad libitum*. Baboons performed in the experimental situation for 500-mg banana-flavored pellets, and supplemental food and fruit were provided daily after each experimental session to maintain normal body weight. Animals were also given full food rations when no experimental sessions were scheduled for the next day (e.g., on Fridays and Saturdays). Each animal's health was checked daily by a trained animal technician, and animals were periodically weighed and physically examined. Baboons CA and HO had no previous auditory discrimination experience prior to this study, while other baboons had prior experience in auditory threshold detection (baboons BE, BV, DR, and ST), frequency discrimination (baboons BE and BV), and in discriminating among different human vowels (baboons DR and BE) and different baboon grunts (baboons BE and BV). Animal care was in accordance with current NIH guidelines on ''Principles of laboratory animal care.''

[a]Electronic mail: bhienz@mail.jhmi.edu

 © 2004 Acoustical Society of America

B. Apparatus

Subjects were tested in a modified baboon squeeze cage fitted within a double-walled sound-attenuating chamber (IAC Inc., model 1201A). A 76-X 97-cm intelligence panel attached to one side of the cage contained a primate lever (BRS/LVE model PRL-003), a red light-emitting diode used as a cue light, and a pellet hopper for delivery of pellets via a mechanical feeder. With the animal positioned facing the panel, the cue light was at eye level, the feeding tube was at mouth level, and the response lever was at waist level in front of the right arm. Auditory stimuli were delivered through a wide-range speaker suspended above the test cage, vertically in line with the animals' normal head position, approximately 20 cm above ear level. Animals were ''shuttled'' into the experimental apparatus for testing via a metal transfer cage. Stimulus presentation, behavioral contingencies, and data recording were carried out automatically by computer.

C. Acoustic stimuli

Baboon grunt vocalizations were synthesized, based upon calls recorded in the wild by Dr. Drew Rendall, who has a large database of recordings from baboons in two populations in Botswana and South Africa that have been analyzed extensively (Owren *et al.*, 1997; Rendall *et al.*, 2000, 2004, 1999). The original vocalizations were collected using a Sony Professional Recorder TC-D5M and Sennheiser microphones under specific observational protocols in which observers noted the positions and identities of callers, apparent receivers, and any salient behaviors and/or events before and after a call. Because the baboons were fully habituated to the presence of human observers, very high quality recordings were made from close range, and were not subject to noise/reverberation problems. They were also obtained under natural ecological and social conditions. Due to similarities in vocal-tract size and structure between baboons and humans (Owren *et al.*, 1997), these sounds could readily be produced using articulatory synthesis software designed to mimic the vocal production process itself [e.g., Klatt (1980) synthesis as implemented in CSRE®]. For the present experiments a number of baboon grunts were synthesized via this method. Such a synthesis recreates the formant structure of the sounds, but does not preserve individual idiosyncrasies in the waveforms (although the laboratory baboons initially responded to these synthetic calls by grunting back to the speaker when the grunts were first played). It has been previously demonstrated that monkeys do respond similarly to both synthesized and natural calls (Norcross *et al.*, 1994; Owren, 1990). The synthetic calls were created using the average acoustic values of 211 different calls of six identified adult male baboons. The /ɛ/ vowel stimuli consisted of synthetic vowels generated using CSRE®, as taken from the data of Peterson and Barney (1952). The standard grunt was 120 ms in duration, with $F_1 = 325$ Hz, $F_2 = 1075$ Hz, $F_3 = 1950$ Hz, $F_4 = 2675$ Hz, and $F_5 = 3750$ Hz. An inflection in F_0 was added to the synthetic calls (F_0 going from 70 to 55 Hz over the 120-ms duration) to mimic that found in the natural calls. Synthetic variants of this synthetic grunt had the same values for F_1, F_3, F_4, and F_5, but with F_2's varying between 1075 and 1475 Hz. The standard vowel /ɛ/ was 250 ms in duration, with $F_1 = 530$ Hz, $F_2 = 1700$ Hz, $F_3 = 2480$ Hz, $F_4 = 3300$ Hz, and $F_5 = 3750$ Hz. An inflection in F_0 was also added (F_0 going from 120 to 80 Hz over the 250-ms duration) to mimic natural vowels. Synthetic variants of /ɛ/ had the same values for F_1, F_3, F_4, and F_5, but with F_2's varying between 1700 and 2000 Hz. The present study concentrated on testing stimulus variants that were higher in frequency than the standard stimuli since previous studies have demonstrated the existence of asymmetries in frequency, intensity, and vowel discriminations for monkeys, birds, and cats (Hienz *et al.*, 1996, 1980; Sinnott *et al.*, 1985); these studies have shown that downward shifts in frequency or intensity are much more difficult to train, and produce much higher thresholds.

The synthetic grunt and vowel sounds were presented via computer-controlled digital audio equipment from Tucker/Davis Technologies (TDT). All stimuli were digitally shaped (20-ms rise/fall) to avoid onset and offset transients. Signals were passed through a programmable attenuator, an amplifier, and then to the wide-range speaker in each test chamber and presented at an average level of approximately 75 dB. The acoustic system was calibrated with sound level meters and condenser microphones placed at ear level. To prevent subjects from responding to possible differences in intensity among stimuli, amplitude levels of all stimuli were varied by randomly selecting an intensity prior to each stimulus presentation. Stimulus intensity was varied over a range that encompassed the range of average intensity differences among all stimuli being tested (e.g., ± 3 dB, in 1-dB steps).

D. Procedure

Baboons were trained using operant conditioning techniques that employ positive reinforcement (Hienz *et al.*, 1993; Hienz and Sheckler, 1987; May *et al.*, 1995). Animals were shuttled into the test cage inside the chamber and trained to detect differences in the signals presented from the overhead speaker. Baboons pressed and held down the lever with their right hand to produce a 2/s pulsed train of a repeating standard sound (e.g., /ɛ/−/ɛ/−/ɛ/−/ɛ/). After a variable time of 1 to 7 s following the maintaining of this holding response, the stimulus train changed by alternating twice between the standard sound and one of the shifted second formant variants of the sound. A release of the lever within a 1.5-s ''window'' from the onset of the first alternating stimulus was considered a correct detection of the change to a variant sound, and was reinforced with a banana pellet. Releases at any other time were considered errors and punished with a brief, 11–15 s timeout from the experimental contingencies. One of four comparison stimuli was randomly selected to occur on any given trial (method of constant stimuli). False alarm rates of the subjects were also measured by presenting ''catch'' trials (during which no stimulus variants occurred), randomly interspersed on 20% of the trials. Releases during catch trials were also punished with a 11–15-s timeout, individually adjusted to minimize false alarms.

TABLE I. Grunt and vowel discrimination (ΔF) values for individual baboons as interpolated from psychometric functions (left), and as computed for a d' value of 1.0 (right).

Interpolated ΔF_2 Thresholds				ΔF_2 Thresholds for a d' of 1.0			
Grunt		Vowel		Grunt		Vowel	
Baboon	ΔF_2	Baboon	ΔF_2	Baboon	ΔF_2	Baboon	ΔF_2
BE	39.96	BE	47.70	BE	29.10	BE	42.30
BV	75.91	DR	89.61	BV	45.10	DR	62.70
CA	56.88	HO	39.60	CA	42.60	HO	30.50
ST	45.48	ST	91.38	ST	31.10	ST	82.30
Average:	54.56		67.07		36.98		54.45
Standard deviation:	15.88		27.26		8.05		22.84
Grunt		Vowel		Grunt		Vowel	
Baboon	$\Delta F_2/F_2$	Baboon	$\Delta F_2/F_2$	Baboon	$\Delta F_2/F_2$	Baboon	$\Delta F_2/F_2$
BE	3.72	BE	2.81	BE	2.71	BE	2.49
BV	7.06	DR	5.27	BV	4.20	DR	3.69
CA	5.29	HO	2.33	CA	3.96	HO	1.79
ST	4.23	ST	5.38	ST	2.89	ST	4.84
Average:	5.08		3.95		3.44		3.20
Standard deviation:	1.48		1.60		0.75		1.34

Baboons were shuttled daily into an experimental chamber, and discrimination performances and reaction times to the stimuli were measured. Sessions were conducted five days per week. Baboons typically completed 400–500 discrimination trials during 100-min sessions. Data obtained during the first 15 min of each session were considered ''warm-up'' data, and were not included in final analyses.

Prior to each testing session, four of the F_2 variant stimuli were selected for presentation during a session, with each variant occurring randomly on 20% of the trials and no stimulus change (i.e., catch trials) occurring randomly on 20% of the trials. The formant variants employed during training were selected to maintain proper stimulus control and yet allow for the bracketing of the likely threshold region. This adjustment was done on an individual basis, since the same set of stimuli might be relatively easy for one animal to discriminate, but difficult for another. Typically, two of the variants were selected to be clearly above threshold and generate 100% detection, a third stimulus was selected to generate responding in the 80–95% range, and a fourth stimulus to generate responding below the 50% threshold level. The final grunt step sizes were 280, 200, 80, and 20 Hz for baboon BE; 400, 240, 120, and 20 Hz for baboon BV; 280, 200, 120, and 20 Hz for baboon CA; and 280, 200, 80, and 10 Hz for baboon ST. The final vowel step sizes were 115, 90, 65, and 15 Hz for baboon BE; 300, 200, 150, and 40 Hz for baboon DR; 160, 110, 60, and 10 Hz for baboon HO; and 230, 180, 130, and 50 Hz for baboon ST. Second formant frequency difference limens (ΔF_2 DLs) were measured by allowing performances to stabilize until seven to ten consecutive sessions were obtained for which false alarm rates were low ($<25\%$) and stable across sessions, and stimulus discriminability (as assessed by percent correct scores and d' estimates) showed no upward or downward trends across sessions. ΔF_2 DLs were calculated by summing the numbers of correct responses and total trials across the last five stable consecutive sessions under each condition and calculating percent correct scores. Psychometric functions were then constructed by plotting these scores as a function of the change in F_2 for each of the comparison stimuli, relative to the standard sound's F_2 value. To make the present results comparable to previous animal psychoacoustic studies, ΔF_2 values were calculated in two manners. First, ΔF_2 DLs were defined as the formant frequency difference producing a detection score halfway between the false alarm rate and 100%. Second, for comparisons across subjects, estimates of discriminability were based upon d' values, which provide an unbiased estimate of the accuracy of a discrimination (Macmillan and Creelman, 1991). ΔF_2 DLs based upon d' values were defined as those formant frequency differences corresponding to a d' value of 1.0. Values of ΔF_2 DLs computed by both of these methods are shown in Table I.

III. RESULTS

Figure 1 shows psychometric functions in which the percentage of lever releases is plotted as a function of changes in F_2 for variants of the baboon grunt (top) and the vowel /ɛ/ (bottom). All baboons produced psychometric functions that showed a decreasing percentage of correct releases for smaller F_2 changes. Individual differences were evident: for grunts, baboons BE and ST were quite sensitive to these F_2 changes while baboon BV was slightly less sensitive than the other subjects; for vowels, baboons BE and DR were most sensitive to the F_2 changes while HO and ST were less sensitive. Differences in false alarm rates (points plotted at zero frequency change) were minimal, with all subjects having rates below 20%.

Table I (top) shows the ΔF_2 values computed as interpolated thresholds corresponding to a detection score halfway between the false alarm rate and a 100% score (left), and thresholds corresponding to a d' value of 1.0 (right). Both the interpolated and the d' ΔF_2 thresholds were lower for the grunt stimuli than for the vowel stimuli, although the

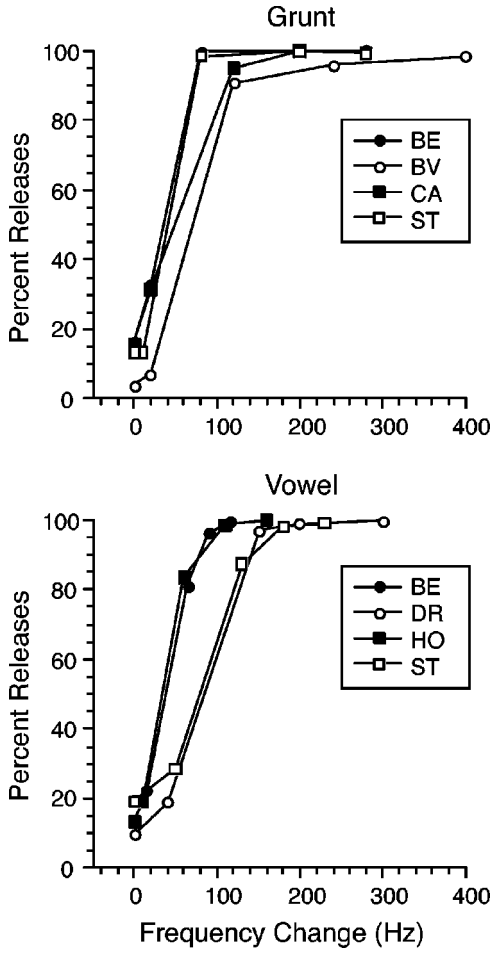


FIG. 1. The mean percentage of lever releases plotted as a function of the change in F_2 frequency between the standard and comparison stimuli for the grunt (top) and vowel (bottom) stimuli. Individual psychometric functions are displayed for each baboon. Each data point is the mean of the last five stable sessions; points plotted at the zero frequency change value represent false alarm rates.

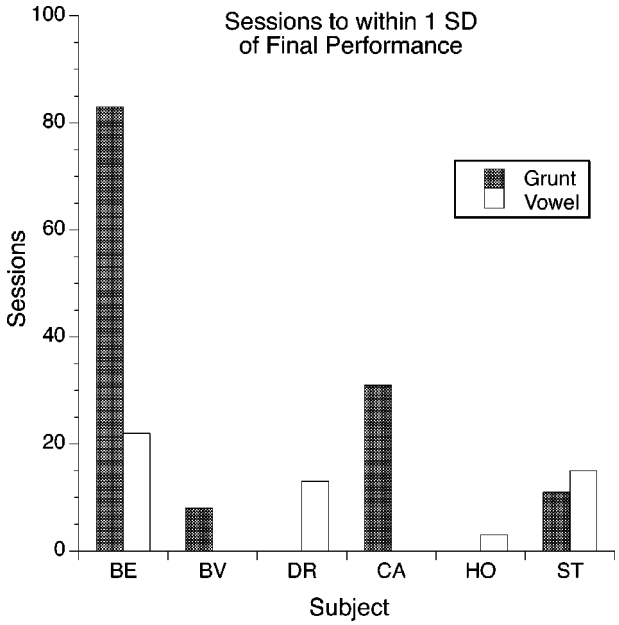


FIG. 2. The number of sessions conducted prior to each baboon producing a discrimination threshold value that was within one standard deviation of his final performance threshold. Within-subject data comparisons are shown for baboons BE (far left) and ST (far right) on both grunt and vowel discrimination threshold performances. Baboons BV and CA only performed the grunt formant discrimination threshold procedure; baboons DR and HO only performed the vowel formant discrimination threshold procedure.

differences were not statistically significant (Student's *t*). Additionally, more between-subject variability in ΔF_2 thresholds occurred for the vowel stimuli than the grunt stimuli, as indicated by the standard deviation measures. The trend of the grunt ΔF_2 values to be lower than the vowel ΔF_2 values, however, could be accounted for by the fact that the grunts were lower in frequency than the vowels. For this reason, Table I (bottom) shows these same ΔF_2 values expressed in terms of the ratio of the measured ΔF_2's to the base frequency (i.e., $\Delta F_2/F_2$). For these ratios, the interpolated grunt ΔF_2 thresholds are slightly higher than the vowel measures while the $d'\Delta F_2$ thresholds are about the same for the grunts and vowels; again, none of these slight differences approached statistical significance.

Further comparisons of the baboons' performances were carried out to determine if any differences in their performances with the grunt and vowel stimuli could be detected. Figure 2 compares how many sessions were required before the first session occurred during which a baboon was within one standard deviation of his final threshold performance level. For this comparison, pairings of baboon BV with DR, and of baboon CA with HO, were made since these pairs had the most comparable ΔF_2 thresholds with one another. For both the within-subject and between-subject pairs, no significant trends were observed to indicate a significant advantage of one type of stimulus over the other. The within-subject comparison for baboon ST showed an advantage for the grunts in that it took fewer sessions for him to reach the criterion with the grunts, compared to the vowels; however, the within-subject comparison for baboon BE showed that he reached criterion more quickly with the vowel stimuli. In the two between-subject comparisons, it required more sessions for baboon CA to reach criterion with the grunts than for baboon HO to reach criterion with the vowels, and subjects BV and DR each required about the same number of sessions to reach with the grunts and vowels, respectively. A similar comparison with a two-standard deviation criterion (data not shown) also showed no significant differences between the two types of stimuli used. Figure 3 shows the two possible within-subject comparisons of response latencies as a function of the percent frequency change for grunt and vowel stimuli for baboons BE and ST. Baboon ST obviously showed much faster responding to the grunt stimuli, while baboon BE showed no differential responding as a function of the stimulus type. Additional comparisons of the false alarm rates measured for grunt and vowel stimuli also showed no significant differences (12.2% versus 15.2% for grunts and vowels, respectively).

Figure 4 presents a comparison of the present interpolated $\Delta F_2/F_2$ values for grunts and vowels from the present study with similar tone and vowel $\Delta F/F$ values for humans, birds, monkeys, and cats. One can clearly see that the tone $\Delta F/F$ values (filled symbols) are strikingly lower for humans than for monkeys, birds, or cats. However, the $\Delta F/F$ values for vowels (open symbols) are much more comparable for all

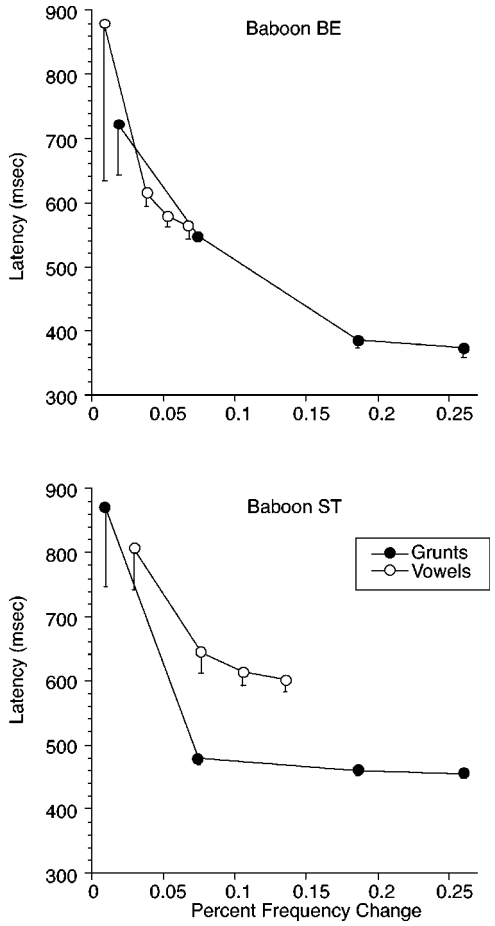


FIG. 3. The latency of responding to grunt and vowel stimuli as a function of the percent change in stimulus frequency for the two baboons tested under both conditions (BE, top graph; ST, bottom graph). Each data point is the mean of the last five stable sessions, with error bars indicating 1 standard deviation.

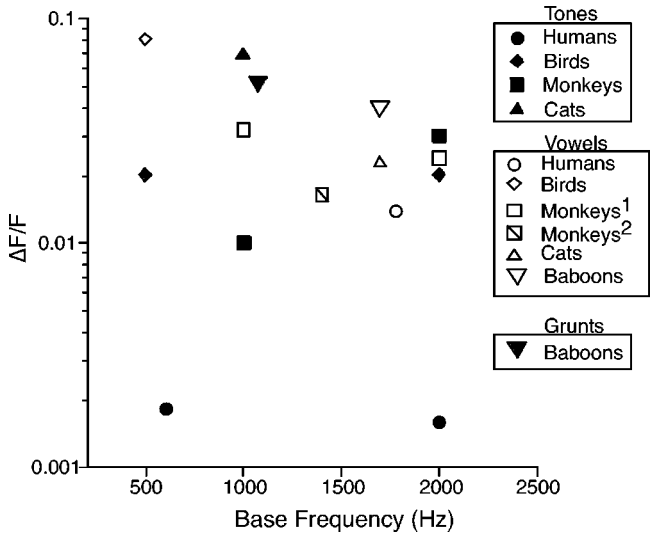


FIG. 4. Average interpolated threshold value ($\Delta F_2/F_2$) changes for the baboon grunt and vowel stimuli of the present study, as compared with similar tone (filled symbols) and vowel (open symbols) ΔF/F values for humans, birds, monkeys, and cats. The sources of the tone ΔF/F data are as follows: humans: Wier *et al.* (1977), birds: Hienz *et al.* (1981), monkeys: Sinnott and Kreiter (1991), cats: Hienz *et al.* (1996). The sources of the vowel ΔF/F data are as follows: humans: Kewley-Port and Watson (1994), birds: Hienz *et al.* (1993) and Hienz and Sheckler (1987), monkeys[1]: Sinnott and Kreiter (1991), monkeys[2]: Sommers *et al.* (1992), cats: Hienz *et al.* (1996).

species shown. The current ΔF/F thresholds for both baboon grunts and human vowels appear to fall within the discriminability range of the other species shown.

IV. DISCUSSION

The present results indicate clearly that baboons have an ability comparable to humans and other animals to discriminate changes in the formant structures of complex signals such as baboon grunt calls and human vowel sounds. Further, the baboons' abilities to discriminate changes in the formant structures of their own species-typical grunt calls is approximately the same as their abilities to discriminate changes in comparable human vowels. The average ΔF_2 thresholds among the four baboons tested ranged from 40 to 76 Hz for grunt stimuli, and from 40 to 92 Hz for vowel stimuli. Similar results have been reported for other nonhuman primates by Sommers *et al.* (1992), who measured difference thresholds in Japanese macaques for vowels that had been filtered to remove frequency information above and below a single formant. They reported an average vowel ΔF of 23 Hz for changes in single-formant vowels centered at 1400 Hz. Sinnott and Kreiter (1991) also measured vowel ΔF's for multiple-formant vowels in macaque monkeys, and reported vowel ΔF's of 32–48 Hz for formants near 1.0 and 2.0 kHz. Both of these studies also employed behavioral procedures similar to the present study in that animals were trained to release a response lever to indicate the detection of changes in a pulsed vowel sound.

It has recently been demonstrated that baboons can readily discriminate between the grunts of adult male and female baboons recorded in the wild (Rendall *et al.*, 2004). In this study baboons also performed a reaction-time procedure in which they were trained to release to a female baboon grunt, and not release to a male baboon grunt. All subjects were able to discriminate new grunt tokens from these two individual calls as well as additional grunts from completely novel males and females. Rendall *et al.* concluded that the discriminations were based on the acoustic-structural differences in the calls associated with the callers' different sex, since the callers were not familiar to the baboon subjects. In this study, the male–female sex differences in mean F_0 were 60 Hz for a base male F_0 of 53 Hz, 120 Hz for a base male F_1 of 332 Hz, and 368 Hz for a base male F_3 of 1062 Hz. The present results indicate that these acoustic differences are all well above the current formant discrimination thresholds of baboons. Of course, the ability of the baboons to correctly discriminate between the grunts of males and females is not surprising given that the magnitude of the acoustic-structural differences between their grunts is likely to be well within the species' frequency discrimination abilities.

The present results also indicate that baboons can discriminate changes in vowel formants at a level roughly comparable to that observed for humans. The present study's vowel ΔFs for baboons were above the average human vowel ΔF of 25 Hz at a comparable frequency (Kewley-Port

and Watson, 1994), although the best baboon's performance of 40 Hz approached human threshold values. Given that many procedural differences exist between the above-noted nonhuman animal studies and the study of Kewley-Port and Watson, it is noteworthy that the vowel ΔF's reported in animals still approach those reported for humans, suggesting that humans and other animals are relatively similar in their ability to discriminate acoustic differences in vowel sounds.

Many comparisons of hearing between humans and nonhuman primates have been made in an effort to determine those aspects of human hearing likely to be species-specific to humans. Thus, Sinnott and Kreiter (1991) compared the vowel discrimination capacities of monkeys and humans, noting that monkeys and humans appear to discriminate vowel formant changes more similarly than they do pure tone frequency changes. In the present study, however, comparisons were made between the ability of baboons to discriminate formant structure changes in baboon grunts versus human vowels to see if baboons might display a species-specific sensitivity to baboon vocalizations. No evidence was found for such a differential sensitivity, although the present study is far from being definitive for numerous reasons (e.g., small sample size, limited within-subject comparisons, the use of synthetic stimuli). Additionally, differences in human-animal hearing sensitivities have been found previously. For example, Sinnott (1987) compared the abilities of red-winged blackbirds and brown-headed cowbirds trained with a go-right/go-left procedure to identify conspecific and alien song themes, and showed via an analysis of response latencies that each bird species exhibits superior identification of conspecific calls, relative to the alien species, and also that humans attend primarily to the final song elements in the bird calls while birds direct their attention primarily to the initial elements. In the present study, no substantial differences were observed between the baboons' performances with grunt versus vowel stimuli, although one baboon did show a large difference in the latency of responding to the stimuli (i.e., latencies shorter by about 150 ms for the grunt stimuli). It should be noted, however, that the Sinnott (1987) study compared the ability of birds and humans to identify between different types of stimuli by having subjects respond on different keys to indicate whether a call was ''conspecific'' or ''alien,'' whereas the present study examined the threshold for discriminability of baboon grunts and human vowels by having subjects release a lever to indicate a change in a repeating background stimulus. Thus, it may not be surprising that differential latency responses were not consistently observed in the present study, since the study was designed to examine discriminability thresholds for baboon grunts and human vowels, and not the relative recognizability of the signals.

ACKNOWLEDGMENTS

This research was supported by Research Grant Nos. DA04731 and DA 12139 from the National Institute on Drug Abuse, and by Grant No. MH65317 from the National Institute of Mental Health to RDH. The authors wish to thank D. A. Pyle and J. E. Johnson for their technical assistance.

Hienz, R. D., and Brady, J. V. (**1988**). ''The acquisition of vowel discriminations by nonhuman primates,'' J. Acoust. Soc. Am. **84**, 186–194.

Hienz, R. D., and Sheckler, D. A. (**1987**). ''A comparison of different methods for determining frequency difference thresholds in cats,'' in Abstracts of the Tenth Midwinter Research Meeting, Association for Research in Otolaryngology, p. 84.

Hienz, R. D., Aleszczyk, C. M., and May, B. J. (**1996**). ''Vowel discrimination in cats: Thresholds for the detection of second formant changes in the vowel /ɛ/,'' J. Acoust. Soc. Am. **100**, 1052–1058.

Hienz, R. D., Sachs, M. B., and Aleszczyk, C. M. (**1993**). ''Frequency discrimination in noise: Comparison of cat performances with auditory-nerve models,'' J. Acoust. Soc. Am. **93**, 462–469.

Hienz, R. D., Sachs, M. B., and Sinnott, J. M. (**1981**). ''Discrimination of steady-state vowels by blackbirds and pigeons,'' J. Acoust. Soc. Am. **70**, 699–706.

Hienz, R. D., Sinnott, J. M., and Sachs, M. B. (**1980**). ''Auditory intensity discrimination in blackbirds and pigeons,'' J. Comp. Physiol. Psychol. **94**, 993–1002.

Kewley-Port, D., and Watson, C. S. (**1994**). ''Formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95**, 485–496.

Klatt, D. H. (**1980**). ''Software for a cascade/parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.

Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

May, B. J., Huang, A. Y., Aleszczyk, C. M., and Hienz, R. D. (**1995**). ''Design and conduct of sensory experiments for domestic cats,'' in *Methods of Comparative Acoustics*, edited by G. M. Klump, R. J. Dooling and R. R. Fay (Birkhauser Verlag, Basil/Switzerland), pp. 95–108.

Norcross, J. L., Newman, J. D., and Fitch, W. (**1994**). ''Responses to natural and synthetic Phee calls by common marmosets (*Callithrix jacchus*),'' Am. J. Primatol **33**, 15–29.

Owren, M. J. (**1990**). ''Acoustic classification of alarm calls by vervet monkeys (*Cercopithecus aethiops*) and humans (*Homo sapiens*): II. Synthetic calls,'' J. Comp. Psychol. **104**, 29–40.

Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): Implications for production processes and functions,'' J. Acoust. Soc. Am. **101**, 2951–2963.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Rendall, D., Cheney, D. L., and Seyfarth, R. M. (**2000**). ''Proximate factors mediating 'contact' calls in adult female baboons (*Papio cynocephalus ursinus*) and their infants,'' J. Comp. Psychol. **114**, 36–46.

Rendall, D., Owren, M. J., Weerts, E. M., and Hienz, R. D. (**2004**). ''Sex differences in the acoustic structure of vowel-like grunt vocalizations in baboons and their perceptual discrimination by baboon listeners,'' J. Acoust. Soc. Am. **115**, 411–421.

Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (**1999**). ''The meaning and function of grunt variants in baboons,'' Anim. Behav. **57**, 583–592.

Sinnott, J. M. (**1987**). ''Modes of perceiving and processing information in birdsong (Agelaius phoeniceus, Molothrus ater, and Homo sapiens),'' J. Comp. Psychol. **101**, 355–66.

Sinnott, J. M., and Kreiter, N. A. (**1991**). ''Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans,'' J. Acoust. Soc. Am. **89**, 2421–2429.

Sinnott, J. M., Owren, M. J., and Petersen, M. R. (**1987**). ''Auditory frequency discrimination in primates: Species differences (*Cercopithecus, Macaca, Homo*),'' J. Comp. Psychol. **101**, 126–131.

Sinnott, J. M., Petersen, M. R., and Hopp, S. L. (**1985**). ''Frequency and intensity discrimination in humans and monkeys,'' J. Acoust. Soc. Am. **78**, 1977–1985.

Sinnott, J. M., Sachs, M. B., and Hienz, R. D. (**1980**). ''Aspects of frequency discrimination in passerine birds and pigeons,'' J. Comp. Physiol. Psychol. **94**, 401–415.

Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (**1992**). ''Formant frequency discrimination by Japanese macaques (*Macaca fuscata*),'' J. Acoust. Soc. Am. **91**, 3499–3510.

Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.

Localization ability with bimodal hearing aids and bilateral cochlear implants

Bernhard U. Seeber[a)]
AG Technische Akustik, MMK, Technische Universität München, Arcisstr. 21, D-80333 München, Germany

Uwe Baumann
Department of Experimental Audiology, University of Munich, D-81377 München, Germany

Hugo Fastl
AG Technische Akustik, MMK, Technische Universität München, Arcisstr. 21, D-80333 München, Germany

(Received 15 March 2004; revised 30 May 2004; accepted 4 June 2004)

After successful cochlear implantation in one ear, some patients continue to use a hearing aid at the contralateral ear. They report an improved reception of speech, especially in noise, as well as a better perception of music when the hearing aid and cochlear implant are used in this bimodal combination. Some individuals in this bimodal patient group also report the impression of an improved localization ability. Similar experiences are reported by the group of bilateral cochlear implantees. In this study, a survey of 11 bimodally and 4 bilaterally equipped cochlear implant users was carried out to assess localization ability. Individuals in the bimodal implant group were all provided with the same type of hearing aid in the opposite ear, and subjects in the bilateral implant group used cochlear implants of the same manufacturer on each ear. Subjects adjusted the spot of a computer-controlled laser-pointer to the perceived direction of sound incidence in the frontal horizontal plane by rotating a trackball. Two subjects of the bimodal group who had substantial residual hearing showed localization ability in the bimodal configuration, whereas using each single device only the subject with better residual hearing was able to discriminate the side of sound origin. Five other subjects with more pronounced hearing loss displayed an ability for side discrimination through the use of bimodal aids, while four of them were already able to discriminate the side with a single device. Of the bilateral cochlear implant group one subject showed localization accuracy close to that of normal hearing subjects. This subject was also able to discriminate the side of sound origin using the first implanted device alone. The other three bilaterally equipped subjects showed limited localization ability using both devices. Among them one subject demonstrated a side-discrimination ability using only the first implanted device. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1776192]

PACS numbers: 43.66.Pn, 43.66.Qp, 43.66.Sr [NFV] Pages: 1698–1709

I. INTRODUCTION

The localization of sounds is a basic attribute of binaural hearing. It enables us to draw attention to sound sources–to look at a certain speaker in conversations, or to react to dangers coming from behind. Hearing impaired persons often suffer from a restriction in these abilities. A restoration of hearing abilities should therefore focus not only on monaural and binaural speech reception but also on localization. Sound localization in the horizontal plane is essentially a binaural process and is based on the evaluation of frequency-dependent interaural time and level differences (Middlebrooks and Green, 1991). Along with the evaluation of these cues, interaural time differences (ITDs) derived from the temporal envelope at each ear serve as an additional cue (Grantham, 1995; Henning, 1974). Normal hearing subjects are able to localize wide-band noise stimuli in the frontal horizontal plane very precisely. At more lateral positions the localization error increases. Blauert (1997) states an underestimation of the lateral position of about 10° for the lateral direction 90° and a shift of 1° to the right for the frontal direction.

Some users of cochlear implants (CIs) continue to use a hearing aid (HA) on the non-implanted side after successful cochlear implantation. They report improved speech perception especially in noise, as well as better perception of music when both devices are used together. Some individuals in this bimodal patient group also report improved localization ability. Similar experiences are reported by the group of bilateral cochlear implantees. These subjective experiences suggest that binaural information is transmitted with bimodal combination of CI and HA as well as with bilateral CIs. The purpose of the present study is to assess the impact of the bimodal and bilateral usage on auditory localization. To avoid quantization of the listeners' response scale (Durlach *et al.*, 1981; Hartmann *et al.*, 1998), that would be introduced by the fixed spacing of speakers used in identification methods, a continuous laser pointing technique with high accuracy and resolution is used (Seeber, 2002). Subjects adjust a laser spot projected onto a curtain covering several loudspeakers to the direction of perceived sound incidence.

[a)]Present address: Auditory Perception Lab, University of California at Berkeley, 3210 Tolman Hall #1650, Berkeley, CA 94720-1650. Electronic mail: bernhard_seeber@gmx.de

The computer-controlled laser spot moves smoothly according to the turning of a trackball, which makes this method fast and intuitive. In contrast to previous studies using identification methods, this new method conveys a continuous distribution of perceived angles. This allows for a direct analysis for the parameters of the localization process: error, bias, and variance.

A coarse localization ability has been demonstrated by some users of unilateral cochlear implants. Gray and Baguley (1993) found with three subjects a 43% correct identification of the quadrant of sound origin in a room if head movements were allowed. The subjects of Luntz *et al.* (2002) identified a 1 kHz sinusoid from five frontal, horizontal speakers with a 45° spacing at 27.5% correct, and through training an improvement to 66.5% accuracy could be achieved.

The use of bimodal aids (one CI and one HA) allows the use of binaural cues. Specifically, Dooley *et al.* (1993) reported the utilization of binaural cues of the CI and HA for speech reception when both devices were controlled by a binaural processor. Improved identification of the side of sound origin using bimodal aids was shown by Tyler *et al.* (2002b). Localization studies or identification studies employing more than two speakers with bimodal aids are currently not reported to our knowledge.

The use of binaural cochlear implants has led to several reports of advantages relative to the monolateral case. A binaural advantage for speech reception was demonstrated by, e.g., van Hoesel *et al.* (2002), van Hoesel and Tyler (2003), Lawson *et al.* (1998, 2000), and Müller *et al.* (2002). Additionally, Tyler *et al.* (2002a, b) showed a good side discrimination ability for speakers at ±45°. Van Hoesel *et al.* (2002) demonstrated, for one subject, a good identification ability of 11 speakers with relative errors of about 9° and rms errors of 15–16°. Recently, van Hoesel and Tyler (2003) showed average rms errors of about 10° for source identification with eight speakers. Most of their subjects identified the two frontal speakers with higher accuracy than speakers located on the side. Using monaural aids, most subjects were able to discriminate the side.

Superior localization ability for sources in the horizontal plane requires inputs from both ears. When using an HA and CI in combination, different stimulation schemes complicate an integrated evaluation of bilateral information. Different compression methods and setups in both devices will affect interaural level cues. Different processing times between the CI and HA will lead to an offset in interaural temporal cues. The pulsatile strategy of the CI will not transmit changes in envelope before the next pulse, which creates an additional temporal uncertainty. Thus, it is interesting to see whether a localization ability with a bimodal arrangement can be achieved based on modified interaural cues.

II. METHODS

A. Cochlear implants and hearing aids

The Combi 40+ implant (Med-El, Innsbruck) consists of an intracochlear array of 12 active electrodes spaced 2.4 mm apart, and the CI24M implant (Cochlear, Melbourne) supports 22 active electrodes with 0.75 mm distance. Both implants are equipped with an extracochlear electrode lying on the skull beneath M. Temporalis. The electrodes are activated by an implanted receiver-stimulator, which receives digitally encoded signals from the speech processor via a transcutaneous inductive link. In the case of the Combi 40+ implant the current pulses are delivered to the electrodes in monopolar mode whereby the extracochlear electrode serves as a reference. The CI24M implant electrode wiring can be configured in different ways. The most common electrode configuration is the MP1+2 mode which is used by subject DT. This mode still produces monopolar stimulation, but with reference to two remote electrodes rather than one.

All of the subjects used a BTE speech processor (Med-El: Tempo+, Cochlear: Esprit). The Tempo+ speech processor delivers a high-rate continuous interleaved sampling strategy (CIS) to the implant with 1515 pulses per second (pps) per channel, whereas the Esprit device employs the SPEAK processing strategy with a lower rate of approximately 250 pps per channel (Skinner *et al.*, 1994; Zierhofer *et al.*, 1995).

Within the scope of the study all subjects of the bimodal group were fitted on the non-implanted ear with an identical, digitally programmable hearing aid with directional microphone (Phonak PZ A 4 ''Power Zoom,'' max. amplification 79 dB, max. sound pressure level on output 144 dB/SSPL, DIN IEC 118-0). Program 1 (P1) was adjusted for the omnidirectional microphone, whereas program 2 (P2) used the same setting with a directional microphone. The first adjustment of the amplification was made according to NAL RP procedure (Keidser *et al.*, 1996). After a habituation time of at least one week, a fine tuning of the fitting of the hearing aid was performed for each subject.

B. Bimodal subjects

Eleven postlingually deafened adults (eight female, three male) participated in the bimodal localization experiments. All but one were implanted with the Combi 40+ device. Subject DT received the Nucleus/Cochlear Corporation implant model CI24M operating in MP1+2 mode. The electrode array was fully inserted for all but one subject. For subject JJ three basal electrodes were not available for stimulation because they were external to the cochlea.

Further demographic information regarding the subjects is listed in Table I in which the subjects are ranked according to their hearing loss at the ear fitted with the hearing aid. All subjects had used their cochlear implant a minimum of 6 months. The speech processor programs were the same during and outside the experiments.

The subjects had a wide range of speech perception abilities. For reference, the audition-alone scores when using either HA, CI, or both for a sentence test in quiet are shown in Table I (Baumann, 2001) [Oldenburg sentence test, Wagener *et al.* (1999a, b, c)].

Patients DT and HS show higher-than-average residual hearing on the non-implanted ear of 73 dB and 66 dB PTA, respectively (PTA: pure tone average at 500 Hz, 1 kHz, and 2 kHz). Subject HS was deafened on his right ear due to a blow to the head in his childhood. The blow caused severe

TABLE I. Overview of subjects with bimodal hearing instruments. Subjects are sorted for PTA.

Subject	Age (years)	CI experience[a]	Deafness duration[b]	Aethiology[c]	CI type	CI side	HA use	Pure tone threshold dB HL 500	1000	2000	4000 Hz	PTA (dB[d])	Sentence HA	CI	CI+HA[e]
HS	28.8	0.8	5.9	Trauma	C40+	r	yes	50	68	80	85	66	100	99	100
DT	79.4	2.0	2.8	pd	CI24M	l	yes	61	78	81	120	73	100	81	100
PG	61.7	3.6	0.3	SHL	C40+	l	no	80	80	75	72	78	2	95	98
BH	34.2	0.5	1.8	pd	C40+	r	no	76	72	90	100	79	0	82	76
JJ	49.8	1.3	1.3	toxic	C40+	l	yes	79	80	80	69	80	0	69	81
RM	27.8	4.2	10.5	pd	C40+	l	yes	91	92	90	72	91	1	72	77
EK	76.2	2.9	0.7	pd	C40+	l	no	105	89	80	81	91	18	94	91
AB	33	0.5	20.9	pd	C40+	l	yes	90	100	110	×	99	44	2	43
RL	63.0	1.1	2.5	pd	C40+	r	no	90	108	118	×	105	6	84	93
EM	22.5	2.0	0.8	Cogan	C40+	l	no	105	105	112	110	107	10	90	95
IS	76.3	0.7	0.7	pd	C40+	r	yes	105	120	×	×	113	6	65	79

[a]Experience with CI in years. AB and RL also conducted a second session 1.9 years later.
[b]Duration of deafness in years before implantation.
[c]For HS and DT see text, pd: progressive degenerative, SHL: sudden hearing loss, toxic: toxic effects of Gentamycin treatment, Cogan: Cogan syndrome.
[d]PTA: Pure tone average at 500 Hz, 1 kHz, and 2 kHz in dB.
[e]Oldenburg sentence test: speech reception in percent (Baumann, 2001).

head trauma and led to a fraction of his skull and cochlea on the right side with immediate deafness and an ongoing tinnitus. His hearing capabilities on the left side were only minorly affected shortly after the accident, but over the years a progressive degeneration of his hearing occurred. The rationale to consider a cochlear implant in the right ear was (1) HS had no residual hearing, which could be damaged by insertion of the electrode; (2) it was expected that due to the progressive hearing loss the left ear will deafen in the next years; and (3) the use of the cochlear implant might mask the annoying tinnitus on the right. Subject DT was not satisfied with the fitting of his hearing aids, although his benefit in terms of speech perception was considerable. He expected a major improvement from the CI. The costs of the implantation were covered by his insurance, and he was implanted at a major German CI center.

As indicated in the row ''HA use,'' six patients (HS, DT,JJ,RM,AB,IS) used their private HA on the non-implanted ear on a regular basis before the beginning of the study. All subjects were fitted with the same type of hearing aid utilized within the study (see above). One week after the first fit of the study HA a fine tuning was accomplished. The localization experiments were executed 2–6 weeks after the fine tuning. All subjects used the study HA/CI combination before the study had started, although BH did not use the study HA on a regular basis.

C. Bilateral subjects

Four subjects wearing cochlear implants bilaterally participated in the localization experiments. Demographic data of the subjects are given in Table II (age 20–65 years, Med-El C40/C40+). All subjects had worn their first implant for 2.2 to 6.7 years. The experience with bilateral CI was at least 0.8 years. All bimodal and bilateral subjects received a payment for their participation.

D. Apparatus

The localization experiments take place in a completely darkened anechoic chamber (dimensions $L\times W\times H=7.5\times 4.2\times 2.8$ m^3) in which the apparatus is installed. Eleven identical closed-cabinet loudspeakers are mounted on a circular tube with a radius of $r=1.95$ m at ear level of the subject. The speakers span an angle of $-50°$ left to $+50°$ right with a spacing of 10°. The frequency response of each speaker is individually equalized by an FIR-filter to 125 Hz to 20 kHz in ± 2.5 dB at the subject's head position. The speakers are switched by a custom-made relay-unit using a PC-type computer. Light emitting diodes (LEDs) are mounted at a distance of 10 cm concentric in front of each speaker. A laser spot is projected onto a curtain in front of the loudspeakers. The curtain is opaque for the subject's gaze but acoustically transparent and translucent for the light of the

TABLE II. Overview of subjects with bilateral CI. Subjects are sorted for experience.

Subject	Age (years)	CI first[a]	CI second[b]	Deafness first[c]	Deafness second[d]	Aethiology First	second	Implant first	second	CI first[e]
BW	49.1	2.9	1.9	1.1	2.1	otosclerosis		C40+	C40+	r
IB	65.1	2.2	1	>20	10	prog. degen.	Menier	C40+	C40+	r
KH	51.3	6.5	0.9	21	21	prog. degen.	head inj.	C40	C40+	r
RL	19.8	6.7	0.8	>10	>10	prog. hereditary		C40+	C40+	l

[a]Experience with first CI in years at time of first session. BW also conducted a second session 1.3 years later.
[b]Experience with bilateral CI in years.
[c]Duration of deafness before implantation on the first implanted ear in years.
[d]Duration of deafness before implantation on the second implanted ear in years.
[e]First implanted ear.

LEDs. The laser beam is deflected in the *x* and *y* direction by two laser-scanner galvanometers which are controlled by a digital-to-analog converter from a computer. Subjects indicate the perceived position of the sound source by a trackball. The projected laser spot is moved smoothly according to the rotation of the trackball on a horizontal track in front of the speakers. The laser spot can be positioned with a precision better than 0.2° at any point on the track within an angle of −70° left to +70° right. A computer performs the coordinate transformation between the angle of the laser spot seen by the subject and the deflection of the laser beam using calibration data. The filtered digital sound data (16 bit, $f_s = 44\,100$ Hz) are written out through a digital soundcard to an external D/A-converter. The speaker signal is amplified and calibrated with a voltmeter before reaching the switching unit. The experimental procedure is controlled by a Matlab routine with the help of customized interface software for the experimental setup. A detailed description of the apparatus and the method can be found in Seeber (2002).

E. Procedures

During the experiments, the subject sat on a chair in the center of the speaker array. The head was stabilized by a head rest and the subject was instructed not to move his or her head. The subject was monitored through an infrared camera which allowed the detection of movements. The sensitivity adjustment of the speech processor for the CI and the amplification of the HA were adjusted prior to the experiments to give the perception of equal loudness at both ears for a speech signal arriving from the front of the subject. However, this was difficult to achieve for some subjects who had a small amount of residual hearing, because feedback problems limited the amplification of the HA.

At the beginning of an experiment, a light appeared directly in front of the subject for 5 s in a completely darkened anechoic chamber. This allowed the subject to align his or her head to the frontal direction. After a pause of 500 ms a target sound was presented in a randomly selected angle within −50° left to +50° right in 10° intervals. Gaussian white noise (125 Hz to 20 kHz) served as a target sound which was divided into five pulses (pulse duration 30 ms, duration of pauses 70 ms, 3-ms Gaussian-shaped slopes). The wide-band noise stimulus was chosen in order to provide the subject with interaural level (ILD) and time difference (ITD) cues used for localization, that are consistent throughout the audible frequency range. Further, the stimulus was pulsed to give information through interaural temporal envelope differences for localization (Middlebrooks and Green, 1991). The pause duration in pulsation was chosen to be long enough for most CI patients to allow for the detection of the gap (Busby and Clark, 1999).

The sound pressure level of the noise was randomly varied in 3-dB steps between 64 and 76 dB SPL. This roving level paradigm should prevent the subject from using information about the sound level obtained by the ear next to the speaker. This range of levels was selected to correspond to the level difference of 12 dB that produces a complete lateralization of wide-band noise for subjects with normal hearing (Blauert, 1997). Due to recruitment effects, the level could not be increased beyond 76 dB SPL. In some conditions subjects reported this stimulus to be very loud whereas few subjects had difficulties in detecting the soft stimulus with the monaural HA.

After the target presentation and a pause of 500 ms, a light spot appeared at 0° in front of the subject. The subject's task was to adjust the light spot with the trackball to the perceived direction of sound incidence. By pressing one of the trackball buttons the subject acknowledged the indicated direction and the light spot disappeared. The procedure was intuitive for all subjects. After a pause of 500 ms, the next trial started with the presentation of the target sound from a different angle. The presentation of all 11 directions formed a block. Between blocks an LED for the head alignment of the subject lit up for five seconds to prevent an incorrect head orientation. In total, subjects conducted ten blocks (110 trials=10 blocks×11 directions). To accomplish the roving level paradigm for each direction five different levels were randomly assigned to the ten repetitions, whereas every level step occurred twice. One experiment with a total of 110 trials was performed in about 11 min.

Prior to the experiment, a randomized presentation of every sound source direction was given to the subject. For a better familiarization with the experimental method, feedback was given in this part through an LED that showed the position of the correct response after pressing the input button.

The bimodal subject group performed the localization experiments in the conditions: CI only (captioned in Table III as CI), CI and HA in omni-directional mode (CI+HA P1), and CI and HA in directional mode (CI+HA P2). Due to the limited time, not all subjects were able to perform the experiment in the HA only condition (HA). Subjects AB and RL conducted the HA only condition 1.9 years later (cf. Table III). The experimental order of the conditions CI only, CI +HA, and HA only was varied between subjects.

The bilateral subject group performed the experiment in three conditions: CI bilateral (CI+CI), or monolateral on both sides (1st CI/2nd CI). Experimental order was varied. Due to time limitations RL conducted only the bilateral test.

Subjects BW and KH were additionally tested in a fixed level condition whereby instead of the roving level paradigm a fixed level of 70 dB SPL was used for the presentation of the stimuli. Subject BW performed a retest 1.3 years later (see Table IV for details).

III. RESULTS

A. Localization with bimodal aids

Figure 1 displays the individual localization results for all subjects. The median responses for the different conditions are depicted with different symbols [CI only (○), HA only (□), and the bimodal configuration CI plus HA (∗), HA with omnidirectional microphone], the errorbars show the quartile range.

A detailed statistical analysis of the localization results is provided in Table III. Purely error-based statistics are given in the first three columns: the absolute localization error calculated for all responses, the mean deviation of the

TABLE III. Statistical analysis of localization results of bimodal hearing aid users: absolute and arithmetical (relative) localization error, quartiles, correlation coefficient, coefficients of regression line, and test for side differentiation.

Subject	Condition	Error (°)		Quartiles (°)[c]	Correlation coefficient[d]	Regression line		Side differentiation[e]
		Absolute[a]	Arithmetical[b]			Slope	Offset	
HS	CI	22.2	10.2	11.4	0.51**	0.41	9.6	**
	CI+HA P1	10.3	8.5	4.5	0.94**	0.80	6.9	**
	CI+HA P2	9.4	6.7	5.1	0.95**	0.83	6.1	**
	HA P1	20.7	−8.9	15.7	0.42**	0.36	−11.0	*
DT	CI	27.3	−9.4	1.5	0.16	0.02	−9.5	+
	CI+HA P1	18.6	8.7	5.7	0.77**	0.36	6.5	**
PG	CI	20.6	10.0	12.4	0.46**	0.29	3.4	**
	CI+HA P1	13.8	−1.4	10.8	0.74**	0.76	−2.3	**
	HA P1	26.1	−9.1	4.8	−0.10	−0.02	−6.6	
BH	CI	31.9	14.5	19.9	0.01	0.01	10.5	
	CI+HA P1	28.4	−0.3	14.8	0.01	0.01	3.0	
	HA P1	36.1	−11.7	19.4	0.11	0.11	−9.6	
JJ	CI	27.6	−14.0	5.6	0.02	0.01	−13.9	
	CI+HA P1	29.4	−10.4	4.5	−0.21+	−0.07	−9.5	+
RM	CI	23.6	−7.4	7.8	0.21+	0.17	−8.2	
	CI+HA P1	33.4	−23.7	9.9	0.03	0.03	−21.8	
	HA P1	23.7	16.9	19.6	0.39**	0.43	20.4	**
EK	CI	20.7	−7.7	12.0	0.36**	0.26	−8.2	*
	CI+HA P1	19.1	0.0	13.2	0.44**	0.36	−0.1	**
	HA P1	24.4	8.3	13.1	0.06	0.03	3.9	
AB	CI	22.8	2.0	13.3	0.36**	0.25	−1.7	*
	CI+HA P1	26.2	−6.2	15.0	0.12	0.08	−4.9	
	HA[f]	28.4	−2.7	16.7	0.19+	0.16	−0.6	
RL	CI	22.3	9.3	13.4	0.08	0.06	8.6	
	CI+HA P1	19.3	4.1	7.8	0.32**	0.21	6.2	+
	HA[f]	22.6	−5.7	13.0	0.30*	0.17	−2.4	+
EM	CI	21.5	−7.6	13.6	0.34**	0.31	−10.7	+
	CI+HA P1	24.2	4.5	12.0	0.43**	0.50	−4.3	**
IS	CI	29.0	10.1	8.8	−0.04	−0.02	13.3	
	CI+HA P1	21.5	2.6	8.0	0.34**	0.36	1.4	**

[a]Absolute error: mean absolute deviation of single localization results from presented direction.
[b]Relative, arithmetical error: mean deviation of single localization results from presented direction.
[c]Average value of single quartiles.
[d]Significance of correlation coefficient: ** 0.001, * 0.01, +0.05.
[e]Differentiation of side of sound origin: Wilcoxon rank sum test on identity of the results of −50° and −40° (pooled) vs. +40° and +50° (pooled). Results for both sides are significantly different at ** 0.001, * 0.01, +0.05.
[f]Results of second session 1.9 years later.

responses from the presented direction (relative error), and the average size of a single quartile. Additionally, the correlation coefficient calculated between responses and presented positions is added in the next column. A linear regression has been carried out furthermore for each observer and condition to calculate the best fitting slopes. The results of this regression are included in terms of slope and offset. Although a single regression line cannot model some of the data, it shows how far the indicated positions follow the presented positions on average for the entire span. The last column of Table III presents a statistical measure for the discriminability of the side of sound origin. The level of significance is given by which the pooled results for −50° and −40° on the left side are not equal to the results for +50° and +40° on the right side (Wilcoxon rank sum test).

The individual localization results of the bimodal subject group in Fig. 1 differ to a large extent. While subjects BH, JJ, AB, and RM show no localization ability at all, subjects EM, IS, EK, and RL are able to discriminate the side of sound origin using both devices. Subjects DT and PG demonstrate limited localization ability, whereas subject HS is able to localize with high accuracy in the bimodal condition.

The left part of Fig. 2 displays the responses of subject HS in the monaural conditions CI only and HA only in detail. Even in the monaural conditions the response positions correlate relatively well with the presented positions (CI

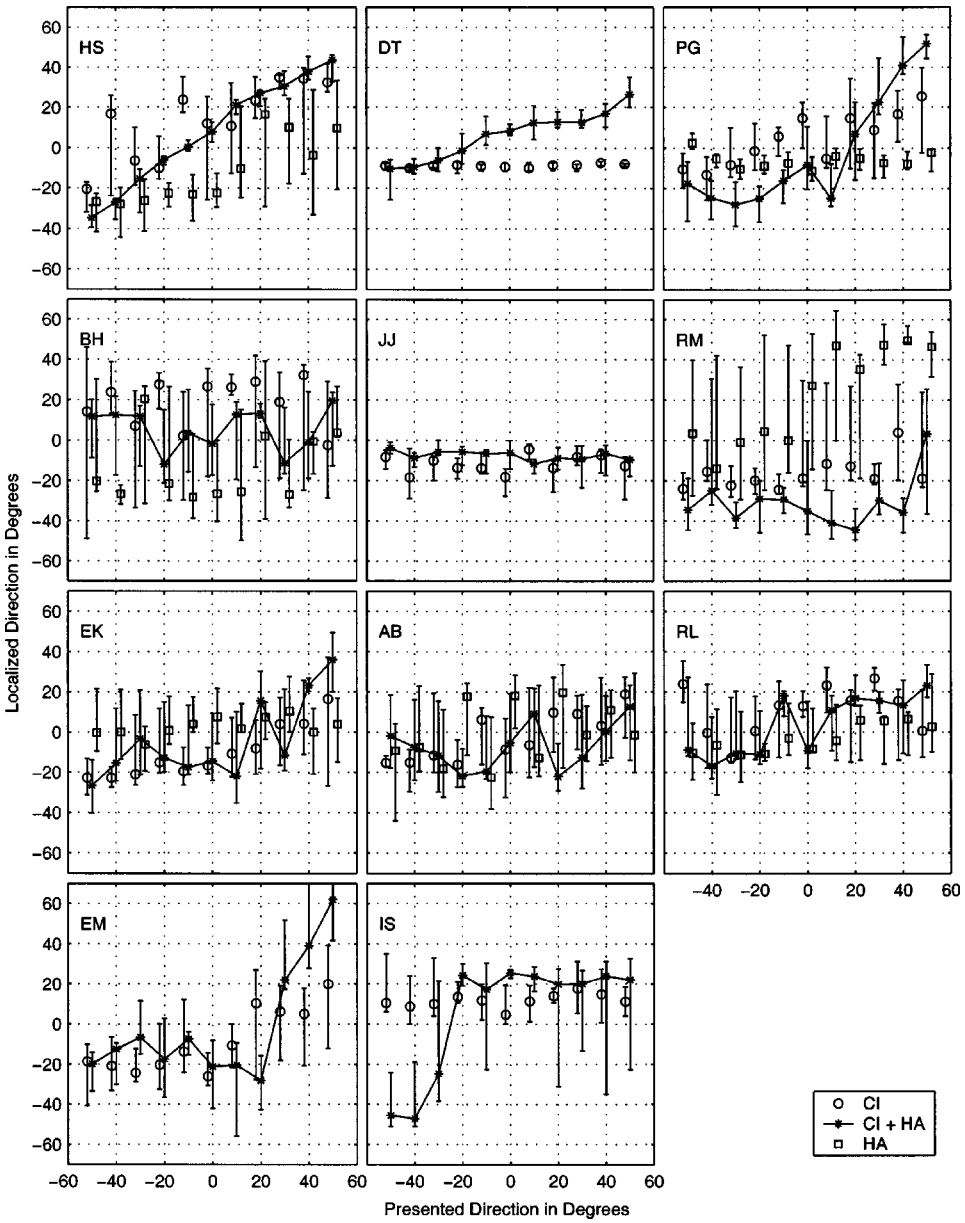


FIG. 1. Medians and quartile ranges of localization responses of the bimodal subject group. Conditions: CI only (○), CI+HA P1 omni (∗), HA only P1 omni (□).

only: 0.51, HA only: 0.42). The monaural information seems to be sufficient to allow for a rough discrimination of the side of sound origin, as can be deduced from the slope of the regression line and further statistical testing (last column of Table III).

The bimodal condition (Fig. 2, right) allows subject HS to respond with high accuracy to the presented position. In the bimodal condition the correlation of response position with presented position increases to 0.94 and the slope of the regression line to 0.8. The average quartiles decrease from

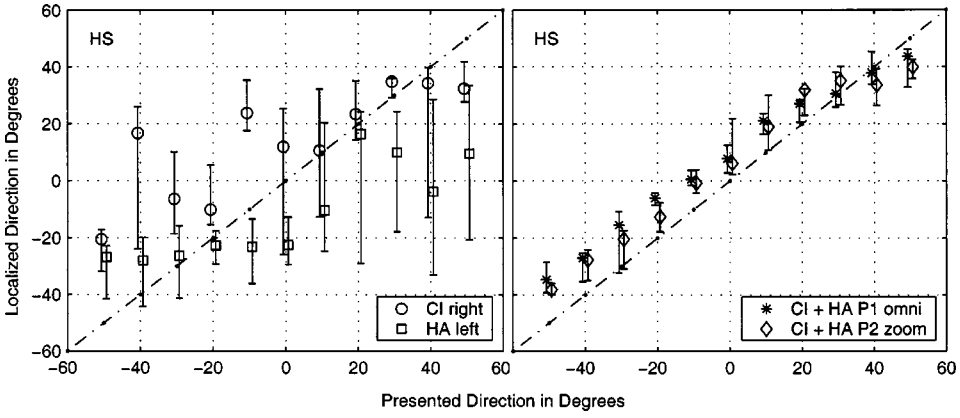


FIG. 2. Localization responses of subject HS. Left: monaural conditions (replotted from Fig. 1): CI only (○), HA only P1 (□). Right: impact of the directional microphone setting, bimodal condition: CI+HA P1 omni (∗) or CI+HA P2 zoom (◇). Medians and quartile ranges.

TABLE IV. Statistical analysis of localization results of bilateral cochlear implant users (CI): absolute and arithmetical (relative) localization error, quartiles, correlation coefficient, coefficients of regression line, and test for side differentiation. The data are calculated as in Table III.

Subject	Condition	Error (°)		Quartiles (°)	Correlation coefficient	Regression line		Side differentiation
		Absolute	Arithmetical			Slope	Offset	
BW	1st CI[a]	30.0	−17.3	25.7	0.36**	0.48	−8.6	*
	CI+CI	6.2	0.6	4.4	0.97**	1.15	−0.5	**
	CI+CI fix SPL[b]	4.9	−0.7	4.0	0.97**	1.06	−1.8	**
	2nd CI	25.9	8.3	19.4	0.14	0.15	5.3	
	2nd CI fix SPL[b]	24.5	10.3	15.9	0.02	0.02	3.3	
IB	1st CI	21.4	−8.4	14.5	0.39**	0.27	−6.8	+
	CI+CI	19.5	−15.4	6.3	0.75**	0.53	−16.4	**
	2nd CI	36.6	−36.6	10.0	−0.18	−0.09	−37.5	
KH	1st CI	38.6	−0.0	18.0	0.01	0.01	−1.1	
	CI+CI	16.9	0.0	8.4	0.72**	0.57	0.2	**
	2nd CI	28.1	10.7	15.5	0.21+	0.23	3.3	+
RL	CI+CI	17.3	−3.8	11.6	0.53**	0.45	0.4	*

[a]Results of second session 1.3 years later.
[b]Condition with fixed level of 70 dB SPL.

11.4° (CI only) or 15.7° (HA only) to 4.5° (CI+HA P1) and 5.1° (CI+HA P2) in the bimodal conditions. The response distributions for the monaural and binaural conditions differ significantly (Wilcoxon rank sum test on responses per direction, α-corrected for 11 directions, significant at 5% for the conditions: CI only vs. HA only, CI only vs. CI+HA P1, and HA only vs. CI+HA P1). The test for side discrimination is highly significant at $p<10^{-7}$. These results confirm the high localization ability of HS in the bimodal condition.

The localization ability of HS remains unchanged if the directional microphone setting (P2) is used instead of the omnidirectional program (P1) in the CI+HA condition (Wilcoxon rank sum test, α-corrected, *not* significant at 30%). As the directional microphone setting shows also no substantial influence on the localization results for the other subjects, further results with CI+HA P2 are not presented.

Subjects DT and PG show limited localization ability with reduced discrimination of sound source direction in the bimodal condition. The slope of the responses of DT is shallow (0.36) with low variance (CI vs. CI+HA P1 significant at 1%, side discrimination for CI+HA P1 at $p<10^{-7}$). In the CI only condition, subject DT is unable to indicate the side of sound origin properly, rather he responded to every stimulus by pointing to approximately −10° with a very small variance (average quartile of 1.5°). Subject PG distinguishes the side using the CI only but not using the HA only. Although the slope of the regression line is higher in the bimodal condition for subject PG, the high variability of the responses indicates limited localization capability (CI vs. CI+HA P1 and HA vs. CI+HA P1 significant at 1%, side discrimination for CI+HA P1 at $p<10^{-7}$).

Subjects EM, EK, IS, and to a limited extent RL, show some benefits in the bimodal condition. Using both devices they are able to discriminate the side of sound origin, especially for more lateral positions (side discrimination test for CI+HA P1 significant at 1% for EM, EK, IS; at 5% for RL; Wilcoxon rank sum test, α-corrected: RL and EK: 5% significance for HA vs. CI+HA P1; EK 5% significance for HA vs. CI). The results of subject EM also show a high regression slope (0.50) in the bimodal condition, but a discrimination of directions can only be seen between +20° and +50°. Subject EK, and with less accuracy EM and RL, are able to discriminate the side of the sound source in a monaural condition.

Subject RM shows a reduction in localization performance in the bimodal condition compared to the HA only condition. However, this reduction is probably not based on a reduction in localization ability; RM reported not being able to hear any directional change in any test condition. The data for RM indicate that this subject used a loudness criterion in the HA only condition; extremely loud sounds were indicated towards the HA side, whereas soft sounds were reported at the contralateral side. In this way, the applied roving level paradigm may have been overridden.

Subjects BH, JJ, RM, and AB were not able to improve their localization skills in the bimodal condition. None of the statistical parameters in Table III show a clear improvement in comparison to the HA only or CI only condition.

In summary, about half of the subjects showed improvement over unilateral performance, but one subject (HS) showed dramatic improvements and was able to perform nearly as well as normal-hearing subjects.

B. Localization with bilateral cochlear implants

The localization results of the bilaterally implanted subject group are presented in Fig. 3. Different symbols depict median responses for the monolateral conditions (either condition 1st CI or 2nd CI only) or the bilateral condition. Errorbars show quartile ranges. The results of the statistical analysis of the responses are shown in Table IV. Compared to the monolateral responses, all three subjects tested both monolaterally and bilaterally were better in the bilateral CI condition in terms of absolute error, quartile range, and correlation coefficient. The slope of the linear regression increased as well.

Using the first implanted CI, subject BW is able to differentiate the side of sound origin; the direction of sounds coming from the nonsupplied side can be roughly estimated, but sounds originating at the CI side are judged to the front. Using only the second implanted CI no side discrimination seems possible, but in the bilateral condition, BW's localization ability improves significantly (Wilcoxon rank sum test, α-corrected: 1st CI only vs. bilateral CI condition at 1%; 2nd CI only vs. bilateral CI condition at 5%). Compared to the results of the other three subjects of the bilateral group, BW's localization ability is very accurate. In the bilateral CI condition the quartile range is 4.4° and the averaged error amounts to 6.2° absolute, and 0.6° relative. The correlation between presented direction and localization response is as high as 0.97 using the roving-level procedure.

Figure 4 displays BW's localization results in the 2nd CI only and bilateral CI condition with fixed level presentation of the target sounds. As can be seen, this additional level cue does not improve localization in either condition. No change in the responses with and without the level rove is observed (Wilcoxon rank sum test, α-corrected: roving vs. fixed level presentation 2nd CI condition: $p>40\%$, bilateral CI condition: $p>10\%$).

Subject IB (Fig. 3, upper right) is also able to discriminate the side of sound origin in the 1st CI only condition. Like BW, IB did not show this ability in the 2nd CI only condition. In the bilateral CI condition, IB is able to perform minimal localization, which is reflected in an increase of the regression slope to 0.53 and also in a reduction of the quartiles from 10–15° to 6.3°. Furthermore, the correlation between presented direction and localization response increases from 0.39 to 0.75 (Wilcoxon, α-corrected, significant at 1% for 2nd CI only vs. bilateral CIs condition and 1st CI only vs. 2nd CI only condition).

Subject KH shows similar results as IB, although KH is not able to discriminate the side of sound origin in the 1st CI only condition. In the bilateral CI condition the average quartile range is substantially smaller than in the two monolateral CI conditions. The correlation between presented direction and localization response increases from 0.21 to 0.72. (Wilcoxon, α-corrected, significant at 5% for 1st and 2nd CI only vs. bilateral CI condition). RL's localization abilities are limited compared to the other subjects. His data shows the weakest correlation between presented direction and localization response (0.52), but he is still able to discriminate the side of sound origin ($p=0.01$).

To summarize, all four subjects show the ability to localize sounds with bilateral CIs, whereas subject BW shows localization ability close to normal-hearing subjects.

IV. DISCUSSION

A. Bimodal aids

The localization results of the bimodal subject group demonstrate an unexpected high localization ability in several subjects. In particular, subject HS shows an excellent localization ability. This is to our knowledge the first report of such highly skilled localization abilities in a subject with a cochlear implant in one ear and a hearing aid in the opposite ear. This is reflected by small average quartiles for his localization responses (4.5°) and by a strong correlation between presented direction and localization response (0.94). The average quartiles are close to the accuracy of normal-hearing subjects (1.7°), as reported by Seeber (2002), obtained with the same experimental setup. Compared to HS's outstanding accuracy, DT and PG show poorer localization performance, although it is well above chance. Subjects EM, EK, IS, and RL can discriminate the side of sound origin correctly, whereas BH, JJ, AB, and RM respond without any relation to the presented sound direction. Considering monolateral performance, five subjects, HS, PG, EK, AB, and EM, discriminated the side of sound origin using the CI alone, whereas only HS did so reliably using the HA alone.

Subjects HS, DT, and PG especially benefited from the bimodal configuration. These subjects have the highest amount of residual hearing and the smallest PTA hearing loss in our bimodal subject group (Table I). Audiograms were collected for all of our subjects to assess the amount of residual hearing on the implanted side. The majority of the subjects including subject HS showed no residual hearing. Several subjects showed a ''left corner audiogram'' (70 dB HL at 125 Hz, 120 dB at 500 Hz). At these high sound levels the mechanical vibration of the headphone and the skull might lead to a sensation of vibration which is often difficult to differentiate from residual hearing. In any case, it seems impossible that the subjects might have used residual hearing at the implanted side during the experiment. This is because the presentation level of the test signals was below their residual hearing threshold. Somewhat surprising is the fact that three subjects (IS,EM,RL), who are able to differentiate sides, are the subjects with the worst PTAs >105 dB. The data are consistent with the hypothesis that a high amount of residual hearing is a prerequisite for bimodal localization although not necessary for a bimodal advantage in terms of side discrimination.

Tyler *et al.* (2002b) showed an improvement of side discrimination with bimodal aids in two of three subjects in a source identification task using two speakers at $\pm 45°$ positions. The subject without improvement in the bimodal condition performed already above chance using the implant alone, whereas the other two subjects were at chance level in the monolateral condition. As the task involved only two speakers, the localization ability of the subjects cannot be inferred in terms of error and variance. Although the study reported here used a more demanding localization task the results for side discrimination of our bimodal subject group seem to be in line with the results obtained by Tyler *et al.* The statistical test for side discrimination showed significant results at a 0.1% significance level in 6 out of 11 subjects of our bimodal subject group.

B. Bilateral cochlear implants

The bilateral CI subject group demonstrated localization abilities with varying accuracy and one top performing subject (BW). BW's localization accuracy in the bilateral CI condition, as shown with our experimental setup, is to our

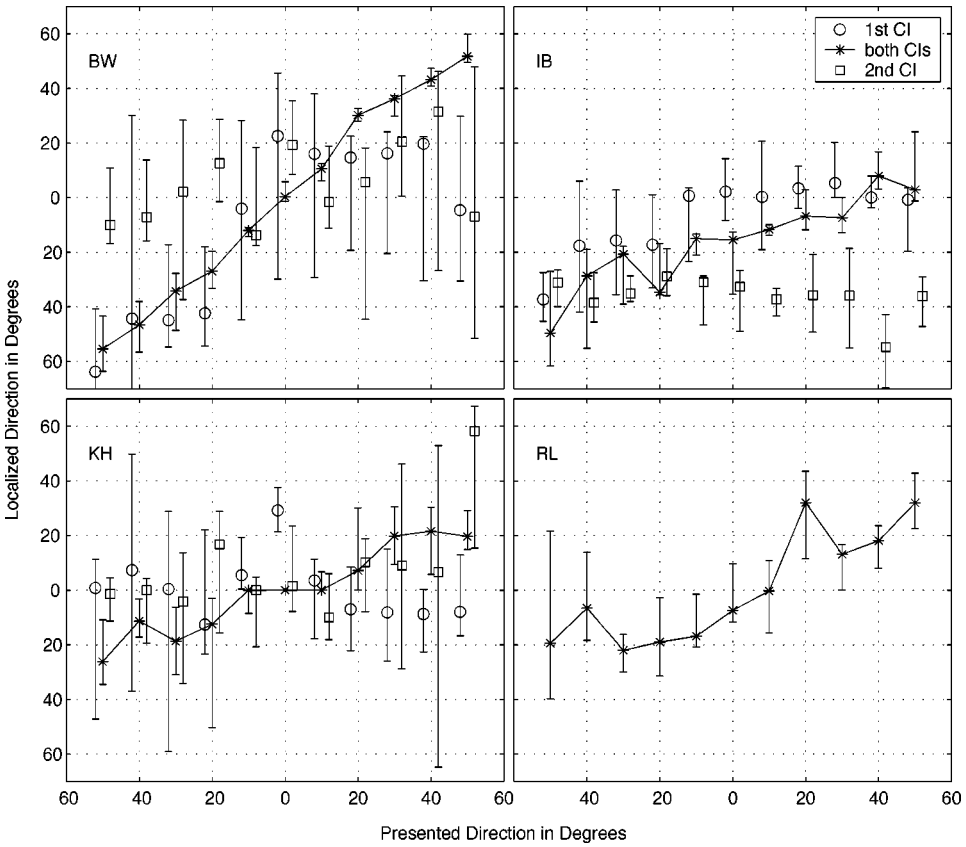


FIG. 3. Medians and quartile ranges of localization results of the bilateral CI subject group for the conditions 1st CI only (○), bilateral CIs (∗), or 2nd CI only (□).

knowledge unparalleled in the CI literature. The average quartiles of BW's localization responses amount to 4.4° and the slope of the regression is 1.15.

Side discrimination tests done by the Iowa group show a bilateral benefit for most of their bilateral CI users. All subjects in one group showed a nearly 100% correct score for the identification of speakers at ±45° (Gantz *et al.*, 2002). In a similar study the performance of six out of seven subjects increased with bilateral CIs (Tyler *et al.*, 2002a). A source identification study with only one subject of van Hoesel *et al.* (2002) in an 11-speaker array showed a level-dependent localization rms-error of about 16° at 70 dB SPL and 8° at 60 dB SPL. The standard deviation was 18° and 13°, respectively. In the monolateral CI conditions the subject was not able to discriminate the side of sound origin. A more recent identification study of van Hoesel and Tyler (2003) showed varying localization ability in five subjects. Average rms errors were about 10°, with one top performing subject who was able to identify the two frontal speakers at 2.5° rms-error. Three of five subjects demonstrated a limited localization ability in the monolateral CI condition. However, the remarkable localization accuracy reported by van Hoesel and Tyler (2003) should not be interpreted in terms of a characterization of the parameters of the underlying localization process, i.e., variance, error, and bias, as the distance of the loudspeakers in the array was much higher (15.5°) than the observed error/variance (down to 2.5°). Using a numerical model of identification methods, Hartmann *et al.* (1998) found that the experimentally observed localization error can be significantly smaller than the actual error of the localization process if the variance of the localization process is small compared to the source's separations (cf. their Fig. 1). According to their simulation the observed rms-error will underestimate the true rms-error by 30% for a source spacing

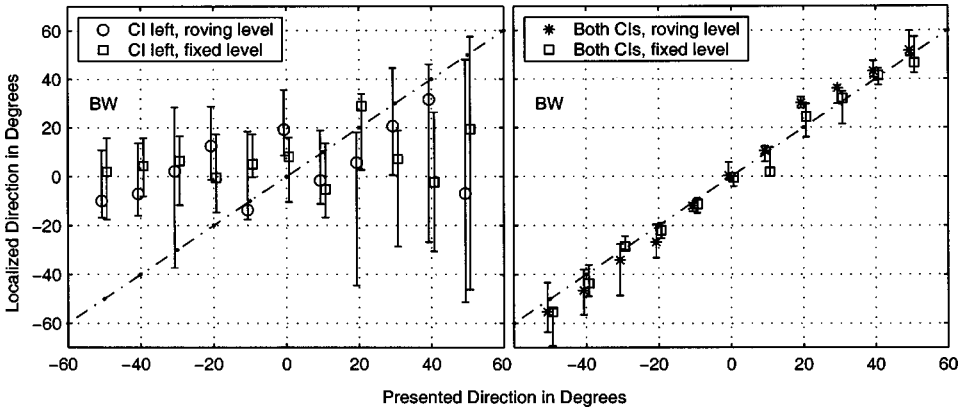


FIG. 4. Influence of the roving level paradigm in the localization results of subject BW. Left: 2nd CI only condition for the roving (64–76 dB SPL, ○) and the fixed (70 dB SPL, □) level condition. Right: bilateral CI condition roving (∗) and fixed (□) level presentation. Medians and quartile ranges.

about four times the error, e.g., for a 15.5° spacing and errors about 3.5° as in van Hoesel and Tyler (2003). The simulation has further shown that in some conditions with bias, the characteristics of the auditory system cannot be recovered from the observed localization responses [cf. Fig. 6 in Hartmann *et al.* (1998)]. Because bias effects are common in CI studies (e.g., with monaural devices or through slightly different adjustment of the AGCs or the amplification of both devices), the results from identification methods should be interpreted with caution. Thus, currently used source identification methods with a fairly large speaker spacing cannot characterize the localization ability of top performing subjects. On the contrary, the continuous localization method employed in our study allows a more straightforward interpretation of the localization data in terms of error and variance as the spacial resolution of the responses is virtually unlimited.

The outstanding localization results of the bilateral CI subject BW might be partly due to an intense localization training in his daily life. BW works as a teacher in a secondary school and needs to address his pupils in response to their comments even when standing at the blackboard with his back to them.

C. Utility of available cues

The observed improvement of localization abilities in the bimodal or bilateral conditions compared to the monolateral conditions indicates the capability of the auditory system to use degraded information from binaural stimulation for localization. Since the overall level is randomized, this information could be derived from (1) monaural spectral information at each ear, (2) interaural time differences (ITDs), and (3) interaural level differences (ILDs).

1. Monaural information

The results of the tests conducted in the monolateral conditions showed that monaural spectral information is sufficient only for the detection of the side of sound origin. Only one subject (HS) was able to use monaural cues in both monolateral conditions. Thus monaural spectral information at two ears is an unlikely cue to explain the high localization accuracy observed in our best performing subjects.

2. ITD cues

Subject BW of our study shows a precision of localization of 4.4° (quartiles) which corresponds to an ITD of about 40 μs (Feddersen *et al.*, 1957).

The mapping of temporal information through cochlear implants is influenced by the pulsatile stimulation employed. The CIS strategy samples the temporal envelope information in frequency bands at a fixed, continuous pulse rate. A population of spiral ganglion cells is stimulated phase-locked to the pulses with an interpulse-interval of about 660 μs or multiples thereof. In normal hearing instead, the discharge pattern on the auditory nerve is phase-locked to certain phases of the sound stimulus (Klinke and Hartmann, 1997). Therefore the electric pulsatile stimulation does not allow a direct temporal coding of phase-related ITD. A temporal change in the envelope in a certain frequency channel, however, will still maintain an amplitude change in the stimulating current. Due to the refractory characteristic of ganglion cells, this will result in a temporal correlation of activity in the auditory nerve to the maximum of the temporal envelope of the band-pass filtered acoustic signal. Assuming that differences of the interaural temporal envelope are integrated over several stimulus pulses, localization could be based on this process in the bimodal and bilateral conditions. An integrative behavior in terms of envelope analysis could overcome the limitations in ITD mapping due to unsynchronized stimulation in the bilateral CI condition. The expected sensitivity to envelope ITDs would therefore depend on the pulse rate and the mechanism that is hypothesized. In the bimodal condition different processing times of CIs and HAs will result in offsets and additionally affect the interpretation of ITDs. However, if these offsets are small, the subjects could potentially adapt their localization to this offset (Gold and Knudsen, 2000; Hofman *et al.*, 1998; Shinn-Cunningham, 2001; Welch, 1986).

Lawson *et al.* (2001) demonstrated the ability to detect changes in envelope-ITDs in unsynchronized pulse trains as low as 25 μs in two of their bilateral CI subjects. Van Hoesel and Tyler (2003) in contrast measured with their subjects detection thresholds of 120 μs for ITDs in envelope and pulse train. For synchronized pulse trains containing a purely envelope-based ITD the threshold rose to 290 μs (one subject).

The ITD detection thresholds for pulse trains using synchronized processors vary highly. Early studies found thresholds of 0.5 to 1.5 ms (van Hoesel and Clark, 1997), whereas the best thresholds are often lower in more recently published studies: 150 μs (Lawson *et al.*, 1998), 90–180 μs (van Hoesel and Tyler, 2003), and 50–150 μs for selected electrode combinations (Lawson *et al.*, 2000). Subject BW also participated in the study of Lawson and co-workers (2000). He showed ITD JND's down to 50 μs for one electrode combination. On most electrode pairs, however, ITD JND's were raised to several hundred microseconds (median 500 μs). In our study, the precision of localization of subject BW would correspond to about 40 μs (quartiles 4.4°, cf. Feddersen *et al.*, 1957) and is thus nearly a magnitude more exact than the average detection threshold with synchronized stimulation. Wightman and Kistler (1996) emphasize that the binaural system always refers to the most reliable available cue. Obviously, the auditory system could base localization on the ITD-evaluation at few electrode combinations with good ITD-detection thresholds, but as ITD-detection thresholds are highly different on different electrodes it seems more likely that this inconsistent ITD information is completely ignored as unreliable. In fact, in a subsequent study BW was not able to localize pulsed low-pass noise that provided ITD cues through evaluation of the interaural phase and the envelope of the signal. In contrast, high-pass noise with a slow envelope rise-time of 200 ms was localized. This emphasizes the importance of evaluation of ILD cues (Seeber, 2003).

3. ILD cues

The CI stimulates regions corresponding to relatively high characteristic frequencies due to its limited insertion depth. The sensorineural hearing loss limits the usable frequency range on the contralateral side fitted with the HA instead to lower frequencies (Table I). Only a limited overlapping frequency range can thus be used for the processing of binaural cues in the bimodal condition. For subjects of the bimodal group with PTAs $>$79 dB hearing loss (7 of the 11 subjects), the hearing aid cannot provide sufficient amplification for frequencies $\geqslant$2 kHz. All these subjects show no localization ability, which might be due to a limited availability or the distortion of ILDs at higher frequencies. Subjects JJ, EM, and EK show a decrease of their hearing loss with increasing frequency. Therefore they might be able to evaluate ILDs. Actually, EM and EK were able to discriminate the side of sound origin.

HS, DT, and PG had considerably lower hearing loss than the other subjects of the bimodal group and the hearing aid allowed a partial loudness compensation. They were able to localize. If the localization process in the bimodal condition relies on ILDs, it has to compare the signals from two different stimulation schemes for intensity. To compensate for the limited pitch matching at both sides, an integrative behavior to form a gross-ILD over several frequency channels could be employed (Macpherson and Middlebrooks, 2002).

Concerning the bilateral CI condition, it has been reported that the ILD-detection thresholds are generally low. The neural ILD thresholds are usually in the range of 1–3 current units, which is equal to 16 μA (0.125 dB) at 1.1 mA (9.5 dB) dynamic range (Lawson *et al.*, 2000, 1998). Since the internal signal processing and generation of the receiver is digital, the resolution of the current pulse amplitude is limited by the size of the digital steps. In the case of the MED-EL Combi40+ device, this resolution is 7 bit linear providing 128 current steps resulting in a current amplitude range of 42 dB. The minimal current step size is adjusted for each electrode channel by the fitting program to either 2.4, 4.2, 7.7, or 13.6 μA depending on the specific current pulse amplitude a patient requires for gaining a comfortable loudness sensation (MED-EL, 2000). Since this fitting is set for each channel individually, a situation can occur where a CI user can detect a small ILD in one channel, whereas the same difference presented on another channel cannot be detected (Lawson *et al.*, 2000).

The preprocessing in the speech processor maps the large acoustical dynamic range to the smaller electrical dynamic range at electrode level. This compression is realized by an analog compression scheme as well as by a mapping of the channel envelope information to current levels according to a logarithmic function. Assuming the same compression ratio is employed in CI processor and the HA, it is obvious that ILDs are reduced. More complicated, the compression change at the onset level (kneepoint) which marks the start of the compressive behavior may also distort ILDs. This distortion occurs if compression is employed on one side while the head shadow effect damps the signal below kneepoint-level on the other side. As a result, the ILD is dependent on the input level, which might affect localization ability (van Hoesel *et al.*, 2002).

To estimate the usefulness of ILD cues for localization the level transformation in the processor has to be considered. If a 3:1 compression of the analog dynamic range of about 40 dB to an electrical dynamic range of 10–20 dB is assumed, an acoustical ILD detection threshold of about 0.4 dB follows for an internal ILD threshold at the nerve of 0.125 dB (Lawson *et al.*, 2000, 1998). For changes of the physical overall ILD of 0.1 dB/° (van Hoesel *et al.*, 2002) a theoretical minimum audible angle of 4° results. This is well in line with the results of top performing subject BW (quartiles 4.5°).

V. CONCLUSIONS

We have shown that a successful restoration of the localization ability in the frontal horizontal plane in deaf patients is possible by means of bilateral cochlear implantation or, in the case of sufficient residual hearing for one ear, by means of a bimodal fitting with a cochlear implant on one ear and a hearing aid on the other ear. The best performing subjects showed an accuracy (average quartiles of localization responses 4.4°) near to normal-hearing subjects (quartiles 1.7°). We believe that ILDs are the most likely cue to enable this high localization precision, when ITD information is restricted to the interaural time differences in the envelope of the signal.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft DFG GRK 267, the Fördergemeinschaft Deutscher Hörgeräteakustiker, and Phonak G.m.b.H. Switzerland. We would like to thank Steve Colburn and one anonymous reviewer for their constructive comments which greatly improved the manuscript. We thank Poppy Crum for language refinements on the final manuscript.

Baumann, U. (**2001**). "Sprachverständnis mit Cochlea-Implantat und Richtmikrofon-Hörgerät (Speech reception with cochlear implant and hearing aid with directional microphone)," in *Fortschritte der Akustik—DAGA '01*, Oldenburg. DEGA. CD-Rom, ISBN 3-9804568-9-7.

Blauert, J. (**1997**). *Spatial Hearing* (MIT, Cambridge).

Busby, P., and Clark, G. (**1999**). "Gap detection by early-deafened cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 1841–1852.

Dooley, G., Blamey, P., Seligman, P., Alcantara, J., Clark, G., Shallop, J., Arndt, P., Heller, J., and Menapace, C. (**1993**). "Combined electrical and acoustical stimulation using a bimodal prosthesis," Arch. Otolaryngol. Head Neck Surg. **119**, 55–60.

Durlach, N., Thomson, C., and Colburn, H. (**1981**). "Binaural interaction in impaired listeners," Audiology **20**, 181–211.

Feddersen, W., Sandel, T., Teas, D., and Jeffress, L. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.

Gantz, B., Tyler, R., Rubinstein, J., Wolaver, A., Lowder, M., Abbas, P., Brown, C., Hughes, M., and Preece, J. (**2002**). "Binaural cochlear implants placed during the same operation," Otolo. Neurotol. **23(2)**, 169–180.

Gold, J., and Knudsen, E. (**2000**). "Abnormal auditory experience induces frequency-specific adjustments in unit tuning for binaural localization cues in the optic tectum of juvenile owls," J. Neurosci. **20(2)**, 862–877.

Grantham, D. W. (1995). "Spatial hearing and related phenomena," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 297–345.

Gray, R., and Baguley, D. (**1993**). "Sound localization with the Ineraid cochlear implant," Adv. Oto-Rhino-Laryngol. **48**, 174–176.

Hartmann, W., Rakerd, B., and Gaalaas, J. (**1998**). ''On the source-identification method,'' J. Acoust. Soc. Am. **104**, 3546–3557.
Henning, G. (**1974**). ''Detectability of interaural delay in high-frequency complex waveforms,'' J. Acoust. Soc. Am. **55**, 84–90.
Hofman, P. M., Van-Riswick, J. G., and Opstal, A. J. V. (**1998**). ''Relearning sound localization with new ears,'' Nature Neurosci. **1(5)**, 417–421.
Keidser, G., Dillon, H., and Byrne, D. (**1996**). ''Guidelines for fitting multiple memory hearing aids,'' J. Am. Acad. Audiol **7(6)**, 406–418.
Klinke, R., and Hartmann, R. (**1997**). ''Basic neurophysiology of cochlear-implants,'' Am. J. Otol. **18**, S7–S10.
Lawson, D., Brill, S., Wolford, R., Wilson, B., and Schatzer, R. (**2000**). ''Speech processors for auditory prostheses,'' Technical report, Research Triangle Park: Research Triangle Institute. Ninth Quarterly Progress Report.
Lawson, D., Wolford, R., Brill, S., Schatzer, R., and Wilson, B. (**2001**). ''Speech processors for auditory prostheses,'' 12th quarterly progress report, Research Triangle Park: Research Triangle Institute.
Lawson, D., Wilson, B., Zerbi, M., van-den Honert, C., Finley, C., Farmer, J. J., McElveen, J., and Roush, P. (**1998**). ''Bilateral cochlear implants controlled by a single speech processor,'' Am. J. Otol. **19(6)**, 758–761.
Luntz, M., Brodsky, A., Hafner, H., Shpak, T., Feiglin, H., and Pratt, H. (**2002**). ''Sound localization in patients with cochlear implant—preliminary results,'' J. Pediatric Oto-Rhino-Laryngol. **64**, 1–7.
Macpherson, E., and Middlebrooks, J. (**2002**). ''Listener weighting of cues for lateral angle: The duplex theory of sound revisited,'' J. Acoust. Soc. Am. **111**, 2219–2236.
MED-EL (**2000**). ''Clinical manual volume 3,'' Technical report, MED-EL G.m.b.H., Innsbruck, Sec. 10-9.
Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.
Müller, J., Schön, F., and Helms, J. (**2002**). ''Speech understanding in quiet and noise in bilateral users of the MED-EL COMBI 40/40+ cochlear implant system,'' Ear Hear. **23**, 198–206.
Seeber, B. (**2002**). ''A new method for localization studies,'' Acust. Acta Acust. **88(3)**, 446–450.
Seeber, B. (2003). ''Untersuchung der auditiven Lokalisation mit einer Lichtzeigermethode (Investigation of auditory localization using a light-pointer method),'' dissertation, TU München. http://tumb1.biblio.tu-muenchen.de/publ/diss/ei/2003/seeber.html.
Shinn-Cunningham, B. (**2001**). ''Models of plasticity in spatial auditory processing,'' Audiol. Neuro-Otol. **6(4)**, 187–191.
Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C., and Arndt, P. L. (**1994**). ''Evaluation of a new spectral peak coding strategy for the nucleus 22 channel cochlear implant system,'' Am. J. Otol. **15**, 215–227.
Tyler, R., Gantz, B., Rubinstein, J., Wilson, B., Parkinson, A., Wolaver, A., Preece, J., Witt, S., and Lowder, M. (**2002a**). ''Three-month results with bilateral cochlear implants,'' Ear Hear. **23**, 80S–89S.
Tyler, R., Parkinson, A., Wilson, B., Witt, S., Preece, J., and Noble, W. (**2002b**). ''Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization,'' Ear Hear. **23**, 98–105.
van Hoesel, R., and Tyler, R. (**2003**). ''Speech perception, localization, and lateralization with bilateral cochlear implants,'' J. Acoust. Soc. Am. **113**, 1617–1630.
van Hoesel, R., Ramsden, R., and O'Drisoll, M. (**2002**). ''Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user,'' Ear Hear. **23**, 137–149.
van Hoesel, R. J. M., and Clark, G. M. (**1997**). ''Psychophysical studies with two binaural cochlear implant subjects,'' J. Acoust. Soc. Am. **102**, 495–507.
Wagener, K., Brand, T., and Kollmeier, B. (**1999a**). ''Entwicklung und Evaluation eines Satztests in deutscher Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a sentence test in German language II: Optimization of the Oldenburg sentence test),'' Z. Audiol. **38**, 44–56.
Wagener, K., Brand, T., and Kollmeier, B. (**1999b**). ''Entwicklung und Evaluation eines Satztests in deutscher Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test in German language III: Evaluation of the Oldenburg sentence test),'' Z. Audiol. **38**, 86–95.
Wagener, K., Kollmeier, B., and Kühnel, V. (**1999c**). ''Entwicklung und Evaluation eines Satztests in deutscher Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test in German language I: Design of the Oldenburg sentence test),'' Z. Audiol. **38**, 4–15.
Welch, R. (**1986**). ''Adaptation of space perception,'' in *Handbook of Perception and Human Performance, I: Sensory Processes and Perception*, edited by K. Boff and L. Kaufman (Wiley, New York), Chap. 24.
Wightman, F., and Kistler, D. (**1996**). ''Factors affecting the relative salience of sound localization cues,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, Mahwah, NJ), pp. 1–23.
Zierhofer, C. M., Hochmair-Desoyer, I., and Hochmair, E. S. (**1995**). ''Electronic design of a cochlear implant for multichannel high-rate pulsatile stimulation strategies,'' IEEE Trans. Rehabil. Eng. **3(1)**, 112–116.

Flow separation in a computational oscillating vocal fold model

Fariborz Alipour[a)]
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

Ronald C. Scherer
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242 and
Department of Communication Disorders, Bowling Green State University, Bowling Green, Ohio 43403

(Received 1 October 2003; revised 9 June 2004; accepted 15 June 2004)

A finite-volume computational model that solves the time-dependent glottal airflow within a forced-oscillation model of the glottis was employed to study glottal flow separation. Tracheal input velocity was independently controlled with a sinusoidally varying parabolic velocity profile. Control parameters included flow rate (Reynolds number), oscillation frequency and amplitude of the vocal folds, and the phase difference between the superior and inferior glottal margins. Results for static divergent glottal shapes suggest that velocity increase caused glottal separation to move downstream, but reduction in velocity increase and velocity decrease moved the separation upstream. At the fixed frequency, an increase of amplitude of the glottal walls moved the separation further downstream during glottal closing. Increase of Reynolds number caused the flow separation to move upstream in the glottis. The flow separation cross-sectional ratio ranged from approximately 1.1 to 1.9 (average of 1.47) for the divergent shapes. Results suggest that there may be a strong interaction of rate of change of airflow, inertia, and wall movement. Flow separation appeared to be ''delayed'' during the vibratory cycle, leading to movement of the separation point upstream of the glottal end only after a significant divergent angle was reached, and to persist upstream into the convergent phase of the cycle. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779274]

PACS numbers: 43.70.Aj, 43.70.Bk [AL] Pages: 1710–1719

I. INTRODUCTION

The mechanics of human phonation involves the coupling of the vibration of the vocal folds with the airflow through the glottis. The pressurized air from the lungs and the biomechanical characteristics of the vocal folds drive the vocal folds into oscillation. The vocal folds modulate the flow of air from the trachea, turning the airflow into cyclic pulsatile jet flow exiting the glottis.

During the closing portion of the phonatory cycle when the glottal walls take on a diverging shape, flow separation is predicted to occur somewhere along the glottal walls (Guo and Scherer, 1993; Alipour *et al.*, 1996a, b; Pelorson *et al.*, 1994; Scherer *et al.*, 2001; Shinwari *et al.*, 2003; Hofmans *et al.*, 2003; Kucinschi *et al.*, in review). It is generally accepted that the flow separation point in a divergent duct is a primary factor in the determination of the pressures of that duct. Flow separation also has been found to be a positive factor in the lowering of the phonation threshold pressure that helps make it easier to create self-oscillation of the vocal folds (Lucero, 1998).

The theoretical notion of the Bernoulli equation that relates the pressure, area, and volume velocity at each cross-sectional area of a physical duct and the quasi-steady approximation of the glottal flow may no longer hold once the flow has separated from the walls. It seems highly likely that the effective glottal (flow) duct, then, becomes nearly uniform for a short distance following the separation point. If this held, then one might assume that the wall pressures past the separation point would equal the pressure in the downstream reservoir just past the glottal exit. This is the basic assumption used in some phonatory models (Titze, 1988; Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998). However, it is now known that the pressure just downstream of the point of separation may continue to rise until reaching a plateau value (Senoo and Nishi, 1977; Scherer *et al.*, 2001; Alipour and Scherer, 2002; Shinwari *et al.*, 2003).

Flow separation from a wall of the glottis takes place when the skin friction or shear stress on the wall approaches zero. The shear stress, a frictional effect exerted by the layers of the fluid on each other, is proportional to the viscosity and the derivative of the flow velocity across the layers (normal to the wall). The pulsatile nature of the glottal flow causes the shear stress distribution to change with time in contrast to that found with constant flow through static models. Since the peak of the shear stress occurs at locations of steep velocity gradients and wall bends, an oscillatory flow with tissue movement may generate high peak values on the walls.

When separation occurs, a slightly lowered pressure region is generated near the flow separation point and the air downstream of the separation point between the glottal wall and the jet flow will circulate at a lower rate of flow. This flow separation may influence the driving pressures on the glottal walls. In a static geometry, flow separation can easily be predicted through the solution of the Navier–Stokes equations. It has been reported that Reynolds number, divergence angle, pressure gradient, and flow deceleration are major determinants of the flow separation location (Senoo and Nishi, 1977; Ojha *et al.*, 1989).

[a)]Electronic mail: alipour@blue.weeg.uiowa.edu

 © 2004 Acoustical Society of America

Pulsatile flow with flow separation has received attention in numerous areas, such as in arterial stenosis. For example, the ultrasonic Doppler shift has been used for the analysis of moving flow separation in arterial stenoses (Tamura and Fronek, 1990; Siouffi *et al.*, 1998). However, an accurate measurement of the wall shear stress is very difficult and an estimated error of 20%–50% can be expected for pulsatile flows (Deplano and Siouffi, 1999). Therefore, in general, computational models may aid in determining important aspects of flow separation. Numerical modeling has been used to investigate pulsatile flow in models of heart valves, carotid arterial bifurcation, and arterial stenosis, where data on oscillatory shear stress and flow separation have been reported (e.g., Perktold *et al.*, 1998; Zhao *et al.*, 2000; Long *et al.*, 2001; Lai *et al.*, 2002). Glottal flow is more complicated than blood flow in arterial stenoses. Both the angle and diameter of the glottis continually change during the open portion of the vibratory cycle, and flow oscillates on the order of 100 times faster. Also, glottal walls oscillate with large amplitudes, and the possible glottal closure within each cycle adds further complication to the analysis and prediction of flow separation locations. In terms of nondimensional parameters, the Strouhal number reflects unsteadiness. The Strouhal numbers for phonation are approximately two to ten times larger than for blood flow (Alipour *et al.*, 1995; Hofmans *et al.*, 2003; Tutty, 1992).

Pelorson *et al.* (1994) reported a quasi-steady model of flow separation that allowed for a moving separation point as the vocal folds oscillated. They based their model on a combination of the unsteady Bernoulli equation and the von Karman equation using the Polhausen cubic polynomial approximation (White, 1974). They reported that at the beginning of the closing cycle, the flow separation point moves upstream and during the last part of the closing (prior to closure) moves downstream. Zhao *et al.* (2002) showed moving separation points in their computational model of forced vocal fold motion, but did not quantify the separation point location. Alipour *et al.* (1996a) and Alipour and Scherer (1998) used computational fluid dynamics and forced oscillatory glottal models and reported asymmetric flow separation within the glottis, as did Shinwari *et al.* (2003).

In the study here, we present a forced-oscillation glottal model in which the locations of the separation points are calculated by solving the time-dependent Navier–Stokes equations and are tracked for a variety of static and vibratory conditions. The model predicts the location of the separation point as a function of the amplitude of glottal motion, flow rate, and phase difference between the inferior and superior glottal margins, conditions that have not been pursued in earlier studies. The outcome of the study should lend conceptual guidance to the notion of the location of glottal flow separation points for phonation.

II. METHODS

A two-dimensional CFD computer code using a pressure-based finite volume method, developed at the Iowa Institute of Hydraulic Research (University of Iowa), was used for the numerical modeling (acoustic effect was not included). Lai and Przekwas (1994), Lai *et al.* (1995), and Lai and Alipour (2002) have described the details of this numerical model. In this method the computational mesh moves freely to conform to the wall motion and provides a solution for the flow with a moving boundary. An unstructured grid technique was adopted in order to use a multiblock-based mesh to represent the geometry of the glottis.

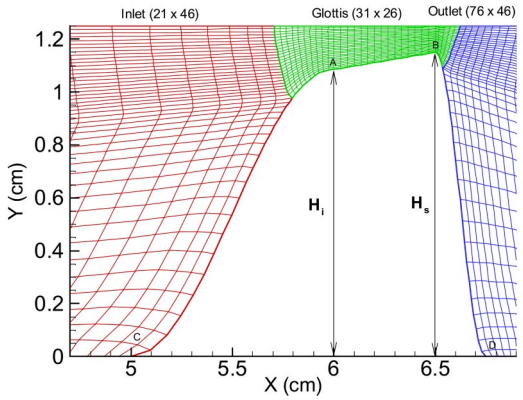


FIG. 1. Design of the oscillating vocal folds and portions of the grids.

A. Design of the flow channel

The channel height was 2.5 cm with a 5-cm inlet duct length and a 50-cm outlet duct length. The vocal fold model had an equilibrium height of 1.15 cm with independent amplitudes ranging from 0 to 0.9 mm. An upstream symmetric flow condition was assumed and only one-half of the channel was discretized for computation. Thus, the glottis moved symmetrically about the midline, and the flow through the larynx was also symmetric. The computational grid was divided into three blocks. Block 1 (inlet) had 21×46 grid points, block 2 (glottis) had 31×26 grid points, and block 3 (outlet) had 76×46 grid points (see Fig. 1).

B. Design of the oscillating glottal walls and the specification of the tracheal flow

The glottal surface was modeled with two sinusoidal curves connected with a tangent straight line. The heights H_i and H_s defined the inferior and superior glottal openings, respectively (Fig. 1). The glottal walls oscillated (were ''forced'') sinusoidally at the inferior and superior margins. The phase between the two locations and the amplitude of each location were controlled. The time variations of the inferior and superior glottal heights were

$$H_i = H_0 + A_i \cos(2\pi f t + \phi), \tag{1}$$

$$H_s = H_0 + A_s \cos(2\pi f t), \tag{2}$$

where H_0 was the equilibrium height value of 1.15 cm from the lateral wall for both the inferior and superior glottal locations, A_i and A_s were the amplitudes of oscillation at the inferior and superior glottal openings, respectively, t was time, f was the frequency of oscillation, and ϕ was the phase lead of the inferior opening relative to the superior opening. At time zero, the glottal diameter of the superior margin was smallest. There was no completely ''closed'' portion to the vibratory cycle.

The flow was not governed by a constant transglottal pressure, but was independently controlled. The flow was synchronized in a sinusoidal fashion with the motion of the glottal walls. The inlet (tracheal) mean velocity (U_T) at the $X=0$ location was set such that at the beginning of the cycle, which occurred when the superior margin was at its minimal diameter, the velocity was zero. The inlet velocity was in phase with the diameter of the upper glottal margin, viz.,

$$U_T = 0.5U_c(1-\cos 2\pi ft), \tag{3}$$

where U_c was the reference maximum center velocity derived from the prescribed Reynolds number, and $\mathrm{Re}=U_cH_T/\nu$, where H_T was the upstream height and ν was the kinematic viscosity.

C. Basic simulation procedure

For a particular Re value, U_c was determined and A_i, A_s, and ϕ were chosen. This permitted the calculation of the dynamic H_i and H_s values, as well as the flow U_T, and the calculation of the pressure and velocity fields for the first time step. At every time step of the simulation, the new glottal wall position was calculated and a new mesh was generated from the previous mesh. The pressure and velocity fields of the previous time step were used as an initial guess in the iteration method. Computations were performed on a Silicon Graphics workstation. A converged solution provided the velocity and pressure distribution in the computational domain for each time interval. The distributions were stored on a computer disk for later analysis. The wall shear stress for every time step was calculated from the velocity distributions. Using a polynomial interpolation technique within the MATLAB package, the location of the flow separation point was calculated for each time step. The separation point was that axial location where the velocity gradient was zero ($du/dy=0$) [MRS criterion discussed in Schlichting and Gersten (2000)], whereby for moving walls the additional criterion $u=0$ is imposed, yielded changes of the separation points within the distance between nodes, and therefore was deemed insignificant for this study.

III. RESULTS

A. Basic characteristics of the mean cross-sectional velocities and transglottal pressure

An example of the kinds of signals obtained from this glottal model is shown in Fig. 2(a), and a sequence of glottal shape change is shown in Fig. 2(b). This figure shows two cycles of the glottal inlet mean velocity, glottal outlet mean velocity, glottal outlet volume velocity, tracheal velocity, transglottal pressure, and the glottal angle (positive values for glottal convergence, negative for divergence). Two phase values, 60° and 90°, with the inferior glottis leading the superior, are shown, while other parameters were held constant (Reynolds number Re=1000, fundamental frequency Fo=100 Hz, and amplitudes of $A_i=0.75$ mm, $A_s=0.90$ mm).

The patterns seen in Fig. 2 resulted from the interaction of the prescribed flow and wall motion. The mean cross-sectional inlet and outlet velocities, as well as the outlet vol-

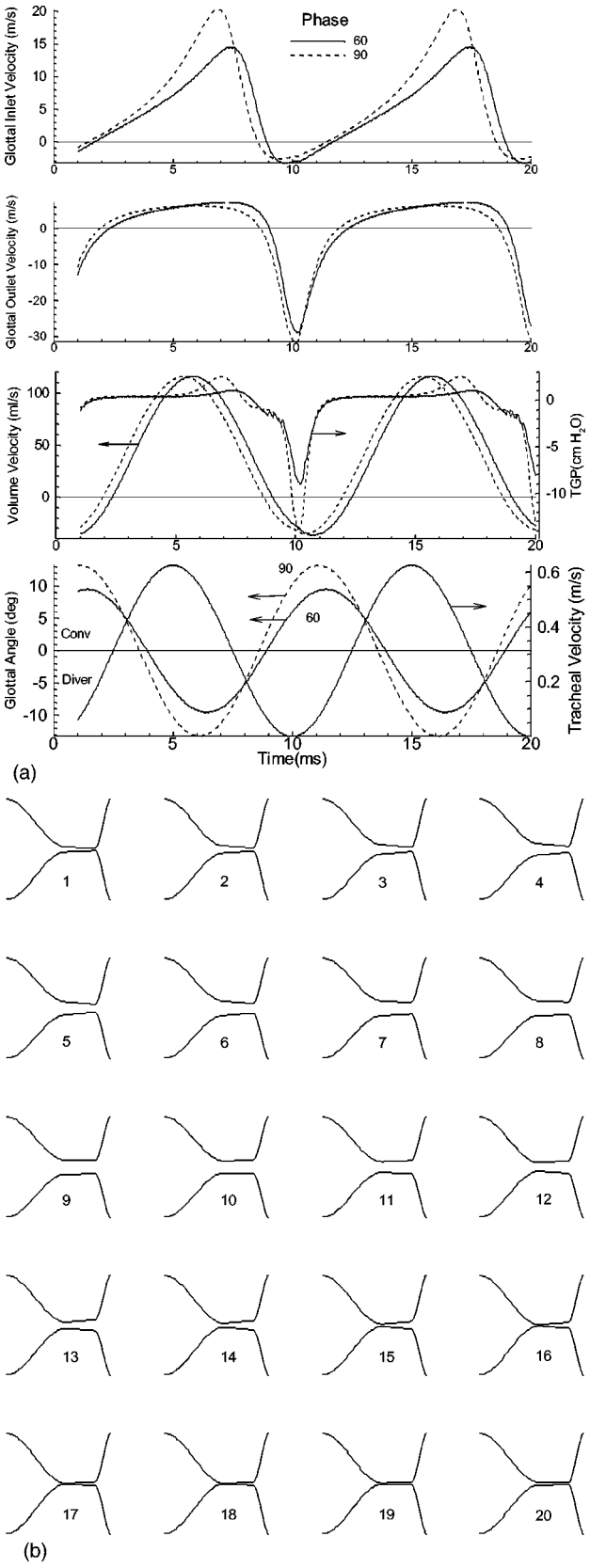


FIG. 2. (a) Glottal and tracheal velocities, angle, and transglottal pressure for the phase lag values of 60° and 90°, and Re=1000, Fo=100 Hz, $A_i=0.75$ mm, $A_s=0.90$ mm. (b) One cycle of motion of the vocal folds for the 60° phase case. The minimal diameter is 0.2 mm. The time between each frame number is 0.5 ms. Pressure units are in cm H_2O (1 cm $H_2O=98.1$ Pa).

ume velocities, are skewed to the right relative to the tracheal velocity. This skewing is consistent with human volume velocity signals. The skewing here appears to be due to the changing shape of the glottis and related to the resulting displacement flows. The maximum glottal velocity peaks occur while both inferior and superior glottal margins are moving medially, and the glottal angle is divergent, even though the tracheal velocity is decreasing. The inward wall movement creates a displacement flow in the downstream direction that increases the glottal velocities and delays the peak glottal velocities relative to the tracheal velocity.

We note that the volume velocity through the glottis and the glottal inlet velocity primarily have positive (downstream) going flow except for a portion of the cycle during which the glottis is opening. The reversal in direction of the flow (toward upstream) is due to the outward movement of the glottis (especially the inferior section) and the corresponding decreasing (and independent) tracheal flow. The outward wall motion, decreasing upstream flow, and incompressibility also create negative transglottal pressures; the absence of a dc flow prevents the maintenance of a positive subglottal pressure. The greatest effect of these conditions is on the glottal outlet velocity, where relatively large negative velocities are seen during glottal opening. It is noted that in the normal larynx, the flow decreases because of greater glottal flow resistance as the glottal diameter reduces during the glottal closing gesture, but here the flow is independently reduced (and has no dc component). This model does not mimic certain features of normal phonation, such as the usual lack of negative volume flows from the glottis and the modulated but positive transglottal pressure (e.g., see Cranen and Boves, 1985), but this model should tend to emphasize the effect of the wall movement on the flow behavior, and will therefore be highly sensitive to the creation of flow separation from the glottal walls. It is assumed that the range of flow separation locations may therefore be greater than expected in normal phonation, and thus this experiment should ''bracket'' that expected range.

Figure 2 also indicates the effect of changing the phase of motion on the inferior and superior portions of the glottis. This effect is to shift the peak of the mean and volume velocities more to the right (later in time) for the 60° phase condition compared to the 90° phase due to timing of the upper and lower diameters.

B. Velocity profiles

The velocity profile across the glottis at any cross section will change as the glottis moves and as the tracheal input velocity changes. Figure 3 shows a snapshot of velocity profiles within the glottis for the nominal conditions (Re $=1000$, Fo $=100$ Hz, $A_i=0.0$ mm, $A_s=0.90$ mm, and $\phi=60°$) and a divergent glottal shape. The velocity magnitudes increase at the minimum cross-sectional area. The velocity vectors are almost flat for most of the channel and vanish on the wall due to the no slip condition. However, downstream of approximately cross section 15, the velocity vectors do not appear for a certain distance away from the wall. The velocities that are not shown are negative and point upstream, and do not appear because of the inability of the plotting routine to show them. The line in the figure designates where the velocity was zero, and is the lateral ''wall'' or boundary of the ''flow tube.''

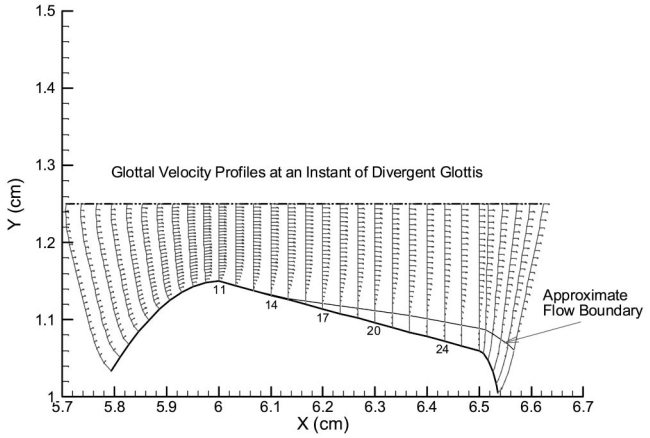


FIG. 3. Velocity vectors within a divergent glottis for Re$=1000$, Fo $=100$ Hz, $A_i=0.0$ mm, $A_s=0.90$ mm, and $\phi=60°$. The length of the small arrows indicates the magnitude of the velocities. Velocities that are not shown near the vocal fold surface are negative and point upstream. The dashed line represents an approximation to the effective ''flow tube.''

Figure 4 shows a few selected velocity profiles from Fig. 3 in an overlaid format with the grid numbers that refer to their locations in Fig. 3. At grid number 11, which is at the minimum glottal area, the velocity has the largest peak value and the largest velocity gradient on the wall ($\delta u/\delta y$) (and consequently the wall shear stress has a large positive value, see below). Further downstream, this velocity gradient on the wall decreases and eventually passes through zero and becomes negative. Once the shear stress becomes negative, the flow reverses. Profiles 17, 20, and 24 show reverse flow below a certain height (negative velocity vectors). Flow separation occurs at the location of the zero shear value.

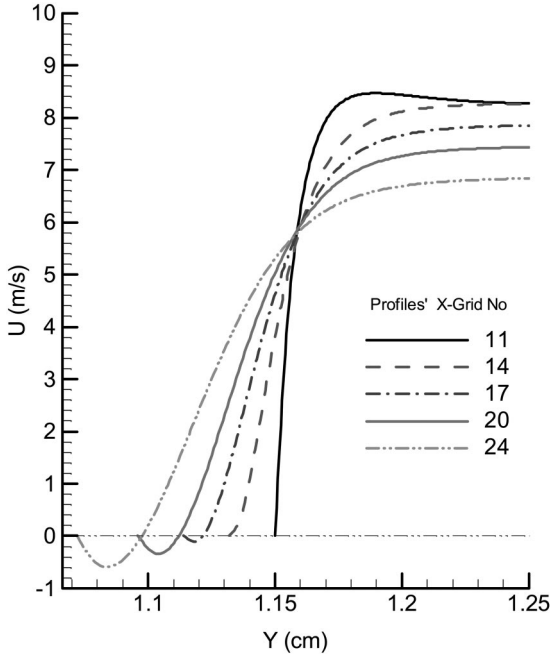


FIG. 4. Glottal velocity profiles at a few selected x-grid locations (from Fig. 3).

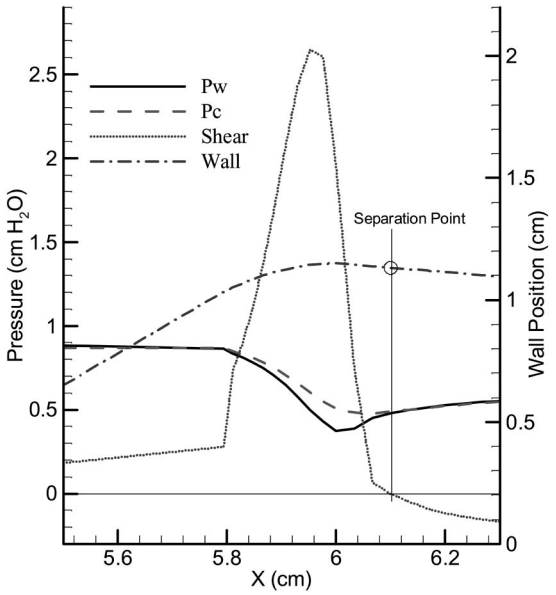


FIG. 5. Variations of wall pressures (Pw) and center pressures (Pc) along with the wall shear stress in a portion of the modeled glottis with flow separation. Also shown is an outline of the glottal surface and the separation point. The conditions are the same as in Figs. 3 and 4. Pressure units are in cm H_2O (1 cm H_2O=98.1 Pa).

Figure 5 demonstrates the variation of centerline pressure, wall pressure, and wall shear stress near a small portion of the glottal wall in the region of the minimum diameter for the same case as in Figs. 3 and 4. The shear stress has a large peak value at the location of the maximum curvature. The wall and centerline pressures are almost identical everywhere except in the region of high shear stress. The vertical line at zero shear stress intercepts the curves at the flow separation location (approximately $X=6.1$). After the separation point, the discrepancy between these two pressures disappears (and is shown to rise slightly, not stay constant). This discussion introduces the topic of flow separation. The following section will show how the separation point changes as phonatory parameters are altered in this model.

C. Flow separation locations

In this section, locations of the flow separation point for divergent glottal shapes are studied relative to motion of the inferior glottis, the superior glottis, and Reynolds number. For the following discussion, the nominal phonatory condition was Fo=100 Hz, $H_0=1.15$ cm, $A_i=0.75$ mm, $A_s=0.90$ mm, and phasing of 60° between the lower and upper glottal margin.

Figure 6 shows the *effects of glottal angle* on flow separation location. In the upper graph (a) the inferior margin is fixed and the superior margin oscillates with an amplitude of 0.9 mm. The lower graph (b) shows the inferior margins oscillating with an amplitude of 0.75 mm with the superior margin fixed. The separation point was tracked over the duration of a phonatory cycle as the inferior or superior amplitude was varied, other parameters having the default values. The symbols indicate the separation point locations for each instant of the oscillation. In both graphs, the glottal angle

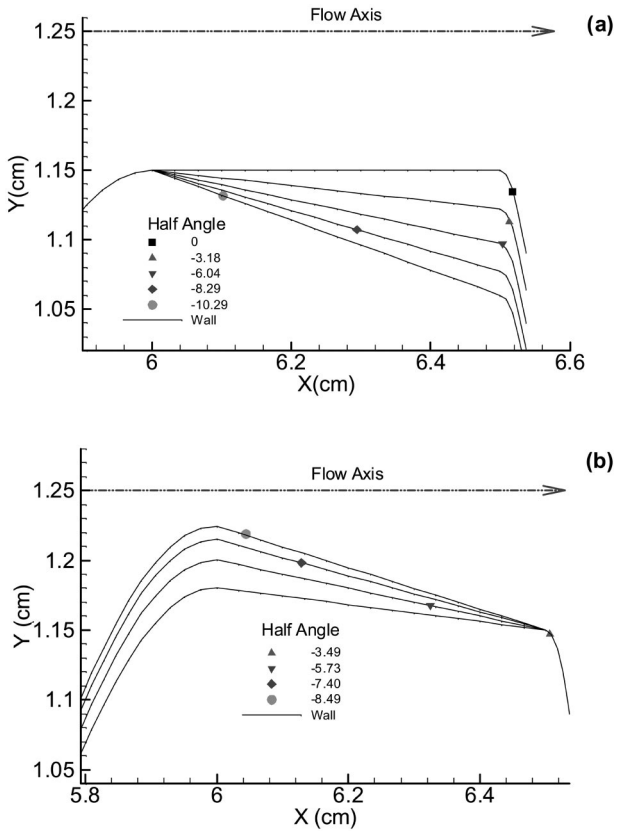


FIG. 6. Effect of the glottal angle on separation point locations: (a) inferior glottis is fixed and superior margin oscillate, and (b) superior glottis is fixed and inferior margin oscillates. Otherwise, the conditions were nominal (Fo =100 Hz, $H_o=1.15$ cm, $A_i=0.75$ mm, $A_s=0.90$ mm, and phasing of 60° between the lower and upper glottal margin).

was a major determinant of the separation point locations. As this angle increased, the separation point moved upstream after a relatively small angle was surpassed.

To show the *effect of glottal wall amplitude*, a case of a divergent glottis with a constant 5° half angle was modeled. The constant angle means that the inferior and superior margins were moved as a unit (zero phase difference between the upper and lower margins), preserving the angle, but allowing glottal motion. In Fig. 7 three panels are displayed. The top panel (a) shows glottal inlet mean velocities for two oscillation amplitudes. The solid curve shows the inlet mean velocity for the static glottis, the flow being symmetric in time. The increase of the amplitude to 0.25 mm (dashed curve) creates a skewing in the flow due to the wall motion (motion inward displaces flow in the downstream direction, skewing the flow).

The second panel (b) shows flow separation points for the static glottis at time instants of 15, 16, 17, and 18 ms. These separation points are on the decreasing portion of the flow, as can be seen from the first panel. The second panel indicates that for this static divergence, greater flow yielded separation points more *downstream* in the glottis, the opposite of the usual finding for the static divergent glottis (Shinwari *et al.*, 2003), thus suggesting that changing flow rates may yield different results than for constant flow rates for a static glottis.

The third panel of Fig. 7(c) includes four solid curves,

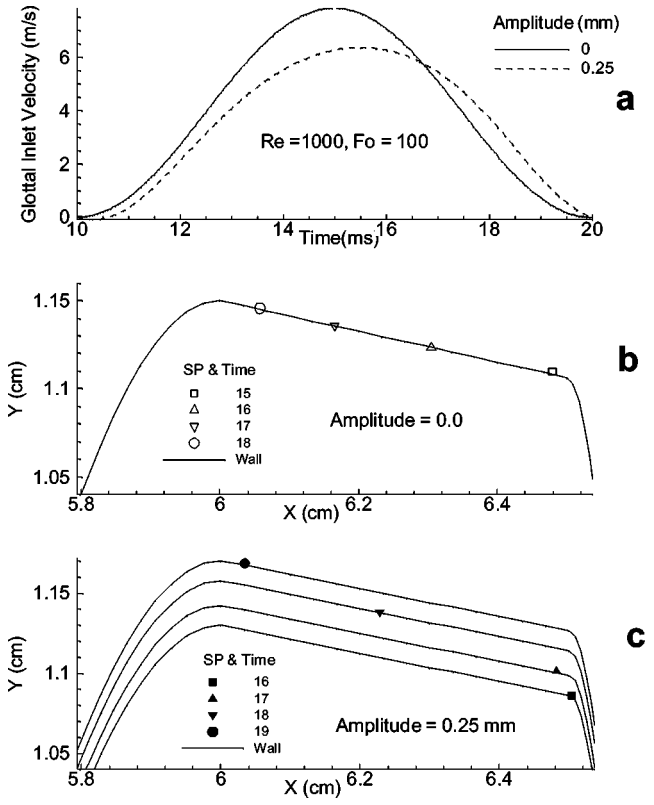


FIG. 7. Effects of the oscillation amplitude on the flow separation location in a 5° half-angle divergent glottis; (a) one cycle of glottal inlet velocities, (b) flow separation points for the static glottis, and (c) flow separation points for a 0.25 mm amplitude. ''SP'' stands for the separation point character, and ''time'' refers to the time of the first panel. Note that the times in panel (c) have been augmented by 1 ms to show the similar range of values as in (b).

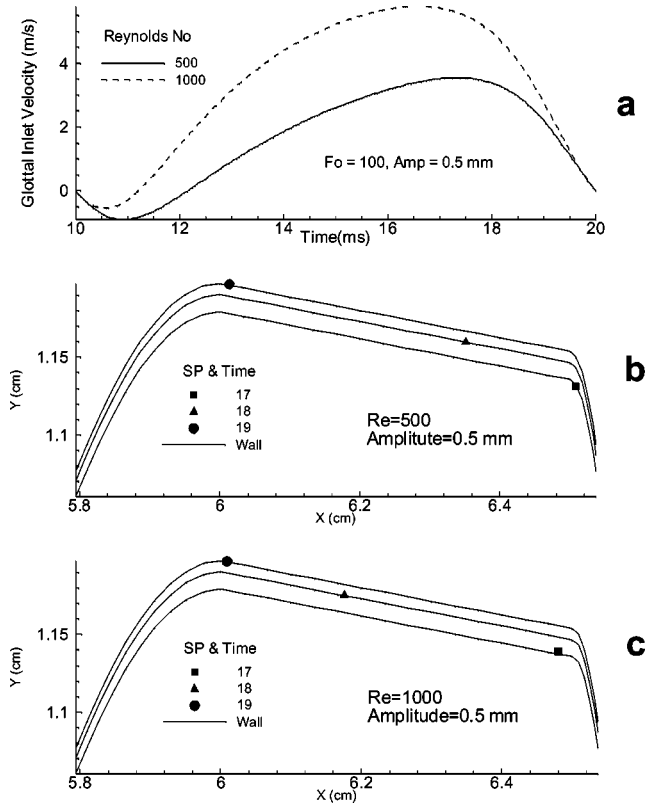


FIG. 8. Effects of the Reynolds number on the flow separation locations in a 5° half-angle divergent glottis: (a) one cycle of glottal inlet velocities, (b) flow separation points for Re=500, and (c) flow separation points for Re =1000.

which display wall locations and separation points at time instants of 16, 17, 18, and 19 ms. A comparison of the times and separation points for the second and third panels indicates that the inward motion of the vocal folds moves the separation point downstream (while still preserving the finding that lower flows yield more upstream separation locations).

The *effect of increasing flow* (Reynolds number) on flow separation is shown in Fig. 8. The top panel (a) shows one cycle of the glottal inlet mean velocities for Reynolds numbers of 500 and 1000. The amplitude of 0.5 mm (again for the zero phase glottis of 5° half angle) is associated with more skewing for the smaller Reynolds number (solid curve). The middle panel (b) shows three frames of wall motion at Reynolds number of 500 and their corresponding separation points for time instants of 17, 18, and 19 ms. Similarly, three wall positions and the separation points for the Reynolds number 1000 are shown in the third panel (c). It appears that as the Reynolds number increases, the flow separation locations move upstream, which is the typical expectation. The amount of the shift is not the same for each of the moments in time, however; the greatest movement was for the separation point at 18 ms, the separation point not near the glottal end points.

Once the flow separates, the ''flow tube,'' that is, the effective spatial region within which the air flows through the glottis, has a larger cross-sectional area than the minimum glottal area (the line cutting through the velocity profiles in Fig. 3 connects points where the flow velocity equals zero and gives the approximate effective boundary of the ''flow tube''). The ratio of the glottal cross-sectional area at the flow separation location and the minimal glottal cross-sectional area, called the *separation cross-sectional ratio*, is plotted in Fig. 9(a) for the 5° (zero phase) divergent glottal model for amplitudes of 0, 0.25, 0.50, and 0.75 mm. There is a trend for increasing ratio values from about 1–1.2 to approximately 1.5–2.0 in the first 25% of the cycle. This is the portion of the cycle in which the tracheal flow is accelerating (up to the inflection point of the velocity curve). From 25% to 50% of the cycle, the ratio value decreases to about 1.4–1.5 or so. This phase of the tracheal flow is when it is decreasing its rate of increase toward its maximum flow, suggesting that this slowing of the velocity increase creates a condition in which the static or outward wall movement pulls the separation point upstream. From 50% to 80%, the ratio value takes different paths depending upon the amplitude. For the zero amplitude (static) case, as shown in Fig. 7, the value of ratio continues to decrease as the flow decreases, meaning that the separation point continues to move upstream on the wall. When the amplitude is nonzero, the wall motion inward deflects (squeezes) the flow, pushing the separation point more downstream, the largest effect being shown by the largest amplitude (0.75 mm), moving the cross-sectional value to approximately 2. From about 80% to 100%, the tracheal flow goes to zero, and the amplitude cases show a fast drop of the ratio value toward 1.0. Figure

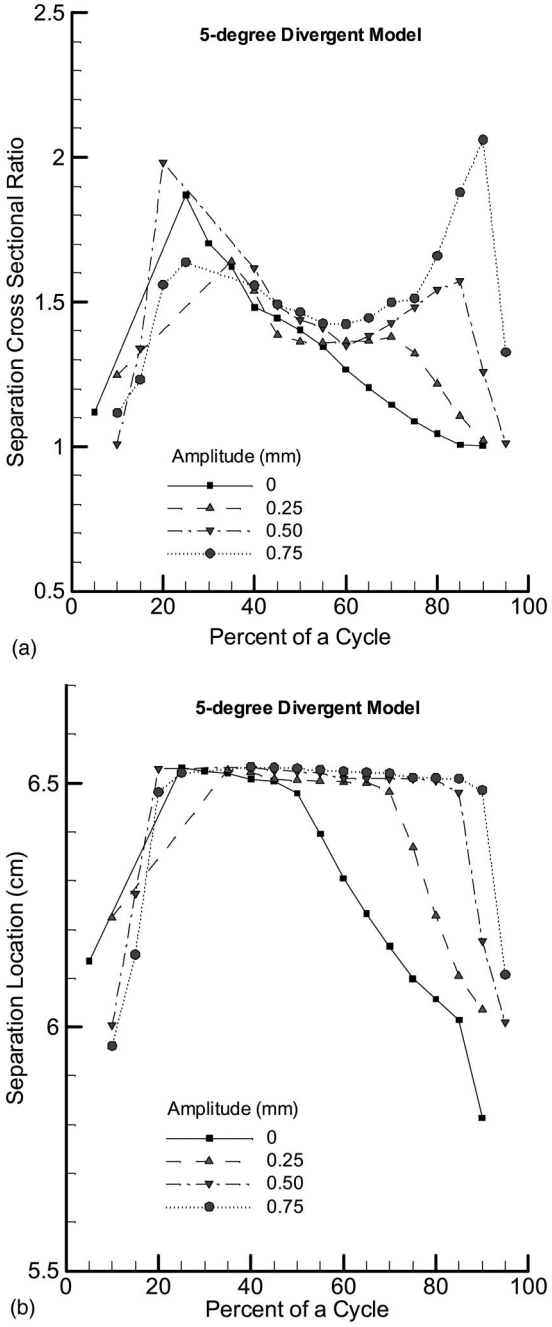


FIG. 9. Effects of amplitude on (a) the separation cross section ratio and (b) the flow separation point location in a 5° half-angle divergent model with zero phase lead between inferior and superior glottal margins. The tracheal flow and glottal excursion reach their maximum when the cycle is at 50%.

9(b) shows the separation locations of the same data shown in Fig. 9(a). For the static condition (zero amplitude), the flow separation location is governed by the flow rate as shown in Fig. 7(a). After a period of stationary condition, the deceleration portion of the cycle causes the separation point to move steadily upstream. Increase of the amplitude makes the stationary period longer.

Separation phenomena are shown for a variety of amplitudes in Fig. 10. Figure 10(a) is a series of four cases for which the Reynolds number was 1000, frequency was 100 Hz, and the amplitudes of the inferior and superior glottal margins were varied as shown on the graph. The values are for divergent glottal angles only. The range of values shown, approximately 1.1 to 1.9, is representative of the data from this study. The overall average was 1.47 (sd=0.23), and is slightly larger than adopted in the voice literature [for example, 1.3 by Lucero (1999)].

Figure 10(b) shows glottal half-angle versus the percent of the cycle for the four cases of oscillation of Fig. 10(a). Thus, the beginning of the cycle is at 0%, and ends at 100%. The figure indicates that the different cases have different beginning glottal angles, and different angles at each percentage of the cycle. Figure 10(c) gives the separation locations versus the glottal angle for the cases in Fig. 10(b). Figures 10(a)–(c) are to be considered in establishing further insight into the dynamic change of the separation point locations. These figures indicate the sensitivity of this model to both the flow and wall motion, and yield results that suggest *delays in the movement of the separation point relative to what might be expected from steady flow in diffusers*. For example, for the four cases in Fig. 10(c), three of them have their cycles start with a convergent shape, but with separation points upstream of the glottal exit (6.5 cm) location (note that the effective glottal entrance is at 6.0 cm). The glottal angle at which flow separation begins to move upstream, away from the exit of the glottis, varies, not happening until the angle reaches a divergent 6°, 3.5°, and 8° half-angle for cases 1, 3, and 4, respectively, and essentially never for case 2, where separation only occurs for *nearly uniform and convergent* shaping near the end of the cycle. Furthermore, cases 2 and 4 end their cycle with a convergent angle but with separation still upstream of the glottal exit. These findings that movement of the separation point upstream away from the exit begins to occur well into the divergent angles, and that upstream separation can persist into the convergent angle portion of the cycle, *strongly suggests inertial forces may act to retard the effects of the flow after the wall movement has occurred*. These findings need further, more complete investigation with more realistic models of phonation.

IV. DISCUSSION

Because this CFD forced dynamic model of phonation controls tracheal flow and wall motion, rather than lung pressure, it may be highly sensitive regarding the resulting locations for the flow separation points in the glottis. The results indicate that increasing the Reynolds number (or flow) tends to move the separation points upstream, as predicted by empirical studies of diffusers and steady flows. However, surprisingly, the direction of movement of flow separation locations appears to be opposite to expectations (or confounded) regarding the dynamic flow changes during a cycle of glottal oscillation. This is shown in a rather pure sense for the divergent glottis, static and moving, with zero phase between the inferior and superior margins, Fig. 9(a). For this condition, while the flow rate increases over the first 25% of the cycle, the separation point moves downstream, which is *opposite* to steady flow expectations (which would be movement of the separation point upstream). After the tracheal flow moves past its inflection point 25% into the cycle, the separation points move upstream during static and outward

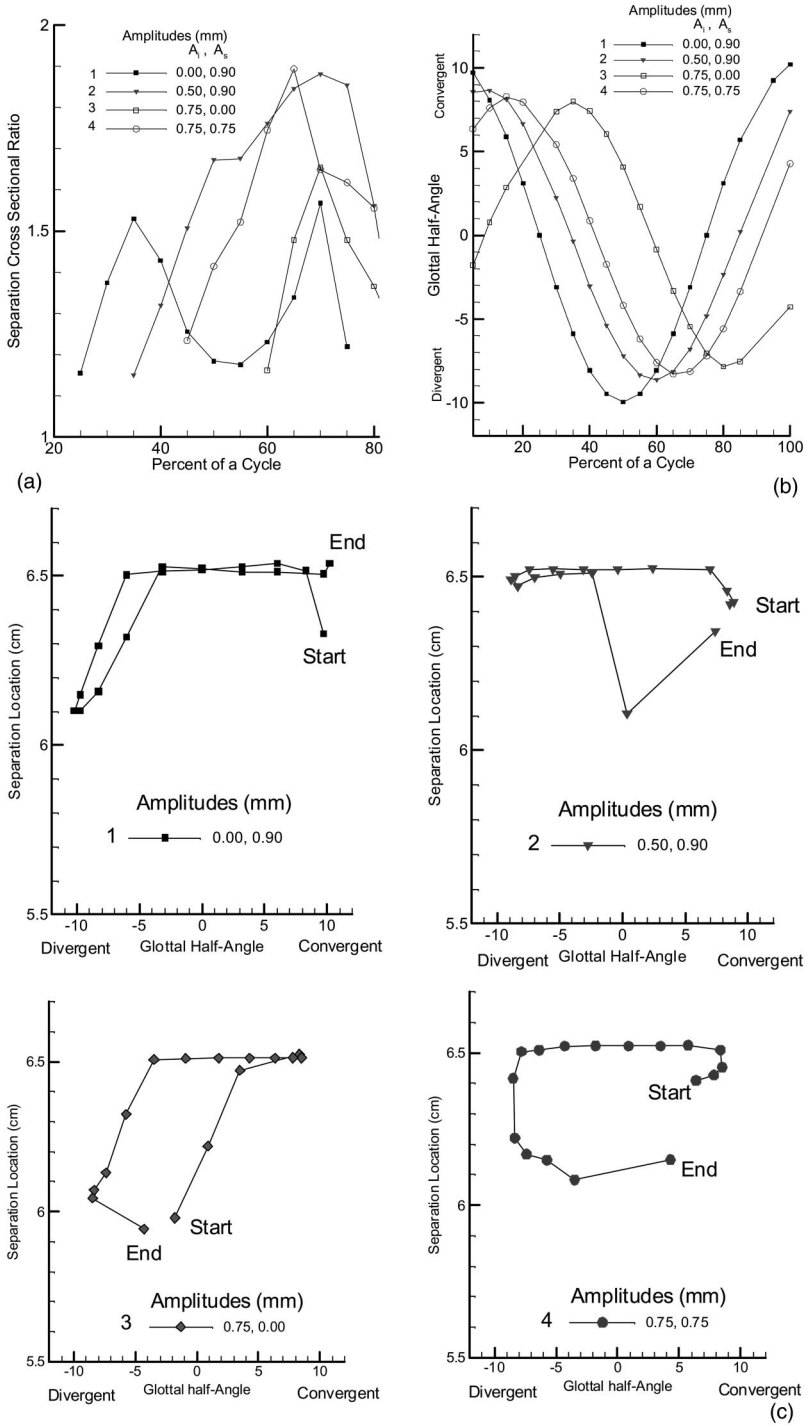


FIG. 10. (a) Separation cross-sectional ratio for the divergent angles for four cases of differing inferior, superior amplitudes. (b) Glottal half-angle for the same four cases of differing inferior, superior amplitudes. (c) Location of flow separation points (separation location relative to Fig. 1 *x* axis) relative to glottal half-angle for the four cases. The location 6.0 cm is at the glottal entrance (A in Fig. 1), and location 6.5 cm is the effective glottal exit (B in Fig. 1).

movement of the glottis (when the flow is still increasing), which is consistent with steady flow expectations, but perhaps not because the flow is increasing, but because the *rate of increase is decreasing*. Past the 50% cycle point, when the flow begins to decrease, for the static duct case, the separation point continues to move *upstream*, contrary to steady flow expectations. However, when there is wall movement, Figs. 9(a) and b suggest a relatively strong influence of the wall movement: the larger the amplitude (and therefore the greater the rate of movement of the glottal walls inward), the more the separation points are moved in the downstream direction because of the ''squeeze'' given to the airflow. These results suggest a strong interplay between the rate of change of the flow through the glottis (relative to the 100-Hz period) and the motion of the walls.

The phenomenon of having the separation point move further downstream in the glottal duct as the velocity increases, and then having a change toward upstream near the inflection point during the velocity increase, is consistent with the phenomenon of the delay in creating the Coanda or

skewing effect in the glottis for accelerated flows as shown by Hofmans *et al.* (2003). That is, their data show that cross-sectional pressure differences, which would reflect a change in status of the separation points in the glottis, occur approximately at or just after the velocity (or tracheal pressure) increase inflection point, similar to the computational findings here for the static divergent glottis with unsteady flow.

Because of the interest in the cross-sectional ratio between the glottal duct diameter (area) at the separation location and the minimum glottal diameter (area), the ratio was mapped for a wide range of conditions (Fig. 9). Those data indicate that the range of this ratio value was from approximately 1.1 to 1.9, with an average of 1.47 (sd=0.27). This number does, however, vary with condition. The findings for the cases with dynamic glottal angle and diameter suggest that the separation point location is a delayed phenomenon relative to the motion of the vocal folds; when divergent angles begin, the flow may stay attached at the glottal exit, to begin to move up the divergent glottis for larger divergent angles, and then retain separation upstream of the glottis exit even as the glottis takes on the convergent shape. Pelorson *et al.* (1994) also show changing separation point locations over a wide range of the glottal duct, but their two-mass model results can not be directly compared here because of the lack of glottal angle information. Even though the computational model presented here may exaggerate this phenomenon of separation point movement due to strong inertial influence, this needs to be hypothesized to exist in more realistic flow-induced oscillation biophysical model.

The shear stress along the glottal walls is strongest at the location of the greatest curvature of the glottis, which would be at the entrance of the glottis for the time when the glottis has a divergent shape. This stress would act tangentially on the vocal folds in addition to the normal air pressures, and thus would add to the forces tending to move the tissue. Although the shear stresses are relatively small, they need to be considered in a more complete modeling process.

V. CONCLUSIONS

This study used a computational fluid dynamics model of phonation with the special characteristics of forced dynamic motion of the vocal folds and a prescribed tracheal flow that varied sinusoidally. The focus on the location of flow separation points during divergent glottal shaping was justified relative to the common application of their location and influence on designating pressures in the glottis. Thus, a general study of the influences on the separation point movement might lend further insight into phonation modeling requirements.

Results suggest the following.

(1) The separation cross-sectional ratio varied from about 1.1 to 1.9 in this study, with an average of 1.47.
(2) Decreasing the inferior-superior phase from 90° to 60° tends to delay the flow and velocity peaks more (greater skewing to the right).
(3) Increasing the average flow rate (in this study, the peak Reynolds number of the upstream flow) tends to move flow separation points further upstream, consistent with steady flow diffusers.
(4) Larger divergent glottal angles tend to move the separation point upstream, consistent with steady flow diffusers.
(5) Inward motion of the divergent glottal walls tends to move the separation point downstream (via ''squeezing'' the flow) and skew the peak velocities.
(6) Mean velocity increase tends to move the separation point downstream in the diverging glottis, and mean velocity decrease can move the separation point upstream, contrary to steady flow diffuser behavior.
(7) The rapid change of airflow through the glottis appears to have a strong influence on the flow separation locations, most likely due to an inertial effect of the air. There appears to be a significant delay in establishing separation point locations during vocal fold motion with changing glottal angle, compared to the separation points expected during steady flow, static diffuser conditions.

Results here suggest that the rate of change of the flow may be important, as well as whether the flow is increasing or decreasing. These aspects need to be studied in more detail. Specifically, what conditions create (and control) reversals of separation location movement relative to expectations from steady flow diffuser results?

This model did not include the influence of the airflow (air pressure) on the tissue *per se*, and therefore a parametric study of flow separation within a fluid-surface interaction model is pertinent.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Yong G. Lai and Iowa Institute of Hydraulic Research for assistance in adapting their CFD code for this project. This work was supported by NIDCD Grant No. DC03566-05.

Alipour, F., and Scherer, R. C. (**1998**). ''Pulsatile flow within an oscillating glottal model,'' J. Irani. Mech. Eng. **3**, 73–81.
Alipour, F., and Scherer, R. C. (**2002**). ''Pressure and velocity profiles in a static mechanical hemilarynx model,'' J. Acoust. Soc. Am. **112**, 2996–3003.
Alipour, F., Fan, C., and Scherer, R. C. (**1996a**). ''A numerical simulation of laryngeal flow in a forced-oscillation glottal model,'' J. Comp. Speech Language **10**, 75–93.
Alipour, F., Scherer, R. C., and Knowles, J. (**1996b**). ''Velocity distributions in glottal models,'' J. Voice **10**(1), 50–58.
Alipour, F., Scherer, R. C., and Patel, V. C. (**1995**). ''An Experimental Study of Pulsatile Flow in Canine Larynges,'' J. Fluids Eng. **117**, 577–581.
Cranen, B., and Boves, L. (**1985**). ''Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production,'' J. Acoust. Soc. Am. **77**, 1543–1551.
Deplano, V., and Siouffi, M. (**1999**). ''Experimental and numerical study of pulsatile flows through stenosis: Wall shear stress analysis,'' J. Biomech. **32**, 1081–1090.
Guo, C. G., and Scherer, R. C. (**1993**). ''Finite element simulation of glottal flow and pressure,'' J. Acoust. Soc. Am. **94**, Pt. 1, 688–700.
Hofmans, G. C. J., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (**2003**). ''Unsteady flow through *in-vitro* models of the glottis,'' J. Acoust. Soc. Am. **113**, 1658–1675.
Kucinschi, B. R., Scherer, R. C., De Witt, K. J., and Ng, T. T. M., (**in review**). ''Flow visualization of the air moving through a model of the larynx,'' J. Acoust. Soc. Am.

Lai, Y., and Alipour, F. (**2002**). ''A computational study of oscillating flow in a model larynx'' in *Proc. 9th Asian Cong. Fluid Mech.*, 27–31 May 2002, Isfahan, Iran.

Lai, Y. G., and Przekwas, A. J. (**1994**). ''A finite-volume method for fluid flow simulation with moving boundaries,'' Comput. Fluid Dynamics **2**, 19–40.

Lai, Y. G., So, R. M. C., and Przekwas, A. J. (**1995**). ''Turbulent transonic flow simulation using a pressure-based method,'' Int. J. Eng. Sci. **33**(4), 469–483.

Long, Q., Xu, X. Y., Rammarine, K. V., and Hoskins, P. (**2001**). ''Numerical investigation of physiologically realistic pulsatile flow through arterial stenosis,'' J. Biomech. **34**, 1229–1242.

Lous, N. J. C., Hofmans, G. C. J., Veldhuis, R. N. J., and Hirschberg, A. (**1998**). ''A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design,'' Acta. Acust. **84**, 1135–1150.

Lucero, J. C. (**1998**). ''Optimal glottal configuration for ease of phonation,'' J. Voice **12**(2), 151–158.

Lucero, J. C. (**1999**). ''A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset,'' J. Acoust. Soc. Am. **105**, 423–431.

Ojha, M., Cobbold, R. S. C., Johnston, K. W., and Hummel, R. L. (**1989**). ''Pulsatile flow through constricted tubes: an experimental investigation using photochromic tracer methods,'' J. Fluid Mech. **203**, 173–197.

Pelorson, X., Hirschberg, A., Vanhassel, 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.

Perktold, K., Hofer, M., Rappitsch, G., Loew, M., Kuban, B. D., and Friedman, M. H. (**1998**). ''Validated computation of physiologic flow in a realistic coronary artery branch,'' J. Biomech. **31**, 217–228.

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.

Schlichting, H., and Gersten, K. (**2000**). *Boundary Layer Theory* (Springer-Verlag, Berlin).

Senoo, Y., and Nishi, M. (**1977**). ''Prediction of flow separation in a diffuser by a boundary layer calculation,'' J. Fluids Eng. Paper No 76-WA/FE-6, 379–389.

Shinwari, D., Scherer, R. C., De Witt, K. J., and Afjeh, A. A. (**2003**). ''Flow visualization and pressure distributions in a model of the glottis with a symmetric and oblique divergent angle of 10 degrees,'' J. Acoust. Soc. Am. **113**, 487–497.

Siouffi, M., Deplano, V., and Pelissier, R. (**1998**). ''Experimental analysis of unsteady flows through a stenosis,'' J. Biomech. **31**, 11–19.

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.

Tamura, T., and Fronek, A. (**1990**). ''Detection of moving flow separation in pulsatile flow and the degree of stenosis by power of Doppler shift signals,'' Circ. Res. **67**(1), 166–174.

Titze, I. R. (**1988**). ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**, 1536–1552.

Tutty, O. R. (**1992**). ''Pulsatile flow in a constricted channel,'' J. Biomech. Eng. **114**, 50–54.

White, F. M. (**1974**). *Viscous Fluid Flow* (McGraw–Hill, New York).

Zhao, S. Z., Xu, X. Y., Hughes, A. D., Thom, S. A., Stanton, A. V., Ariff, B., and Long, Q. (**2000**). ''Blood flow and vessel mechanics in a physiologically realistic model of human carotid arterial bifurcation,'' J. Biomech. **33**, 975–984.

Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (**2002**). ''Computational aeroacoustics of phonation, part I: Computational methods and sound generation mechanisms,'' J. Acoust. Soc. Am. **112**, 2134–2146.

Sound generation by steady flow through glottis-shaped orifices

Zhaoyan Zhang,[a] Luc Mongeau,[b] Steven H. Frankel, Scott Thomson, and Jong Beom Park

Ray W. Herrick Laboratories, Purdue University, 140 South Intramural Drive, West Lafayette, Indiana 47907-1077

(Received 21 January 2004; revised 4 June 2004; accepted 17 June 2004)

Although the signature of human voice is mostly tonal, it also includes a significant broadband component. Quadrupolelike sources due to turbulence in the region downstream of the glottis, and dipolelike sources due to the force applied by the vocal folds onto the surrounding fluid are the two primary broadband sound generating mechanisms. In this study, experiments were conducted to characterize the broadband sound emissions of confined stationary jets through rubber orifices formed to imitate the approximate shape of the human glottis at different stages during one cycle of vocal fold vibrations. The radiated sound pressure spectra downstream of the orifices were measured for varying flow rates, orifice shapes, and gas mixtures. The nondimensional sound pressure spectra were decomposed into the product of three functions: a source function F, a radiation efficiency function M, and an acoustic response function G. The results show that, as for circular jets, the quadrupole source contributions dominated for straight and convergent orifices. For divergent jets, whistling tonal sounds were emitted at low flow rates. At high flow rates for the same geometry, dipole contributions dominated the sound radiated by free jets. However, possible source-load acoustic feedback may have hampered accurate source identification in confined flows. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779331]

PACS numbers: 43.70.Aj, 43.28.Ra [MSH] Pages: 1720–1728

I. INTRODUCTION

Voice production involves sound generation by turbulent confined jets through an orifice (the glottis) with a time-varying area. The glottis is a slitlike orifice between the vocal folds, located in the airway and connected to the vocal tract. In voice production, the glottis is excited into self-oscillations by the airflow through it, causing a confined pulsating air jet downstream. The motion of the vocal folds is complex, and involves a continuous evolution of the orifice profile. One cycle of the orifice time history may be idealized using a succession of three representative glottal wall profiles: convergent, straight, and divergent. The acoustic waves generated by the pulsating jet flow, which represent the main voice source, are effectively filtered by the articulators forming the vocal tract downstream, such as the mouth and nasal cavity, before being radiated from the lips.

The radiated sound signature in voice production consists of the sum of a periodic component and a random broadband component. The periodic component due to the pulsating flow forced by glottis oscillations dominates at low frequencies in sonorant sounds. This basic tonal voice production mechanism is generally well understood, and can be modeled using a quasi-steady approximation (Zhang *et al.*, 2002b). The (modulated) broadband component, which is presumably related to the turbulence in the air stream, is much smaller in amplitude for phonation at low frequencies. However, broadband sound is an important component of human voice at high frequencies (above approximate 7000 Hz). Similar jet flows are also involved in the generation of fricative consonants, another important component of speech. A better understanding of the broadband sound generation mechanisms is useful in speech applications such as voice diagnostics and high-quality speech synthesis.

There have been relatively few studies on the production of broadband sound from air flow through the larynx and the vocal tract. Early measurements suggested that the broadband noise spectrum is relatively constant in the mid-audio frequency range (Flanagan, 1965). Experimental work by Meyer–Eppler (1953) has suggested that, for fricative consonants, a critical Reynolds number exists below which there is negligible turbulent sound generation. Meyer–Eppler has reported that the sound pressure related to turbulence in the airflow increases with the square of the Reynolds number. Such a scaling law, however, was postulated without rigorous verification through systematic changes in physical model size or working fluid.

Stevens (1998) has reported that the radiated sound pressure from turbulent flow through constrictions is proportional to the third power of jet velocity and the square root of the constriction area. The recent work of Harper (2000) on breathing sounds was again based on a Reynolds number power law. Such models neglect the possible influence of acoustic loading on sound production. The spectrum of the sound pressure for a source such as turbulent flow past an obstacle located in an open space tends to have a broad peak at a constant Strouhal number (Goldstein, 1976). To account for the influence of the acoustic response of cavities and ducts for sources within confined spaces on power law exponents, inverse filtering methods are needed. Many of these

[a]Also at the Institute for Systems Research, University of Maryland, College Park, MD 20742; Electronic mail: zhaoyan@glue.umd.edu

[b]Currently on leave at Ecole Centrale de Lyon, 36 Avenue Guy de Collongue, 69134 Ecully, France.

 © 2004 Acoustical Society of America

methods may not yield accurate results when the sound source region is spatially distributed, random, and spatially partially coherent.

Broadband sound generation by confined stationary jets through circular orifices was recently studied by Zhang *et al.* (2002a). An acoustic similarity law was derived, in which the nondimensional sound pressure spectrum was formulated as a product of a source spectral distribution function, a radiation efficiency function, and an acoustic frequency response function. The three functions were separated using a spectral decomposition method, thus allowing the identification of source type and source spectral characteristics. Sound radiation was found to be similar to that of a quadrupole source in tubes, with negligible dipole contribution.

For sound generation by stationary jet flows through glottis-shaped orifices, dipole sources due to the fluctuating pressure on the orifice walls may play a role in addition to the quadrupole sources due to flow turbulence in the region downstream of the orifice (Zhang *et al.*, 2002b). It is believed that the shape of the glottal orifices has an important impact on the behavior of these two sound production mechanisms. In particular, there may exist interactions between vortical flow structures and the sound field in cases where the orifice walls form a divergent diffuser.

Vortex structures in jet flows through orifices have been well described. Johansen (1929) studied water flow through a sharp-edged circular orifice mounted concentrically in a glass pipe. Flow visualization showed an axisymmetric jet with vortex rings. At low Reynolds number, as the flow began to transition from laminar to turbulence, these vortices were shed at a constant Strouhal number. Vortices were also observed to emerge from a sharp-edged two dimensional slit as well as a sharp-edged circular orifice in a rectangular channel by Beavers *et al.* (1970). Vortex sheets were formed with a constant Strouhal number (0.43 for two-dimensional jets and 0.63 for circular jets) over Reynolds numbers ranging from 500 to 3000. Flow visualization of the jet formation process has been performed by Shadle *et al.* (1987) using a dynamic model of the vocal folds consisting of two rectangular shutters. It was found that when the orifice was slightly off-center, the jet attached to the wall closest to the orifice. When obstructions representing false folds were added downstream, however, the jet plume ''straightened out.'' Recently, flow visualization observations have been made for flow through a rigid static divergent glottis model (Dabiri *et al.*, 2001). The jets were found to be skewed, and oscillated randomly. Vortical structures were also observed. The flow separation points moved along the glottis walls as the transglottal pressure was varied, and the flow was very unstable.

Instability waves and related vortex structures are possible sound sources. Relationships between the existence of forced shear layer instability and radiated sound were identified in the experiments of Moore (1977). McGowan (1988) used an aeroacoustic approach to study the source of sound during phonation. He postulated that, in addition to the unsteady volume velocity monopole source, an oscillating force (i.e., a dipole source) arises from the interaction between vortical structures shed from the glottis and the velocity field itself. In measurement using a dynamic mechanical model of the larynx and the vocal tract, Barney and co-workers (1999) found evidence of a vortex train, which contributed to sound radiation due to the interaction with the vocal tract downstream. Hofmans (1998) experimentally and computationally investigated the starting flow through various mechanical models comprising nozzles and orifices. A very complex behavior of interacting vortices due to the jet instability was observed. Frequencies of pressure perturbations associated with the vortex pairing and merging process were in the audible range. Contributions to speech production were postulated.

The present study is an extension of the work of Zhang *et al.* (2002a). The objective was to investigate the possible effects of the orifice geometry on broadband sound generation. The investigation of the source spectral characteristics for different orifice geometries could also allow the development of advanced models for broadband voiced sound in speech synthesis applications. Whistlelike tone generation for geometries featuring a divergent orifice cross-section was encountered, and is also discussed.

II. THEORETICAL BACKGROUND

A brief description is given in this section of the acoustic similarity law and spectral decomposition method derived by Zhang *et al.* (2002a). The readers are referred to the original paper for a detailed analysis and validation.

Consider sound generation by a finite region of turbulence in a long rigid rectangular tube. Assuming a compact source region, the radiated sound pressure spectral density can be expressed as the product of a source spectral distribution function F, a radiation efficiency function M, and a far-field acoustic response term G. The latter describes the effects of sound propagation phenomena from the source region to the observer. The nondimensional sound pressure spectrum, $E=S_{pp}(x,f)/(\frac{1}{2}\rho U^2)^2 d/U$, is presumed to obey the product law (Zhang *et al.*, 2002a):

$$E=F\left(\mathrm{St},\mathrm{Re},\frac{d}{D}\right)M(\mathrm{He})G\left(\mathrm{He},\frac{x_R}{D},R\right), \tag{1}$$

where S_{pp} is the sound pressure spectral density, D is the tube cross-sectional dimension, d is the orifice hydraulic diameter, x_R is the observer location, ρ is the fluid density, and R is the acoustic boundary condition. The hydraulic diameter was defined for the slit-shaped orifice cross sections as 2 times the orifice width. The dimensionless groupings are defined as $\mathrm{Re}=Ud/\nu$, $\mathrm{St}=fD/U$, $\mathrm{He}=fD/c$, where U is the centerline velocity of the jet, ν is the fluid viscosity, f is frequency, and c is the speed of sound.

The radiation efficiency is defined as the ratio of the radiated sound power and the mechanical power driving the sound generation phenomena. For a multipole source of order n in free space, the radiation efficiency function is of the order of $(kD)^{2n}$ (Howe, 1998), where k is the wave number. The radiation efficiency decreases with ascending multipole order. The presence of the confinement changes both the source characteristics and the sound field. At frequencies below the duct cut-on frequency, in which only plane wave

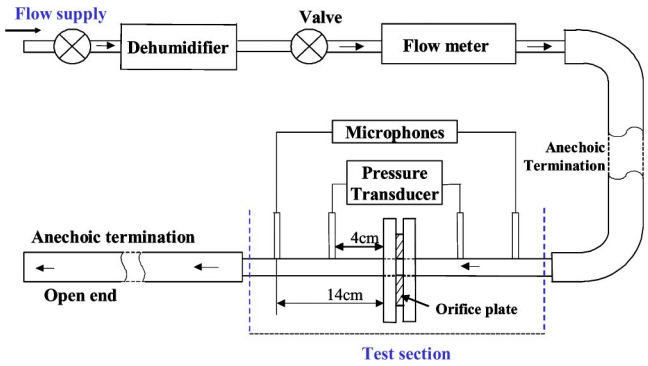


FIG. 1. Schematic of the experimental apparatus.

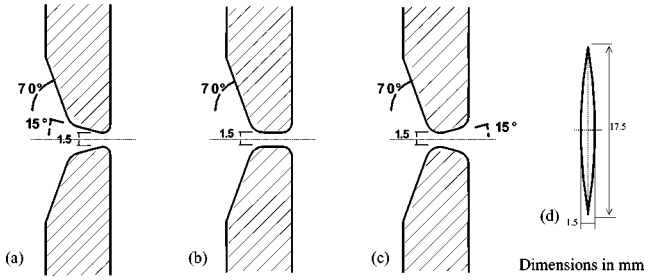


FIG. 2. Cross section of the three orifice passages (a) convergent, (b) straight, and (c) divergent. The flow direction is from left to right. (d) Elevation of the orifice showing the frontal opening area.

propagation is possible and the sound field is essentially one-dimensional, sound radiation is more efficient in tubes than in free space (Howe, 1998). For example, the radiation efficiency of a quadrupole source in tubes is in the form of $(kD)^2$, in contrast with $(kD)^4$ for a quadrupole source in free space $(kD<1)$. As frequency increases above the cut-on frequency, the sound field is three-dimensional and the sound radiation behavior approaches that for free space (Davies *et al.*, 1968). The radiation efficiency accordingly transitions from $(kD)^2$ to $(kD)^4$.

The M and G functions are functions of only the Helmholtz number. Their product forms a new function G_1. For a fixed microphone position, constant acoustic boundary conditions and orifice geometry, Eq. (1) becomes

$$E = F(\mathrm{St},\mathrm{Re})G_1(\mathrm{He}). \tag{2}$$

The three functions can be separated from measured sound pressure spectral data using a spectral decomposition method (Zhang *et al.*, 2002a), assuming a linear source-filter model for the sound generation system. Expressing the nondimensional spectrum using a logarithmic scale, Eq. (2) is rewritten as

$$10\log F(\mathrm{St},\mathrm{Re}) = 10\log E - 10\log G_1(\mathrm{He}). \tag{3}$$

Nondimensional sound spectra may be plotted against the St number and the Re number for values of f and c forming a constant He number. Since the F and G_1 functions are assumed to be independent, surfaces for consecutive He number values should be parallel to each other. They should have the same functional shape and only differ by a constant level difference over the St–Re plane. This level difference is equal to the difference in the value of $10\log G_1$ corresponding to the two He number values considered. By arbitrarily imposing a zero value to the function G_1 at the minimum value of the He number in the experimental data, the G_1 function can then be uniquely constructed over the entire range of He number. This allows $10\log F^k$ functions to be obtained for each combination of the Strouhal number and the Reynolds number using Eq. (3). The final value of the function $10\log F$ is obtained by ensemble-averaging all $10\log F^k$ functions.

It has been shown that the acoustic frequency response function, G, is in good agreement with the transfer function between the radiated acoustic pressure at one fixed location and the strength of an equivalent velocity source located at the orifice, and has a comparatively uniform spectral shape (Zhang *et al.*, 2002a). For a multipole sound source of order n, the G_1 function then follows closely the trend of the M function. Therefore, it is possible to identify the dominant order of the source multipole from the trends in the measured G_1 functions.

III. EXPERIMENTAL APPARATUS

Hardware from a previous study was used (Zhang *et al.*, 2002a). The test section consisted of two rectangular tubes, as illustrated in Fig. 1. Each tube had an inner cross-section 2.25×2.25 cm and was 22.5 cm long. A 6 mm thick rubber glottis-shaped orifice plate with two cover aluminum plates was inserted between the two rectangular tubes. The rubber glottis-shaped orifice plates were built to simulate the human vocal folds geometry, and acted as a constriction to the airflow. Three different orifice geometries (straight, convergent, and divergent) (Fig. 2) were used in the experiments, each representing a typical stage of the phonation cycle. The two tubes were connected to anechoic terminations on both the flow supply and the discharge ends, respectively. The same anechoic terminations were used as in previous studies (Zhang *et al.*, 2002a; Zhang *et al.*, 2002b). The reflection factor of the downstream system was measured using the two-microphone method. Its magnitude is shown in Fig. 3 over the plane wave frequency range. The reflection factor

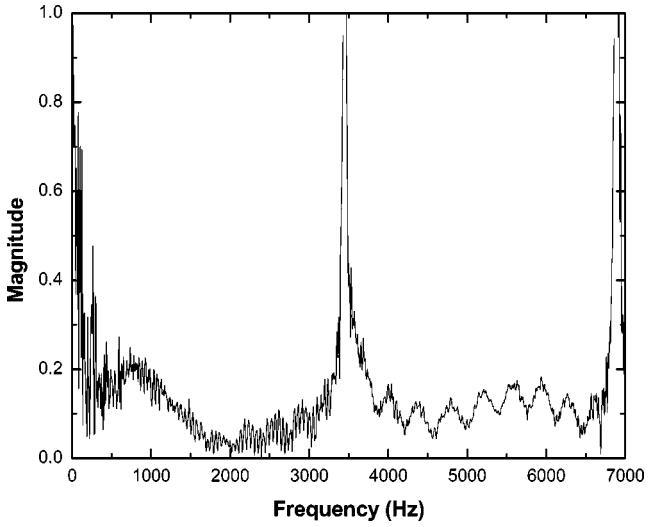


FIG. 3. Magnitude of the reflection factor of the downstream system.

magnitude lies below 0.3 at most frequencies. The peaks around 3500 Hz and 6900 Hz are believed to be due to the fact that the microphone spacing is a multiple of half-wavelengths and the calculation of reflection factor is indeterminate. More resonances may exist at frequencies beyond the cut-on frequency of the acoustic high modes (approximate 7000 Hz). The flow supply was designed for the use of different gas mixtures in the experiments. Use of different working fluids allows the variation of fluid properties, such as the viscosity and the sound speed. Three different helium–CO_2 mixtures, with helium mole ratios of 0, 0.33, and 0.57, were used in addition to air, allowing the speed of sound to be varied between 268 m/s and 409 m/s, and the viscosity to be varied between 8.4×10^{-6} m/s and 72.69×10^{-6} m/s^2.

Two flush-mounted microphones (6.35 mm diameter, B&K 4938) were installed along the tubes, each 14 cm upstream and downstream from the orifice plate, for the measurements of the radiated sound pressure. The background noise of the system, measured with the flow supply turned off was found to be negligible. Preliminary measurements were made using two microphones, with the second one installed 9 cm from the orifice plate. Measurement of the pressure frequency transfer function between the two microphones showed that the pressure disturbances propagated at the speed of sound. At very low frequencies, the microphone may have been in the acoustic near field and recorded pseudosound due to the turbulent structures which may not have been dissipated completely. This is believed to have induced some errors at low frequency in the identified M function and G_1 functions.

The output signals from the microphones were acquired at a sampling rate of 32 768 Hz using an HP356XA data acquisition system. The volumetric flow rate through the orifice was measured using a precision mass-flow meter (Baratron type 558A) at the inlet of the test section. The time-averaged pressure gradient across the orifice was measured using a pressure transducer (Baratron type 220C), with pressure taps installed along the tube as close as possible to the orifice plate.

IV. BROADBAND SOUND GENERATION

A. Straight orifice

Figure 4 shows radiated sound pressure spectra in air for different values of orifice pressure drop, Δp. The spectral level increases with the pressure drop, or the centerline velocity calculated using Bernoulli's equation

$$U=\sqrt{2\Delta p/\rho_0}, \tag{4}$$

where $\rho_0=1.2$ m^3/s is the medium density. Note that pressure heads in voice production studies are customarily reported in units of cm water.

The spectral decomposition method discussed in Sec. II was used to separate the acoustic frequency response function, the radiation efficiency function, and the source spectral distribution function. Figure 5 shows the resulting G_1 function. Additional curves showing He^2 and He^4 power laws are also shown in the figure for comparison. Following the discussion in Sec. II, the M function approximately obeys He^2 and He^4 trends at low and high He numbers, respectively, indicating a dominant quadrupole-type sound source. The corresponding source function, F, is shown versus the Reynolds number for constant Strouhal number, and versus the Strouhal number for constant Reynolds number in Figs. 6 and 7, respectively. Recall the hydraulic diameter is used as the characteristic dimension in the Reynolds number. The effects of Reynolds number were nearly negligible over the range from 200 to 20 000. This was expected since large scale turbulence contains most of the flow energy, and is sound radiation efficient. Increasing the Reynolds number implies finer grain turbulence in the flow, produced through the well-known cascading process. Small grain turbulence is spatially incoherent and thus contributes little to sound generation. Regressions show that the source function decreased approximately with the third power of the Strouhal number

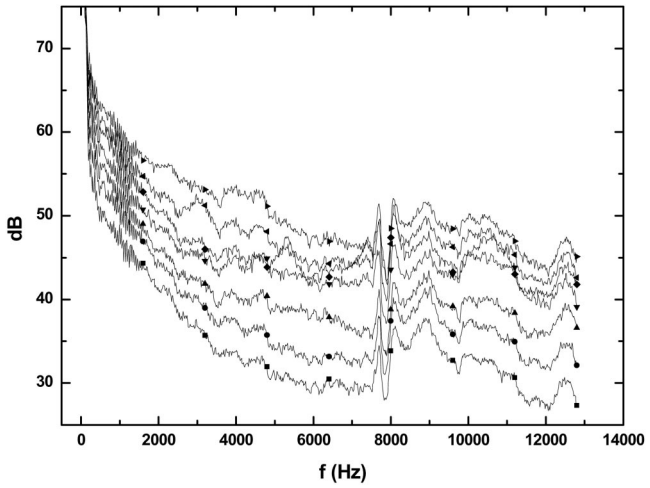


FIG. 4. Sound spectra measured for different transglottal pressure using air. Straight glottis-shaped orifice. ■, $\Delta p=8$ cm H_2O; ●, $\Delta p=10$ cm H_2O; ▲, $\Delta p=12$ cm H_2O; ▼, $\Delta p=14$ cm H_2O; ◆, $\Delta p=16$ cm H_2O; ◀, $\Delta p=18$ cm H_2O; ▶, $\Delta p=20$ cm H_2O.

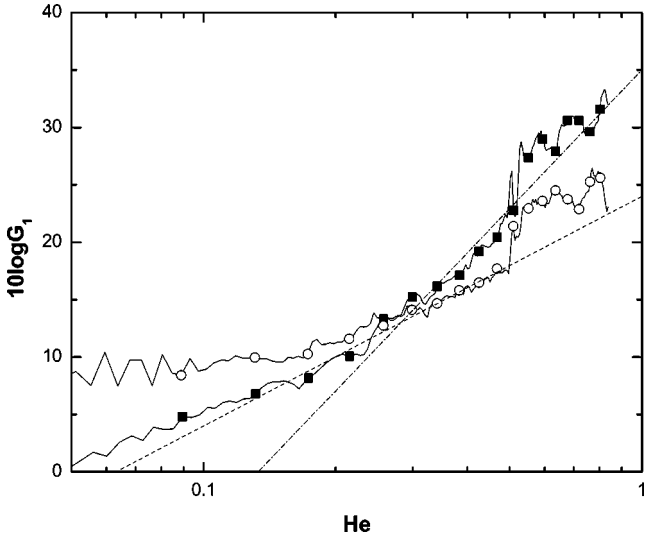


FIG. 5. Acoustic frequency response versus Helmholtz number. ■, straight glottis-shaped orifice; ○, convergent glottis-shaped orifice; - - -, 20 log He; -·-·-, 40 log He.

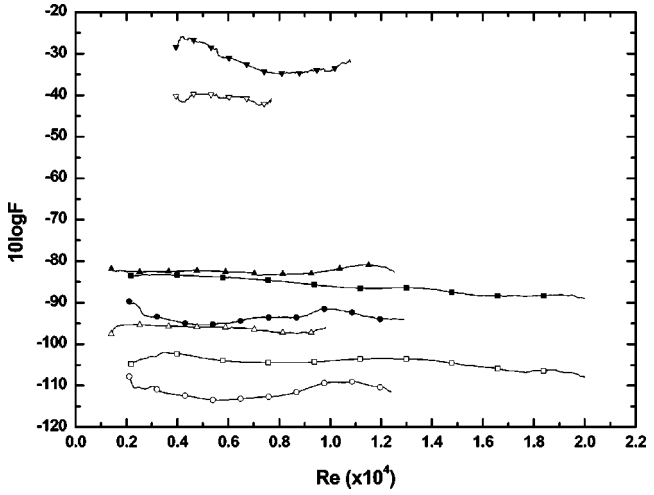


FIG. 6. Source spectral distribution function vs the Reynolds number at constant Strouhal numbers for different orifices. ■, circular orifice; ●, straight glottis-shaped orifice; ▲, convergent glottis-shaped orifice; ▼, free jets from divergent glottis-shaped orifice. Solid symbols, St=2; open symbols, St=5.

for Strouhal numbers smaller than one, and with the fifth power for Strouhal numbers greater than two, which is similar to the circular orifice case (Zhang *et al.*, 2002a).

B. Convergent orifice

Similar experiments were performed using a convergent glottis-shaped orifice. Figure 8 shows the sound spectra measured for air and different pressures. As in the case of straight orifices, the spectral amplitudes increase with jet velocity. The spectral decomposition method was applied and the resulting G_1 function is shown in Fig. 5. The G_1 function follows a He^2 trend at low Helmholtz numbers (up to around 0.45, the cut-on Helmholtz number for higher order acoustic modes), indicating a quadrupole-like sound source dominant at low frequencies. This is not surprising, as the flow within

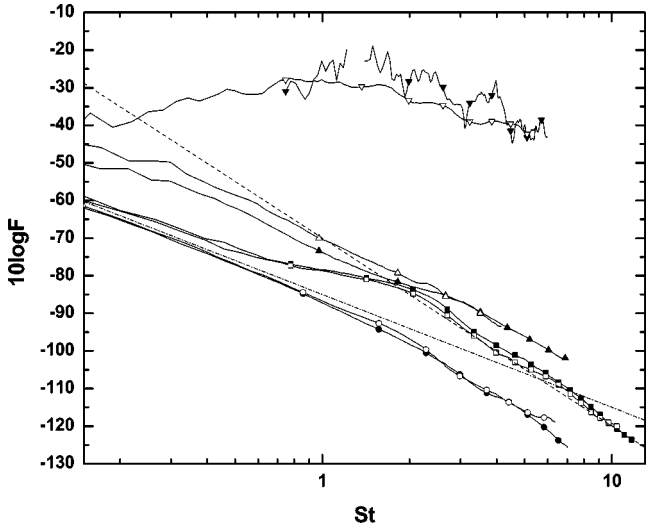


FIG. 7. Source spectral distribution function vs the Strouhal number at constant Reynolds numbers for different orifices. ■, circular orifice; ●, straight glottis-shaped orifice; ▲, convergent glottis-shaped orifice; ▼, free jets from divergent glottis-shaped orifice. Solid symbols, Re=4000; open symbols, Re=7000; - - -, −50 log St; -·-·-, −30 log St.

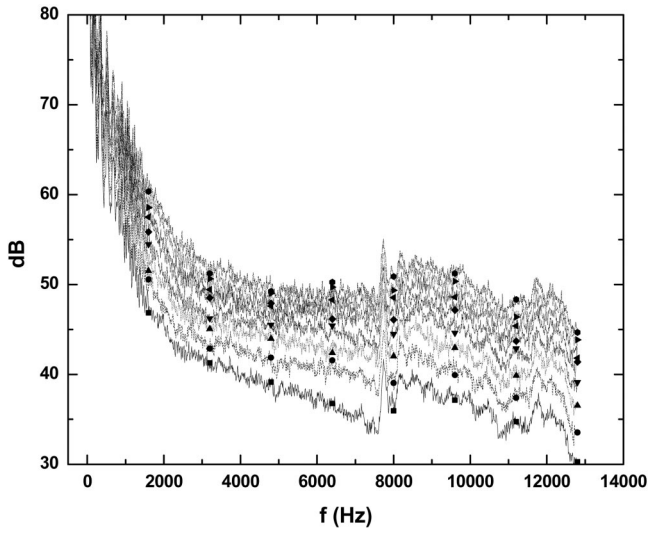


FIG. 8. Sound spectra measured for different transglottal pressure using air. Convergent glottis-shaped orifice. ■, $\Delta p = 10$ cm H_2O; ●, $\Delta p = 12$ cm H_2O; ▲, $\Delta p = 14$ cm H_2O; ▼, $\Delta p = 16$ cm H_2O; ♦, $\Delta p = 18$ cm H_2O; ◀, $\Delta p = 20$ cm H_2O; ▶, $\Delta p = 22$ cm H_2O; ⬢, $\Delta p = 24$ cm H_2O.

and immediately downstream the orifice is very stable (or ''laminar''), and turbulence appears further downstream into the tube (see discussion of Fig. 17 below). For greater Helmholtz numbers, the G_1 function deviates from a He^2 trend. The radiation function approaches at high frequency a He^4 law characteristic of a free field. Accurate regression at high Helmholtz numbers is difficult due to the presence of many resonances and sparse data points. At very low Helmholtz numbers the G_1 function is independent of Helmholtz number, which is partially due to flow supply noise. Another source of errors is the possible pseudosound recorded by the microphone, as discussed earlier in Sec. III.

The F function is plotted versus the Reynolds number and the Strouhal number in Figs. 6 and 7, respectively. Again, the effects of the Reynolds number were negligible. The spectral shape of the F function was very similar to that for the case of a straight orifice, following a St^{-3} trend at low Strouhal numbers, and St^{-5} at high Strouhal numbers.

C. Divergent orifice

The sound radiation characteristics of divergent orifices were different from those of the straight and convergent orifices. Figure 9 shows the sound spectra measured with air for mean pressure drops from 9 cm H_2O to 30 cm H_2O. Broadband levels seem to jump up for pressures above 10 cm H_2O. As the mean pressure drop further increases, level changes are frequency dependent. The sound levels are nearly constant within specific frequency bands, for example around 4 kHz. Near the Nyquist frequency of 12.8 kHz, the spectral level culminated for intermediate pressures around 18 cm H_2O.

Figure 10 shows similar sound spectra measured with CO_2 as the working fluid for norminal pressure drops from 10 cm H_2O to 26 cm H_2O. The nominal pressure drop is defined based on the density of air instead of the density of

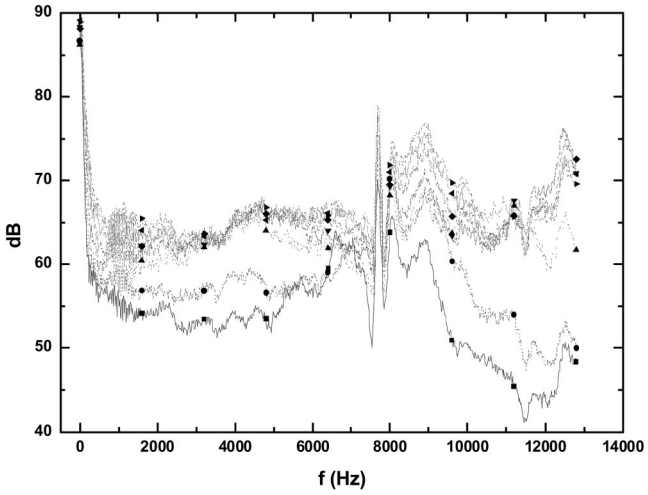


FIG. 9. Sound spectra measured at high flow rate in air. Divergent glottis-shaped orifice. ■, $\Delta p = 9$ cm H_2O; ●, $\Delta p = 10$ cm H_2O; ▲, $\Delta p = 14$ cm H_2O; ▼, $\Delta p = 18$ cm H_2O; ♦, $\Delta p = 22$ cm H_2O; ◀, $\Delta p = 26$ cm H_2O; ▶, $\Delta p = 30$ cm H_2O.

the specific gas mixture. Thus values of nominal transglottal pressure correspond to a given jet velocity, regardless of gas mixture. Since the sound speed is smaller in CO_2, the spectra extend to a larger He number. Two resonance frequencies (around 6000 Hz and 12 000 Hz) may be detected from the spectra. Again, the sound level at high frequencies culminated around 18 cm H_2O. The behavior above 8 kHz indicates that there may exist a tonal component of comparable amplitude at high flow rates at a frequency in between the two tube resonance frequencies. The tonal frequency increases with the jet velocity, and, when superimposed on the broadband component, may contribute to the complexity in the overall spectral shape and amplitude. This is discussed further in Sec. V.

Efforts were made to separate the source function using the spectral decomposition method. The resulting G_1 function, shown in Fig. 11, does not seem to obey a multipole power law of the type He^n as for the straight and diverging cases. Significant scattering from the ensemble mean

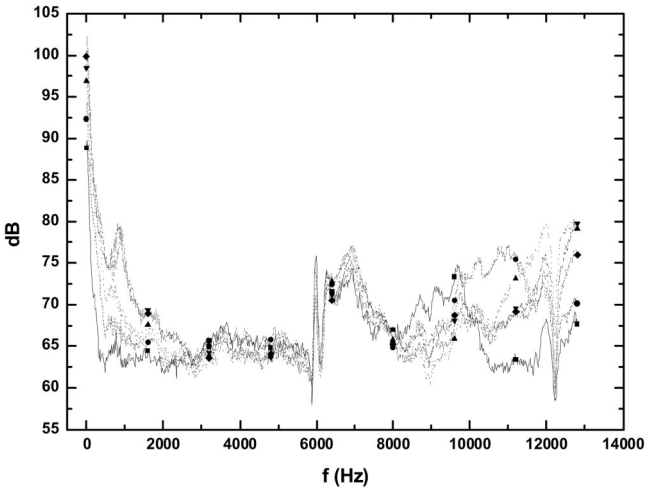


FIG. 10. Sound spectra measured at high flow rate in CO_2. Divergent glottis-shaped orifice. ■, $\Delta p = 10$ cm H_2O; ●, $\Delta p = 14$ cm H_2O; ▲, $\Delta p = 18$ cm H_2O; ▼, $\Delta p = 22$ cm H_2O; ♦, $\Delta p = 26$ cm H_2O.

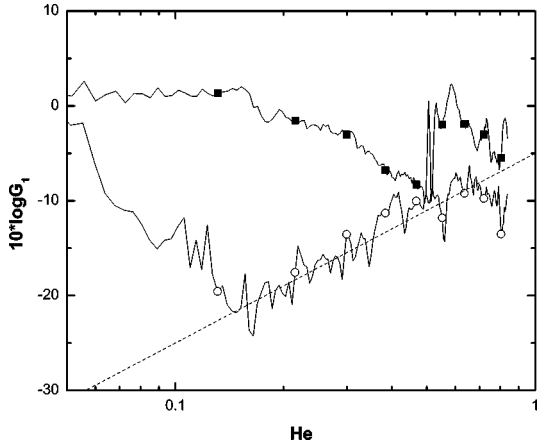


FIG. 11. Acoustic frequency response versus Helmholtz number. ■: divergent glottis-shaped orifice, confined flows; ○, divergent glottis-shaped orifice, free jet flows; - - -, 20 log He.

$10 \log F$ function was observed among the individual $10 \log F^k$ functions, with a standard deviation of 5 dB as compared to the approximate 15 dB dynamic range for $10 \log F$. Since it is unlikely any other sound generating mechanisms was present, this characteristic behavior was attributed to some kind of flow-acoustic interaction in this case. Vortex shedding, pairing, roll-up, and growth of instability waves are very sensitive to small acoustic disturbances. Any small disturbances reflected back to the orifice region could affect the near field flow, in particular flow separation and vortex shedding. A feedback loop may therefore exist. The confined tube configuration could definitely increase acoustic disturbances and enhance this kind of interaction, invalidating the basic source-filter model assumption.

This is substantiated by results obtained with the downstream tube removed, in an open jet configuration, shown in Fig. 12. The measurements were made on the jet axis, 16 cm away from the jet exit plane. The spectra are more predictable, with seemingly constant level increases with pressure,

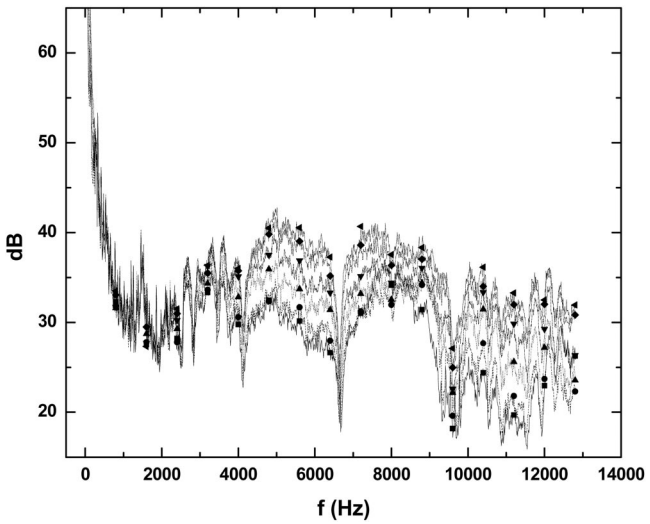


FIG. 12. Sound spectra measured at different high flow rate in air. Free jet flow through a divergent glottis-shaped orifice. ■, $\Delta p = 10$ cm H_2O; ●, $\Delta p = 12$ cm H_2O; ▲, $\Delta p = 16$ cm H_2O; ▼, $\Delta p = 20$ cm H_2O; ♦, $\Delta p = 26$ cm H_2O; ◀, $\Delta p = 30$ cm H_2O.

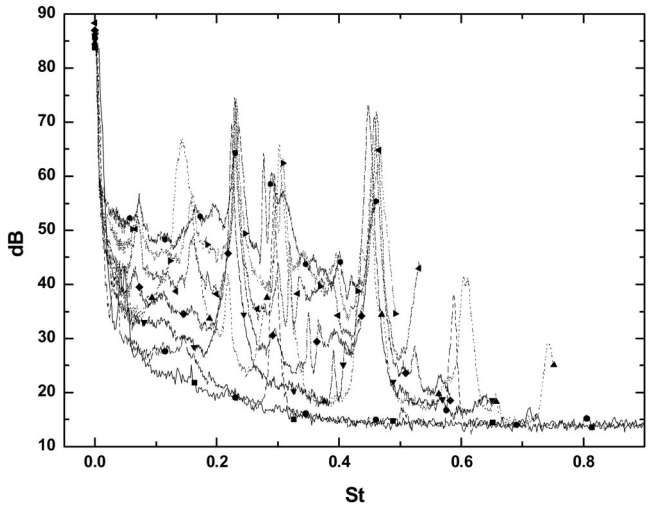


FIG. 13. Sound spectra measured at low transglottal pressures in air. Divergent glottis-shaped orifice. Pure tones were observed. ■, $\Delta p = 1$ cm H_2O; ●, $\Delta p = 2$ cm H_2O; ▲, $\Delta p = 3$ cm H_2O; ▼, $\Delta p = 4$ cm H_2O; ♦, $\Delta p = 5$ cm H_2O; ◀, $\Delta p = 6$ cm H_2O; ▶, $\Delta p = 7$ cm H_2O; ⬢, $\Delta p = 8$ cm H_2O.

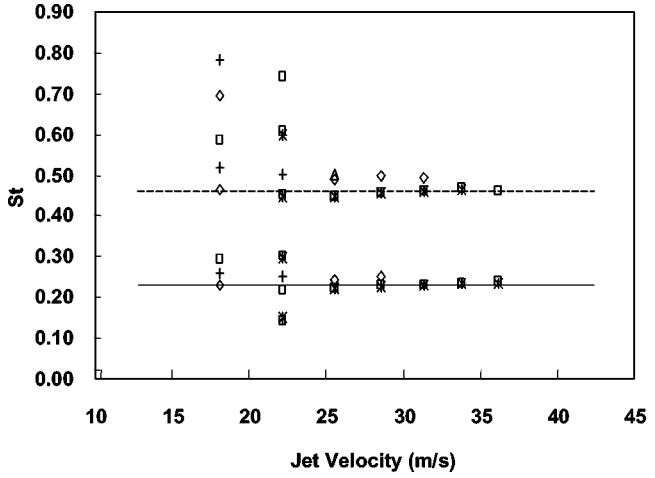


FIG. 14. The Strouhal number of the observed jet tones for a divergent glottis-shaped orifice versus the jet velocity. □, confined air jet; △, confined CO_2 jet; ◇, confined He–CO_2 jet;*, free air jet; +, free CO_2 jet;—, St = 0.23; - - -, St = 0.46.

except at low frequencies because of background noise attributed to the flow supply. The spectral decomposition method was again applied for this case. As shown in Fig. 11, the G_1 function follows a He^2 power law for mid- and high-Helmholtz numbers. This suggests that the sound source is dipolelike in this case (i.e., for a divergent orifice). Due to the unstable flow separation within the glottis, there exists significant interaction between the shed vortices and the orifice wall. The resulting dipole source then tends to dominate the sound field. The quadrupole source, which is associated with the turbulence region downstream of the orifice, is therefore less important. At lower Helmholtz number, the G_1 function tends to decrease with the Helmholtz number, which may again be due to significant background noise in the free jet configuration. Other errors may occur since the sound pressure measurements were made in the acoustic near field at low frequencies. The fluctuations in the G_1 function are believed to be due to reflections off surfaces near and downstream of the jet exit. The F function is plotted against the Reynolds number and the Strouhal number in Figs. 6 and 7, respectively. The source spectra culminate at midfrequencies. For Strouhal number values lower and higher than the peak Strouhal number, the F function approximately follows St^2 and St^{-2} trends, respectively.

The significant difference observed in the sound generation between the confined jet and free jet configurations for large flow velocities indicates that significant source-tube acoustic interaction may exist in the confined jet configuration for the divergent orifice shape. Such interaction was not evident in the results for the other orifice geometries studied.

V. TONAL SOUND GENERATION

Tonal sounds were produced for low flow rates through the stationary divergent orifice configurations, for flow velocities ranging from 17 to 37 m/s. Figure 13 shows typical sound pressure spectra measured with air for transglottal pressures between 1 cm H_2O and 8 cm H_2O. The frequency of the tones increased nearly linearly with the jet centerline velocity. As the jet velocity was increased, broadband sound increased gradually in level until it reached that of the tonal component. Eventually, as the transglottal pressure was increased above $\Delta p = 9$ cm H_2O (Fig. 9), the broadband component dominated although tones were still present in the spectra in some cases.

Figure 14 shows the Strouhal number of the tonal sounds versus the jet velocity. The hydraulic diameter of the orifice was used as the characteristic dimension in the definition of the Strouhal number. Both the fundamental, St ≈ 0.23, and the first harmonic, St ≈ 0.46, of the tonal sounds are shown. These values correspond to values reported for vortex shedding in shear layers, for example, for flow past a cylinder (Blake, 1986). It was found, by varying the working fluid, that the effects of viscosity on the tonal sounds Strouhal number were insignificant. The effects of the orifice dimensions are unknown, since the experiments were performed with the orifice area kept constant. At low flow rates (2 and 3 cm H_2O in air, or 18.08 and 22.14 m/s for the jet velocity), the fundamental Strouhal number was only 0.15 in one case, and Strouhal number values for tonal components were scattered over a range.

The tone Strouhal numbers and their onset velocities were consistent for different tube configurations. For example, radiated sound spectra measurements were repeated with the downstream tube removed to eliminate possible flow-acoustic loading feedback. The Strouhal numbers of the tonal sounds, shown in Fig. 14, were found to coincide with their confined jets counterparts. This suggests that the jet tones may be caused by a hydrodynamic phenomenon, and not by flow-acoustic interactions as the case of broadband sound generation at high flow rates discussed in Sec. IV C.

For one case, the instantaneous spanwise velocity profile within the jet was measured using a gated hot-wire probe located 0.5 mm downstream from the orifice exit. The probe was traversed every 0.1 mm along a 3 mm span, which in-

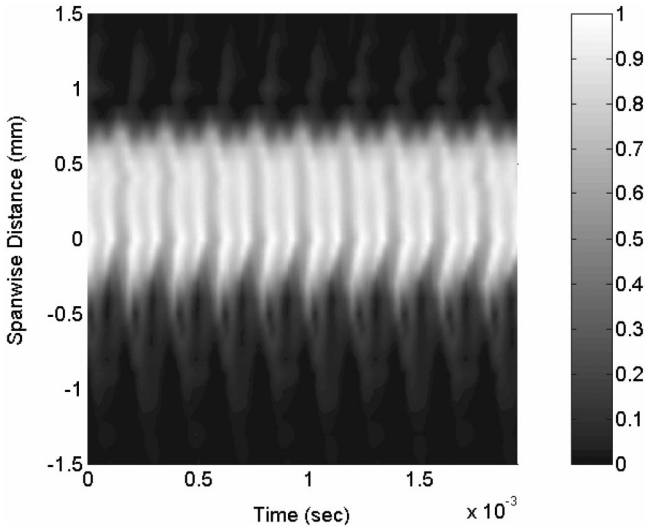


FIG. 15. Time history of the normalized jet velocity along the spanwise direction in presence of the jet tone. The center of the orifice is at the zero spanwise distance.

cluded the entire jet plume. The data acquisition was triggered and phase-locked at each location by using a filtered, nearly sinusoidal sound pressure signal from the microphone located in the tube upstream of the orifice. The instantaneous velocity profile for a few cycles is shown in Fig. 15. The frequency of the tone was 5.1 kHz, and the transglottal pressure 6 cm H_2O. The velocity oscillated in phase throughout the span of the orifice. The harmonic component at twice the fundamental frequency was strong. There was no apparent fluctuation in the jet core location over one cycle, and thus no apparent flipping of the jet plume from one diffuser wall to the other. The jet was permanently attached to one wall, as shown in Fig. 16, as found using an oil-based smoke injected into the air stream, and a strobe light. For a transglottal pressure of 10 cm H_2O, the jet plume is clearly skewed as a result of attachment to one of the glottal walls. It is not clear whether this was due to asymmetry in the orifice, or to a so-called Coanda effect. The jet plume discharge angle was bi-stable, and could be manipulated through slight adjustments in the orifice geometry. With very careful positioning, the jet plume was normal to the exit plane in a state of unstable equilibrium.

It is possible that shear layer instabilities within the jet may cause a slight shift in the flow separation point within

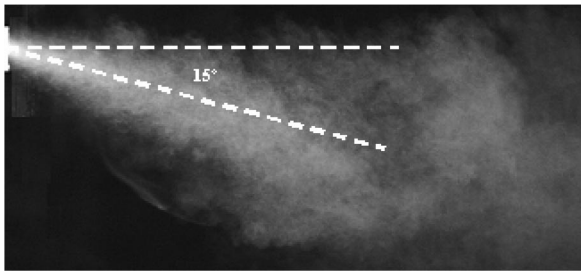


FIG. 16. Flow visualization of steady free jet just beyond divergent orifice. The orifice exit plane roughly corresponds to the left edge of the image. It can be seen that the jet is here attached to the lower surface of the orifice, causing it to exit at an angle corresponding to orifice wall angle (15°).

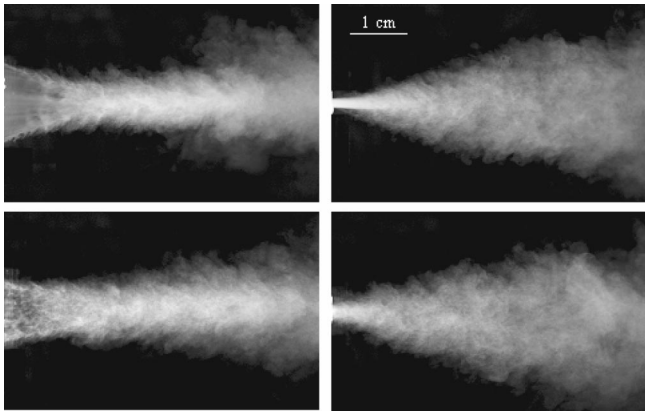


FIG. 17. Flow visualization of free jet just beyond the convergent (top) and divergent (bottom) orifices. $\Delta p = 4$ cm H_2O. The jet in the first 5 to 10 mm is clearly more stable in the convergent case than for the divergent case.

the divergent orifice. This could change the flow resistance of the orifice and result in a surge, or periodic ''blocking'' of the orifice. Note that this phenomenon did not occur for convergent orifices. Flow separation, or stall, is known to occur under adverse pressure gradient for jet flows through divergent channels. Over a certain range of diffuser angle and ratio of orifice length to orifice throat width, large transitory stalls occurs and the entire flow region is characterized by a large, low frequency surge or stall (Kline, 1959). Such dynamic stall induces pressure fluctuations and unsteady flow behavior both upstream and downstream of the orifice (Smith, 1975). Similar observations were made in studies of jets flow through knife-edge orifices by Johansen (1929), and Beavers and Wilson (1970). Vortical structures shed at a constant Strouhal number over the range $200 < \mathrm{Re} < 3500$ were observed. The Strouhal number observed of the vortical structures for the two-dimensional orifice was around 0.43, which is approximately twice the value for the jet tone observed in present study. This could be accounted for by vortex pairing, with subharmonics excited and contributing to the sound radiation. At low flow rates, paired vortices may pair again with another vortex giving rise to oscillations at one-third of the vortex shedding frequency in the sound field.

VI. EFFECTS OF ORIFICE GEOMETRY

For the cases of a circular straight orifice (Zhang *et al.*, 2002a), as well as for the straight and convergent glottis-shaped orifices, the flow within and immediately downstream of the orifice is laminar and stable. Instability waves appear several orifice diameters downstream, beyond which the flow transitions to turbulence abruptly. Figure 17 shows two pictures of the flow just downstream of the orifice exit plane for the convergent orifice (unconfined flow). The laminar core may be seen, along with vortical structures generated downstream due to shearing and mixing. The sound generation process for these cases exhibits all more of a quadrupole type behavior, which may due to the fact that turbulence appears several glottal diameters downstream in the tubes and thus there is less random force fluctuations on the orifice. The source functions for these three cases follow

similar spectral similarity laws. Note for convergent shapes, Fig. 17, the merging or collision of the corner vortices could affect the quadrupole source strength.

For a divergent orifice, the flow is turbulent right beyond the glottis exit as shown in Fig. 17 for one free divergent jet and 4 cm H_2O transglottal pressure. The flow separates within the orifice and is very sensitive to small disturbances. The resulting hydrodynamic and acoustic feedback loops are responsible for the tonal sound observed at low flow rates and the turbulent sound spectral behavior at high flow rates, respectively.

Although the divergent orifice geometry has significant effects on the flow field and sound generation process in the steady jet flows, it is not clear that whether these observed feedback mechanisms would occur in the unsteady flow configuration, as in voice production. It is possible these feedback loops may not have enough time to develop considering the constantly changing pulsating flows. On the other hand, the source spectral distribution functions for the other two orifice geometries and the circular orifices have been shown to follow similar trends. This scaling law in the source spectral shape and amplitudes may be used for the noise modeling in high-quality speech synthesis.

VII. CONCLUSIONS

Experiments were performed to study sound radiation from stationary jet flows through glottis-shaped orifices. Three orifice geometries were studied with straight, divergent, and convergent profiles. For the straight and convergent orifices, the sound generation characteristics were found to be similar to those of circular orifices, i.e., quadrupole type of source. The source spectral distribution functions for the two orifice geometries follow similar trends, decreasing approximately with the third power and fifth power of the Strouhal number at low and high Strouhal numbers, respectively. For the divergent orifice, jet tones were produced at low flow rate, with a fundamental Strouhal number around 0.23, which is probably related to a dynamic surge phenomenon. At high flow rates, the sound field is dominated by broadband sound. Significant differences in the sound field between the confined divergent jets and the free divergent jets were observed at high flow rates. The free jet sound was found to be a dipole type of source. The confined jets exhibit a different complex behavior, which is probably due to flow-acoustic interactions and the existence of a feedback loop. Extended flow visualization and detailed flow measurements are needed to further understand the flow field and characterize possible sound generation mechanisms.

ACKNOWLEDGMENTS

This study was supported by Research Grant No. RO1 DC03577-06 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

Barney, A., Shadle, C. H., and Davies, P. O. A. L. (**1999**). ''Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory,'' J. Acoust. Soc. Am. **105**, 444–455.

Beavers, G. S., and Wilson, T. A. (**1970**). ''Vortex growth in jets,'' J. Fluid Mech. **44**, 97–112.

Blake, W. K. (**1986**). *Mechanics of Flow-Induced Sound and Vibration* (Academic, New York).

Dabiri, J., Kucinschi, B., Afjeh, A., Dewitt, K., Scherer, R., and Agarwal, M. (**2001**). ''Smoke flow visualization for the flow through a symmetrical and an oblique glottis with and without the presence of the false vocal folds,'' Research report, Optical flow diagnostics laboratory, University of Toledo.

Davies, H. G., and Ffowcs Williams, J. E. (**1968**). ''Aerodynamic sound generation in pipe,'' J. Fluid Mech. **32**, 765–778.

Flanagan, J. L. (**1965**). *Speech Analysis, Synthesis, and Perception* (Springer-Verlag, Berlin).

Goldstein, M. E. (**1976**). *Aeroacoustics* (McGraw-Hill, New York).

Harper, V. H. (**2000**). ''Respiratory tract acoustical modeling and measurements,'' Ph.D. thesis, Purdue University, West Lafayette.

Hofmans, G. C. J. (**1998**). ''Vortex sound in confined flows,'' Ph.D. thesis, Universiteis-drukkerij, TUE, Ben Mobach, TUE.

Howe, M. S. (**1998**). *Acoustics of Fluid-Structure Interactions* (Cambridge University Press, Cambridge).

Johansen, F. C. (**1929**). ''Flow through pipe orifices at low Reynolds numbers,'' Proc. R. Soc. London, Ser. A **216**, 231–245.

Kline, S. J. (**1959**). ''On the nature of stall,'' J. Basic Eng. **81**, 305–320.

McGowan, R. S. (**1988**). ''An aeroacoustic approach to phonation,'' J. Acoust. Soc. Am. **83**, 696–704.

Meyer-Eppler, W. (**1953**). ''Zum Erzeugunsmechanismus der Geraus-chlaute,'' Zeitschrift fur Phonetik und allgemeine Sprachwissenschaft, **7**, 196–212.

Moore, C. J. (**1977**). ''The role of shear-layer instability waves in jet exhaust noise,'' J. Fluid Mech. **80**, 321–367.

Shadle, C. H., Elliott, S. J., and Nelson, P. A. (**1987**). ''Visualization of the air flowing through a dynamic model of the vocal folds,'' ISVR Technical Report 154.

Smith, C. R. (**1975**). ''A note on diffuser generated flow unsteadiness,'' J. Fluids Eng. **97**, 377–379.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002a**). ''Broadband sound generation by confined turbulent jets,'' J. Acoust. Soc. Am. **112**, 677–689.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002b**). ''Experimental verification of the quasisteady approximation for aerodynamic sound generation by pulsating jets in tubes,'' J. Acoust. Soc. Am. **112**, 1652–1663.

An acoustic description of the vowels of Northern and Southern Standard Dutch

Patti Adank[a)]
School of Psychology, University of Liverpool, Bedford Street South, L69 7ZA, Liverpool, United Kingdom

Roeland van Hout
Department of Linguistics, University of Nijmegen, PO box 9103, 6500 HD, Nijmegen, The Netherlands

Roel Smits
Max Planck Institute for Psycholinguistics, PO box 310, 6500 AH Nijmegen, The Netherlands

(Received 9 December 2003; revised 1 May 2004; accepted 15 June 2004)

A database is presented of measurements of the fundamental frequency, the frequencies of the first three formants, and the duration of the 15 vowels of Standard Dutch as spoken in the Netherlands (Northern Standard Dutch) and in Belgium (Southern Standard Dutch). The speech material consisted of read monosyllabic utterances in a neutral consonantal context (i.e., /sVs/). Recordings were made for 20 female talkers and 20 male talkers, who were stratified for the factors age, gender, and region. Of the 40 talkers, 20 spoke Northern Standard Dutch and 20 spoke Southern Standard Dutch. The results indicated that the nine monophthongal Dutch vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ can be separated fairly well given their steady-state characteristics, while the long mid vowels /e o ø/ and three diphthongal vowels /ɛɪ ɔu œy/ also require information about their dynamic characteristics. The analysis of the formant values indicated that Northern Standard Dutch and Southern Standard Dutch differ little in the formant frequencies at steady-state for the nine monophthongal vowels. Larger differences between these two language varieties were found for the dynamic specifications of the three long mid vowels, and, to a lesser extent, of the three diphthongal vowels. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779271]

PACS numbers: 43.70.Fq, 43.70.Kv, 43.72.Ar [AL] Pages: 1729–1738

I. INTRODUCTION

This paper presents a description of the acoustic characteristics of the 15 vowels of Standard Dutch as spoken in the Netherlands and in Belgium (Flanders). Previous descriptions of the vowels of Standard Dutch are given in Pols *et al.* (1973) and Van Nierop *et al.* (1973).[1] These studies describe the Dutch vowels in terms of the average frequencies and standard deviations of the first three formants. Pols *et al.* recorded 1 token of the 12 monophthongal vowels in an /hVt/ context, produced by 50 male talkers, and Van Nierop *et al.* recorded one token of each of the 12 monophthongal vowels in an /hVt/ context, produced by 25 female talkers.

The acoustic measurements by Pols *et al.* (1973) and Van Nierop *et al.* (1973) have been recognized as representing the standard formant values for the vowels of Dutch and were used in studies on vowel perception and talker normalization (e.g., Disner, 1980; Syrdal and Gopal, 1986). However, in our view the description of the acoustic characteristics of the vowels of Dutch as presented by Pols *et al.* and Van Nierop *et al.* is limited in four respects:

(a) Pols *et al.* and Van Nierop *et al.* do not provide information about dynamic properties such as vowel duration and spectral change (formant measurements were presented for a single time slice); information that has been found to play a central role in vowel perception (Di Benedetto, 1989a, b; Hillenbrand and Gayvert, 1993; Nearey, 1989, Strange *et al.* 1983; Strange, 1989), and that reflects differences in language varieties (Hagiwara, 1997).

(b) Not all vowels of Dutch were included by Pols *et al.* and Van Nierop *et al.*, as no description was given of the acoustic characteristics of the three diphthongs /ɛɪ ɔu œy/.

(c) No information is given in the two studies about the regional background of the talkers; the formant measurements described in Pols *et al.* and Van Nierop *et al.* may therefore display uncontrolled regional variation.[2]

(d) No recordings were made of talkers from Flanders, the Dutch-speaking part of Belgium, thus excluding the southern variety of Standard Dutch from the description.

The purpose of the present study was twofold. First, we aim to give a description of the Dutch vowel sounds. In this description we want to improve on the four limitations mentioned above. Second, we aimed to give an overview of the similarities and dissimilarities of the vowel systems of Northern Standard Dutch and Southern Standard Dutch.

The present study was set up as follows. A total of 1200 vowel tokens were recorded, spoken by 40 talkers, 20 male talkers and 20 female talkers. These talkers were screened for their age, gender, regional background, and socioeconomic status. Each talker produced two tokens of all 15 vowel sounds of Dutch, /ɑ a ɛ ɪ i ɔ u ʏ y e o ø ɛɪ ɔu œy/. Each token was produced in a carrier sentence, in a /sVs/

a)Electronic mail: adank@liv.ac.uk

context. For each token, measurements were made of the duration, the fundamental frequency, and the first three formants at nine time slices in the vowel duration.

II. MATERIALS

A. Database design

Our materials were taken from a large database, comprising recordings of 160 talkers of standard Dutch who were stratified for the following sociological variables: speech community (country), regional background, gender, and age. All talkers were teachers of Dutch at secondary education institutes at the time the interview was recorded. They were required to be teachers of the Dutch language for three reasons. First, Dutch teachers are professional language users; they are expected to speak standard Dutch on a daily basis. Second, they are instructors of the standard language and can thus be regarded as having a normative role. Third, Dutch teachers' speech was expected to show more regional variation than broadcasters', whose speech is used in other pronunciation studies of the standard language (cf. Bell, 1983; Deterding, 1997).

There is a long-standing discussion whether the speech communities of the Netherlands and Flanders have one or two Dutch standard languages. It is evident that there is a Netherlandic and a Flemish standard variety, with clear differences on all linguistic levels, certainly on the phonetic level (cf. Van de Velde *et al.*, 1997). The historical background of the development of the Dutch standard language in relation to the speech communities of the Netherlands and Flanders is too complex to summarize here. Van de Velde *et al.*, (1997) present a short overview, including the relevant literature. Research data available produce convincing evidence that many differences exist between the two speech communities on the level of pronunciation. The data set was therefore split into a Dutch and a Belgian component, referred to as Northern Standard Dutch (NSD) and Southern Standard Dutch (SSD), respectively.

In each community, four regions were selected: a central region, an intermediate region, and two peripheral regions. The central region is the economically and culturally dominant region in each of the speech communities. For the Netherlands, the central region is the west, consisting of the provinces of Northern-Holland, Southern-Holland, and Utrecht, also known as the ''Randstad,'' referred to as ''N-R'' (Netherlands-Randstad). The cities Amsterdam, Rotterdam, Utrecht, and The Hague are part of the Randstad. In Flanders, the central region is ''Brabant,'' or ''F-B'' (Flanders-Brabant). Brabant comprises the Belgian provinces ''Antwerpen'' and ''Flemish-Brabant,'' with the cities of Antwerpen and Leuven, respectively.

In this paper, we limit our description of the acoustic characteristics of the vowel tokens produced by the 40 Dutch teachers from the two ''central'' regions N-R and F-B, respectively. Therefore, whenever we refer to Northern Standard Dutch, the N-R region is meant, and whenever we refer to Southern Standard Dutch, the F-B region is meant. See Adank (2003) for a description of the remaining six regions.

Several towns were selected per region, following three criteria. First, the selected towns in each region had a comparable socioeconomic profile. Second, the towns within each region belonged to the same dialect group. Third, the Dutch spoken in that town was regarded as characteristic of that region. No major cities were selected, because it was expected that the Dutch spoken in major cities is influenced by dialects (or languages) other than the dialects spoken in the surrounding region, due to migration. For N-R, two towns were selected: Alphen aan de Rijn and Gouda; for F-B the two towns selected were Lier and Heist-op-den-berg.

The teachers who participated in the interview taught at schools for secondary education in the selected towns. They had to meet the following requirements. First, at the time of the interview, they all lived in one of the selected towns, or near that town in the dialectal region characteristic for that region. Second, they were born in the region or moved there before their eighth birthday. Third, they had lived in the region for at least eight years prior to their 18th birthday. This last requirement was formulated on the basis of studies by Payne (1980) and by Scovel (1988). Payne stated that children younger than 8 years old have no difficulty acquiring the phonological system of the place they moved to. Scovel reported that learners of a second language generally do not acquire near-native pronunciation of this language after puberty. This last requirement was therefore used to make sure that the talkers had lived in the town/region from an age at which they had no difficulties in learning the language variety spoken in that region/town. Finally, the talkers were divided into two age groups, a younger group and an older group. The talkers in the younger group were between 22 and 44 years old at the time of the interview and talkers in the older group were between 45 and 60 years old. NSD and SSD thus both consisted of 20 talkers: five young men, five older men, five younger women, and five older women.

B. Recordings

Dutch vowels are traditionally divided into phonologically short vowels, /ɑ ɛ ɪ ɔ ʏ/, phonologically long vowels /a e i ø o u y/, diphthongs, /ɛɪ ɔu œy/, and schwa, /ə/ (Booij, 1995). Recordings were made for all vowels except for schwa, because it does not occur in stressed syllables in Dutch.[3] All target vowels were produced in a carrier sentence. The sentences had the following generic structure for the short vowels (''V'' indicates the target vowel):

In sVs en in sVsse zit de V
/ɪn sVs ən ɪn sVsə zɪt də V/
[*In sVs and in sVsse is the V*]

The sentences had the following structure for the long vowels and the diphthongs:

In sVs en in sVze zit de V
/ɪn sVs ən ɪn sVzə zɪt də V/
[*In sVs and in sVze is the V*]

Of the three different consonantal contexts (CVC, CVCV, or V), the CVC contexts were selected for further processing. The CVC structure /sVs/ can be regarded as a neutral context for Dutch vowels. It was decided not to use the ''traditional'' /hVt/ structure (/hVd/ for English), because it could not be predicted how people from Flanders would

pronounce the word-initial /h/. Almost all people interviewed in Flanders are originally dialect talkers; they usually speak Southern Standard Dutch as well as their (native) dialect. Some Flanders dialects do not have /h/ in their phoneme inventory. When asked to produce a word-initial /h/, Flemish dialect talkers tend to replace /h/ by either a glottal stop, or by /χ/. The /sVs/ context was adopted to make sure that no additional sources of variation were added to the data.

The vowel tokens were recorded as a task in a so-called ''sociolinguistic interview'' in which vowels and consonants were elicited in a wide variety of tasks. The carrier sentences were presented to the talker on a computer screen, with a three-second interval between sentences. When the talker made a mistake, the interviewer interrupted the computer program and went back at least two sentences and asked the talker to read these sentences again, to make sure that all sentences were recorded correctly. The carrier sentences task was performed twice during the interview; each vowel token was therefore available twice in each syllabic structure. A total of 1200 vowel tokens were recorded: two tokens of each of the 15 vowel categories of Dutch, produced by 40 talkers.

The recordings were made on DAT with a TASCAM DA-P1 portable DAT-recorder, with an AKG C420 Headset condenser microphone. The recordings were digitized through a Lucid Technology PCI24 digital audio card, and stored at 48 kHz on a PowerMac 7500/100. The neutral context sentences were down-sampled to 16 kHz, using an non-causal FFT anti-aliasing filter set to 8 kHz.

Recording conditions were different for each of the talkers. Some were interviewed in an empty classroom and others were interviewed at their own home. Due to these differences in recording conditions, in rare cases, background noises were audible. Whenever this was the case, the speech segment was excluded from further analysis.

C. Acoustic measurements

1. Duration

The start and end times for the duration of each token were labeled manually in the digitized speech wave, using the program Praat, version 4.02 (Boersma, 2001). We define the duration as the interval between the onset and offset of the glottal vibrations of the vocalic portion of the /sVs/ syllable. When labeling, it was ensured that the surrounding speech sounds were not audible in the remaining signal. Segment labels were placed at zero crossings. The duration of each vowel segment was defined as the interval between the segment labels at the start and end of the vocalic portion.

2. Fundamental frequency

For each vowel token, $F0$ was extracted automatically with Praat using an autocorrelation-based procedure that was evaluated as the best option available in Praat (Boersma, 1993). The upper and lower limits for the autocorrelation peaks were set at 50 and 300 Hz for male talkers and at 100 and 500 Hz for female talkers, respectively. The size of the Hanning window was chosen such that it included at least three glottal pulses and was set to 60 ms for the male talkers and to 33.3 ms for the female talkers. $F0$ was extracted at the temporal mid-point of the vowel token's duration.

The resulting measurements were checked for outliers (cases 1.5–3 times the interquartile range, or IQR), and extreme values (cases >3 IQR) for each talker. Every outlier and extreme value was manually verified. Eleven cases were replaced by the mean $F0$ for the talker, usually because voice characteristics (e.g., hoarseness) of the talker prevented reliable $F0$-estimation.

3. Formant contours

The formant contours for $F1$, $F2$, and $F3$ were estimated using a program for automatic formant tracking (Nearey *et al.*, 2002).[4] Nearey *et al.*'s program consists of two parts: a formant tracking algorithm and a user interface that allows the user to verify, and, where necessary, adjust the formant tracks generated by the tracking algorithm. Its preprocessing consists of applying a 25-ms cosine4 window with a time-step of 2 ms. Subsequently, three formant candidates are estimated by means of root extraction, using a version of Markel and Gray's (1976) ''FORMNT'' algorithm, followed by a five-point running median smoothing. The number of LPC coefficients is fixed at nine. For each vowel token, these settings were identical.

Some settings for Nearey *et al.*'s tracking algorithm are to be chosen by the user. The tracking program cycles through several cutoff frequencies for each individual vowel token. The frequency range consists of two parts, a lower range and an upper range. The lower range is fixed between 0 and 3000 Hz, while the upper range is to be set by the user between 3000 Hz and the highest cutoff frequency (hc) to be evaluated. Within this upper range (3000-hc) the user must also provide the total number of cutoff frequencies (nc) to be evaluated. Finally, the user has to decide whether the distance (d) between the cutoff frequencies is spaced logarithmically (Log) or linearly (Hz) across the upper range.

We evaluated several combinations of settings for hc, nc and d and found that the best results were obtained with hc set to 4200 Hz, nc set to 5, and d set to Log (Adank, 2003). Formant tracks for all vowel tokens were measured with these settings.

The resulting formant tracks for each vowel token were verified by hand in the user interface of the program. Here, the tracks were plotted on the smoothed spectrogram. The course of each track could be altered, which was done in 20%–25% of the cases.

Once the experimenter was satisfied with the position of the formant tracks, the frequencies of the first three formant tracks were stored at nine points ($t1-t9$) of the vowel's duration, with the first point, $t1$, at the start of the vocalic portion of the token, and the ninth point, $t9$, at the end of the vocalic portion, and the remaining points spaced at equal-sized intervals, relative to the absolute duration of the token. Whenever we refer to the vowel's temporal midpoint, the fifth point of the total of nine points, or $t5$, is meant. We only present results for points $t3$ (25%) through $t7$ (75%), because we suspected that, the measurements at that point were influenced less by the consonantal context than at $t1$ and $t9$, or even at $t2$ and $t8$.

TABLE I. Average durations (ms), fundamental frequencies (Hz), and formant frequencies (Hz) for the first three formants for the vowels tokens produced by the 20 NSD talkers and the 20 SSD talkers. F, female talkers; M, male talkers, Dur, duration. *N* per cell is 20. *F*0, *F*1, *F*2, and *F*3 were sampled at 50% (*t*5) of the steady-state portion of each vowel token.

			/ɑ/	/a/	/ɛ/	/e/	/ø/	/ɪ/	/i/	/ɔ/	/u/	/o/	/ʏ/	/y/	/ɔu/	/ɛɪ/	/œy/
NSD	Dur	F	94	214	101	177	184	89	92	96	98	183	89	96	205	199	200
		M	96	203	95	181	184	82	94	90	98	184	88	93	195	192	189
	*F*0	F	226	194	220	207	201	221	248	218	249	201	246	245	201	200	201
		M	149	134	154	131	142	154	157	152	164	139	154	162	127	133	136
	*F*1	F	758	912	535	442	445	399	294	419	286	445	417	305	715	659	693
		M	578	670	475	400	375	361	278	402	259	412	366	259	580	543	574
	*F*2	F	1280	1572	1990	2343	1713	2276	2524	918	938	964	1830	1918	1280	2097	1679
		M	1172	1425	1739	1995	1563	1919	2162	821	805	929	1595	1734	1117	1804	1515
	*F*3	F	2895	2852	2871	2908	2550	2883	2911	3013	2736	2417	2711	2635	2586	2816	2618
		M	2435	2485	2492	2583	2241	2536	2665	2851	2253	2306	2345	2205	2284	2498	2280
SSD	Dur	F	107	240	101	192	200	88	147	97	128	210	89	153	245	238	241
		M	90	204	86	169	175	76	96	83	99	182	77	109	212	199	212
	*F*0	F	225	203	224	219	217	256	234	233	237	215	249	236	196	202	201
		M	126	116	128	119	121	135	148	136	149	125	138	144	113	124	114
	*F*1	F	725	868	581	436	439	455	317	475	321	418	457	337	696	670	696
		M	555	717	475	384	374	364	278	398	266	369	353	265	549	545	583
	*F*2	F	1262	1640	1932	2420	1804	2115	2647	987	1019	968	1785	2077	1282	2159	1762
		M	1066	1429	1616	1993	1539	1745	2179	850	978	862	1492	1825	1127	1779	1484
	*F*3	F	3041	3031	2978	3021	2666	2948	3312	3133	2871	2992	2884	2634	2808	3040	2855
		M	2655	2651	2572	2616	2377	2566	2787	2665	2422	2540	2514	2348	2444	2533	2415

III. RESULTS

Table I gives the average measurement results for the true monophthongal vowels, for the five acoustic variables (duration, *F*0, *F*1, *F*2, and *F*3), taken at the temporal midpoint *t*5 (50%), for the NSD and the SSD talkers. The average values at *t*3 (25%) and *t*7 (75%) for /ɔu œy ɛɪ/ and the long mid vowels /e ø o/ are displayed in Table II.

1. Duration

Figure 1 shows the average durations for all 15 vowels for the male talkers and Fig. 2 for the female talkers. A first observation is that the vowels in Figs. 1 and 2 can roughly be divided into two groups based on their duration: long ones, i.e., the three diphthongal vowels /ɛɪ ɔu œy/, the three long mid vowels /e o ø/ plus /a/, and short ones, i.e., all full monophthongal vowels /ɑ ɛ i ɪ ɔ u y ʏ/ except /a/. A series of *t*-tests was carried out to establish which vowel pairs differed significantly in duration. A total of 105 pairwise comparisons were carried out for the 15 vowels. These analyses were done separately for the NSD men, the NSD women, the SSD men, and the SSD women. Because of the large number of analyses, a Bonferroni correction was carried out, setting the significance level to $\alpha = 0.001$. The results showed that /a e ɛɪ ɔu o ø œy/ all have significantly longer durations than /ɑ ɛ ɔ u y ʏ/.

Figures 1 and 2 further show longer durations for female SSD talkers, for a larger subset of the vowels. To test the significance of possible effects of regional background and gender, we carried out a repeated measures analysis of variance (ANOVA) on the duration measurements. In the ANOVA, the within-subject factor VOWEL consisted of the duration measurements for the 15 vowels, while the regional

TABLE II. Average formant frequencies (Hz) for the first three formants for the vowels tokens produced by the 20 NSD talkers and the 20 SSD talkers. F, female talkers; M, male talkers. *N* per cell is 20. *F*1, *F*2, and *F*3 were sampled at 25%: "*t*3" and 75%: "*t*7" of the steady-state portion of each vowel token.

			/ɔu/		/ɛɪ/		/œy/		/e/		/ø/		/o/	
NSD			*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7
	*F*1	F	763	565	692	527	740	530	452	402	468	418	485	407
		M	633	466	557	450	592	457	415	361	407	329	444	365
	*F*2	F	1419	1128	2019	2225	1675	1675	2241	2425	1726	1737	1078	937
		M	1254	994	1733	1920	1485	1555	1887	2075	1544	1592	1015	848
	*F*3	F	2656	2616	2817	2823	2679	2594	2890	2940	2574	2552	2489	2417
		M	2320	2340	2502	2586	2357	2261	2545	2597	2254	2229	2258	2371
SSD			*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7	*t*3	*t*7
	*F*1	F	741	564	666	587	679	603	438	430	437	424	435	408
		M	612	428	575	478	588	488	386	376	377	355	385	347
	*F*2	F	1398	1158	2111	2270	1742	1856	2349	2425	1796	1828	1077	927
		M	1225	985	1704	1899	1449	1584	1942	2020	1534	1562	965	844
	*F*3	F	2806	2888	3042	3081	2868	2825	3007	2997	2697	2685	2922	3040
		M	2441	2544	2539	2553	2439	2404	2616	2611	2422	2387	2472	2576

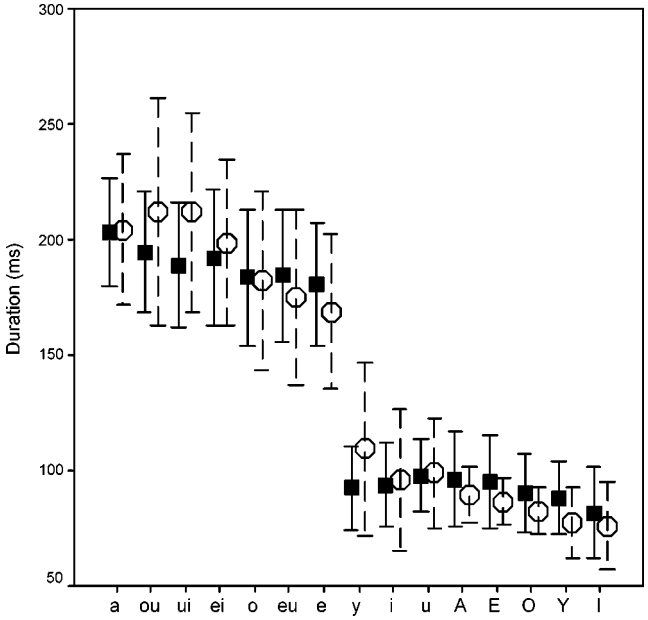


FIG. 1. Error plot of average duration (ms) per vowel, for the men for the two regions: NSD (filled squares) and SSD (open circles) (“ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

background of the talker (region) and the talker’s gender (gender) served as between-subject factors. The analysis showed that the longer durations for the female talkers in Figs. 1 and 2 are significant: gender had a significant effect on duration ($F[1,36]=5.77$, $p<0.05$). Region, the interaction of gender×region, and the three-way interaction of vowel×region×gender were just not significant. Restricting the ANOVAs to pairs out of the four groups clearly indicated that one group was different from the other three: SSD women, who on average produce longer durations. *Post-hoc*

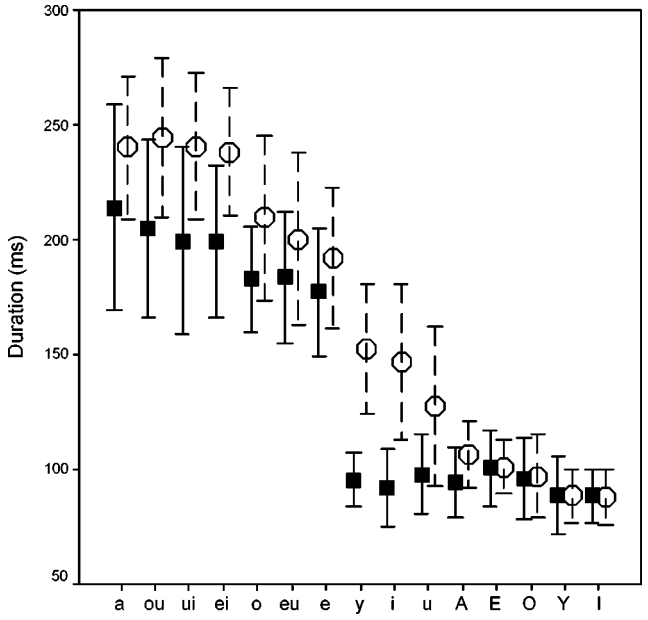


FIG. 2. Error plot of average duration (ms) per vowel, for the women for the two regions: NSD (filled squares) and SSD (open circles) (“ei”=/ɛi/, “ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

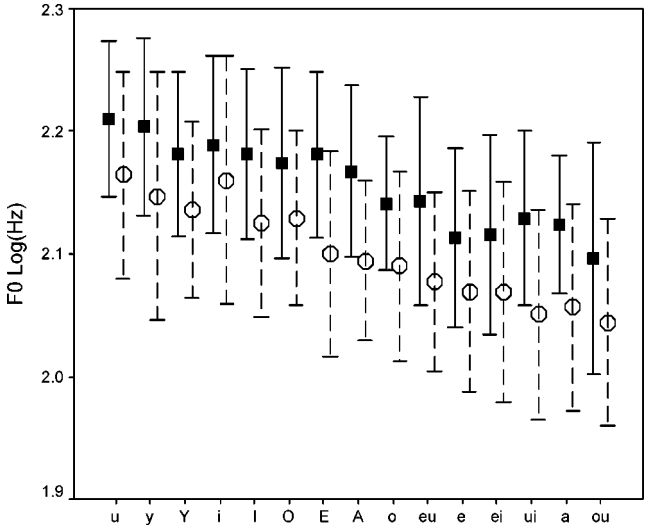


FIG. 3. Error plot of average $F0$ in Log(Hz) per vowel, for the men for the two regions: NSD (filled squares) and SSD (open circles) (“ei”=/ɛi/, “ou”=/ɔu/, “ui”=/œy/, “eu”=/ø/, “A”=/ɑ/, “O”=/ɔ/, “Y”=/ʏ/, “I”=/ɪ/). Bars represent one standard deviation.

analyses on the four groups for all 15 vowels reveal a clear pattern (Tukey, $\alpha=0.05$): SSD women have significantly longer durations compared to the other three groups for the diphthong “ɛi” and for the three high vowels /y ʏ u/.

The above results for the duration measurements can be summarized as follows. First, the 15 vowels fall into two groups based on their duration: longer ones /a ɛɪ ɔu œy e o ø/ and shorter ones /ɑ ɛ i ɪ ɔ u y ʏ/. Second, female SSD talkers produced longer durations than male talkers for several vowels, indicating a more general effect for the whole set of long vowels and the high vowels. Longer vowel durations for female talkers were reported earlier in Hillenbrand *et al.* (1995). We, like Hillenbrand *et al.*, cannot account for these gender-specific differences and we cannot predict whether a similar difference between male and female talkers in SSD would also be found in spontaneous speech.

Finally, it is not feasible to compare our results with those reported in Pols *et al.* (1973) and Van Nierop *et al.* (1973), because they did not include duration measurements in their description of the Dutch vowels.

2. Fundamental frequency

The average $F0$ values for the male talkers are displayed in Fig. 3, whereas Fig. 4 gives the values for the female talkers. First, a regional difference can be observed in Fig. 3: average $F0$ values for all vowels are lower for SSD, with differences ranging between 8 Hz (for /i/) and 25 Hz (for /ɛ/). No such pattern can be observed for the female talkers in Fig. 4. A repeated measures ANOVA was carried out for the male and female talkers separately to evaluate whether there was an effect of region on $F0$. In this ANOVA, vowel was composed of the 15 vowels and the regional background of the talker (region) served as a between-subject factor. The results show no significant effects for region. Further analysis showed that $F0$ and duration correlate: vowels with a low average $F0$ in Figs. 3 and 4 generally show longer durations in ms in Figs. 1 and 2. Pearson’s r was computed for $F0$ and

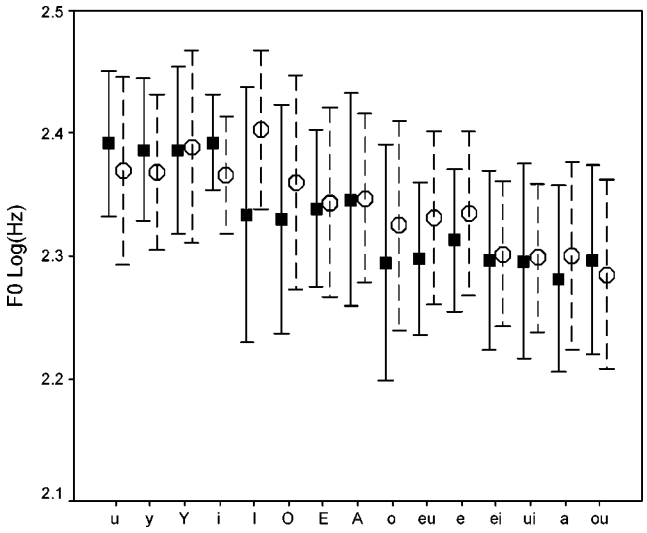


FIG. 4. Error plot of average $F0$ in Log(Hz) per vowel, for the women for the two regions: NSD (filled squares) and SSD (open circles) (''au''=/ɔu/, ''ui''=/œy/, ''eu''=/ø/, ''A''=/ɑ/, ''O''=/ɔ/, ''Y''=/ʏ/, ''I''=/ɪ/). Bars represent one standard deviation.

duration for the NSD women (−0.468), NSD men (−0.480), SSD women (−0.284), and SSD men (−0.331), across all vowel tokens. All correlations were significant at $p<0.05$. This phenomenon is also known as vowel-intrinsic $F0$, and it was reported to occur in many languages [cf. Whalen and Levitt (1995) for an overview]. Finally, it was again not possible to compare our results with those in Pols *et al.* (1973) and Van Nierop *et al.* (1973), because they did not include fundamental frequency measurements in their description.

3. Formant frequencies

a) Steady-state. Figures 5 (male talkers) and 6 (female talkers) show acoustic vowel diagrams for the nine monophthongal vowels at 50% ($t5$); the long mid vowels /e ø o/ and the three diphthongal vowels /ɛɪ ɔu œy/ were excluded, because we suspected that these six vowels show considerable diphthongization.

It is not easy to provide a concise description of the variation patterns in Figs. 5 and 6. Still, some general tendencies can be observed. Overall, Fig. 5 shows that the shape and size of the vowel systems for both language varieties for the male talkers are roughly similar, although /u/ for NSD shows a higher average $F2$. Large differences can also be seen for /ɪ/, /a/, /ɛ/, and /i/, for $F1$ as well as for $F2$. For the female talkers in Fig. 6, the overall shape and size of the vowel systems for NSD and SSD appear to differ little; the locations of the point vowels /i a u/ are roughly equal for both varieties. Larger differences can be observed for /ɪ/ and /y/; average $F2$ is higher and average $F1$ is lower for /ɪ/ for NSD, and average $F1$ and $F2$ are lower for NSD for /y/. Finally, Fig. 6 shows that /ɪ/ and /ɛ/ differ mainly in their average $F1$ and less in their average $F2$.

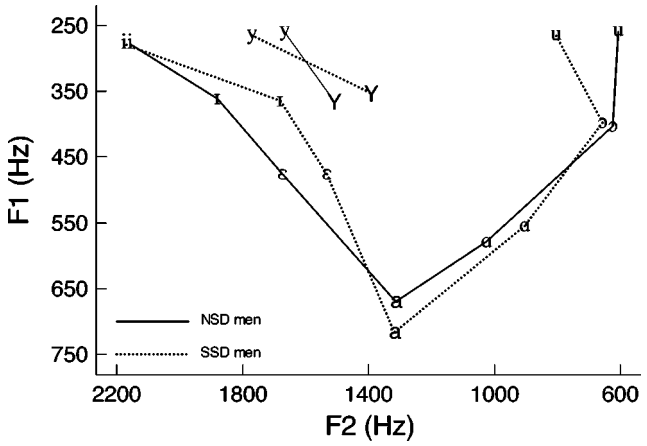


FIG. 5. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD and SSD for the men. The averages were taken at 50% of the vowel duration.

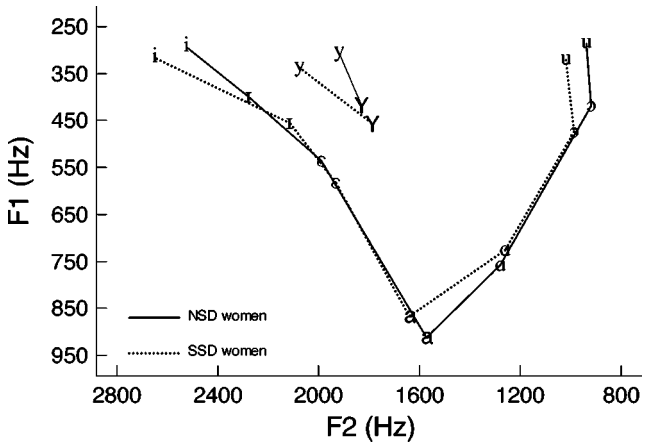


FIG. 6. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD and SSD for the women. The averages were taken at 50% of the steady-state portion of the vowel duration.

b) Spectral change. Figure 7 gives the average frequencies for $F1$ and $F2$ at 25% and 75% in the vowel duration ($t3$ and $t7$, respectively). Figure 7 presents spectral change for the three full diphthongs, for the male and female talkers, respectively. Figure 8 shows the spectral change for the three long mid vowels, for the male and female talkers.

To evaluate whether there are systematic differences in the spectral change patterns of the two regions and the two genders depicted in Figs. 7 and 8, five ANOVAS (repeated measures) were carried out, three for the diphthongal vowels /ɛɪ ɔu œy/ and two for the long mid vowels /e o ø/. The within-subject factor consisted of a measure for the spectral change in each vowel token, which was defined as the absolute difference of the formant frequency between 25% and 75% of the vowel duration in Hz. Thus $\Delta F1$ is the absolute difference in Hz between the values of $F1$ at $t3$ and at $t7$, and $\Delta F2$ is the absolute difference in Hz between the values of $F2$ at $t3$ and at $t7$. The first ANOVA for the diphthongal vowels tested possible effects of the between-subject factors region and gender and the within-subject factor vowel, on the change in the first formant frequency $\Delta F1$. The results showed an effect for region ($F[1,36]=7.15$, $p<0.05$) and an interaction effect for gender×region ($F[1,36]=4.42$, $p<0.05$). The significance of the interaction between gender and region indicates that the effect of region on $\Delta F1$ is

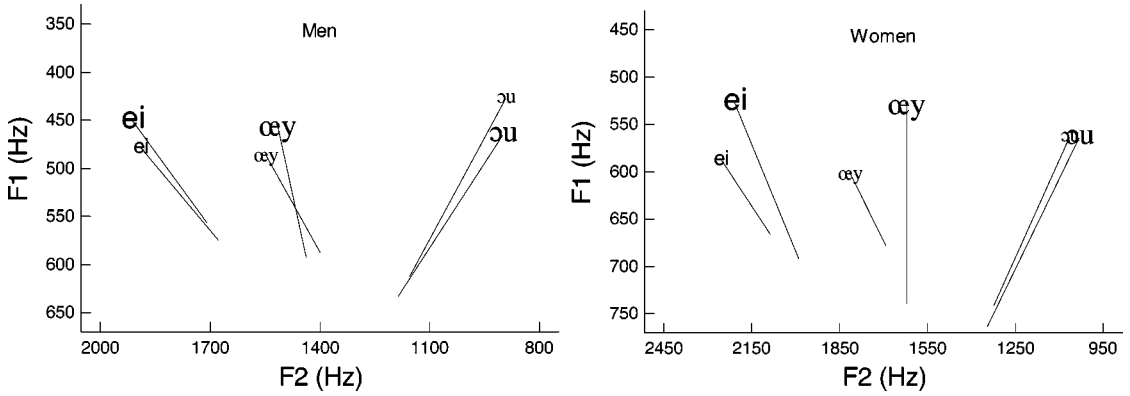


FIG. 7. Spectral change patterns for the three diphthongal vowels /ɛi œy ɔu/ for NSD and SSD for the men (left panel) and the women (right panel). The phonetic symbol at the end of each line is plotted at the average formant frequency at 75% of the vowel duration and each line originates from the average formant frequencies at 25% of the vowel duration. The larger symbols refer to NSD and the smaller symbols refer to SSD. ei refers to /ɛi/.

gender-specific. To investigate the interaction further, the last ANOVA was repeated for the male and female talkers separately, with region as the sole between-subject factor. The results showed an effect of region for the female talkers ($F[1,18]=11.32$, $p<0.05$) for $\Delta F1$, but not for the male talkers. In the next analysis, vowel represented measurements of $\Delta F2$ for the vowels /ɛɪ ɔu œy/, and region and gender as between-subject factors. The results showed no significant effects for $\Delta F2$. The final two analyses were set up as the first two, only this time the analyses were carried out on the vowels /e o ø/ for $\Delta F1$ and $\Delta F2$, respectively. The results for the ANOVA with $\Delta F1$ for /e o ø/ showed an effect for region ($F[1,36]=33.53$). An effect for region was also found for the ANOVA with $\Delta F1$ for /e o ø/ ($F[1,36]=18.51$, $p<0.05$).

In summary, the following spectral change patterns were found. First, for the diphthongal vowels /ɛɪ ɔu œy/, greater diphthongization of $F1$ was found for the female NSD talkers as opposed to the female SSD talkers, whereas no such region effect was found for the male talkers. Second, the long mid vowels /e o ø/ showed more diphthongization of $F1$ and $F2$ for NSD than for SSD, for both genders.

c) Comparison with Pols *et al.* (1973) *and* Van Nierop *et al.* (1973). Figure 9 gives the acoustic vowel diagram for the nine monophthongal vowels at 50% ($t5$) for our male talkers plus the Pols *et al.*'s male talkers. Figure 10 gives the acoustic vowel diagram for the nine monophthongal vowels at 50% ($t5$) for our female talkers plus Van Nierop *et al.*'s female talkers. A first general observation from Fig. 9 is that all average formant values for Pols *et al.*'s data are higher for $F1$. The vowels /a ɑ ɔ/ show lower average values for $F2$ for Pols *et al.*'s data. Figure 10 shows a different variation pattern in the vowel systems for the women than the one observed for the men in Fig. 9. Higher average values for $F1$ for /a ɛ ɪ ɔ ʏ/ and lower values for $F2$ for the vowels /a ɑ ɛ ɔ u y ʏ/ can be observed in Fig. 10. Overall, the vowel diagrams for our data appear smaller than Pols *et al.*'s and Van Nierop *et al.*'s, possibly indicating a more ''casual'' speaking style in our data. It is unclear what caused the other differences between the two data sets. One possible cause may be that the pronunciation of the Dutch vowels has changed in the three decades since Pols *et al.* and Van Nierop *et al.* made their recordings. In our opinion, however, it seems more likely that the observed differences are due to either differences in the consonantal context of the vowels (/hVt/ vs. /sVs/), or to uncontrolled regional variation in Pols *et al.*'s and Van Nierop *et al.*'s data, or to differences in the techniques used to estimate the formant frequencies. Given these variation sources, we decided not to further analyze differences between our data and data described in Pols *et al.* and Van Nierop *et al.*

IV. DISCRIMINANT ANALYSIS

We carried out a series of quadratic discriminant analyses (QDAs) to establish how well the vowels could be separated based on various combinations of acoustic characteristics. We evaluated steady-state characteristics, i.e., $F0$ and

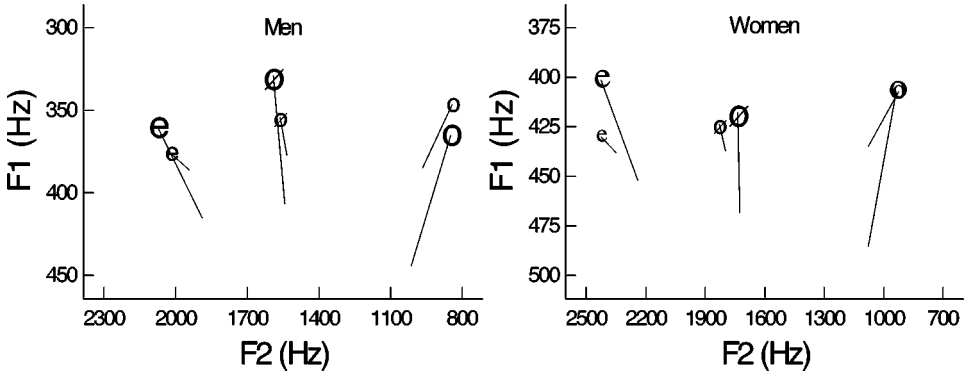


FIG. 8. Spectral change patterns for the three long mid vowels, /e ø o/, for NSD and SSD for the men (left panel) and the women (right panel). The phonetic symbol at the end of each line is plotted at the average formant frequency at 75% of the vowel duration and each line originates from the average formant frequencies at 25% of the vowel duration. The larger symbols refer to NSD and the smaller symbols refer to SSD.

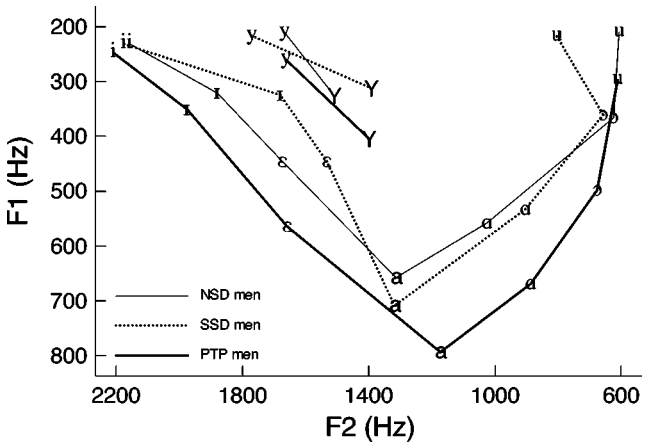


FIG. 9. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD, SSD, and from Pols, Tromp, and Plomp (1973), "PTP," for the men. All averages were taken at 50% of the vowel duration. The averages for NSD and SSD were based on 20 talkers and the averages from Pols *et al.* were based on 50 talkers.

$F1-F3$ sampled at 50% of the vowel duration ($t5$), as well as dynamic specifications, i.e., the duration and $F1-F3$ sampled at 25% and 75% of the duration ($t3$ and $t7$). Table III shows that including only $F1$ and $F2$ at $t5$ yields 59.0% for NSD and 54.2% for SSD. Including $F0$ in the set led to a small improvement for NSD and a considerable improvement for SSD. Entering the parameter set consisting of $F1$, $F2$, and $F3$ led to improvements for both language varieties as compared to entering only $F1$ and $F2$. Entering $F0$, $F1$, $F2$, and $F3$ led to another improvement in the scores. To find out whether the results would improve when the three mid vowels and the three full diphthongal vowels are excluded, the QDA for the parameters $F1$ and $F2$ was repeated using only the nine monophthongal vowels. The results show a remarkable improvement over those carried out using all 15 vowels, for both language varieties. However, the improvement is considerably higher for NSD. Adding $F3$ led to a further improvement, especially for SSD. The next QDA was

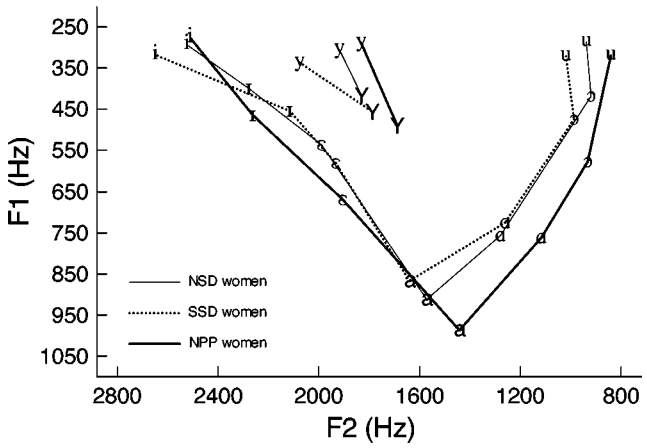


FIG. 10. Acoustic vowel diagram showing average formant frequencies for the nine monophthongal vowels /a ɑ ɛ i ɪ ɔ u y ʏ/ for NSD, SSD, and from Van Nierop, Pols, and Plomp (1973), "NPP," for the women. All averages were taken at 50% of the vowel duration. The averages for NSD and SSD were based on 20 talkers and the averages from Van Nierop *et al.* were based on 25 talkers.

TABLE III. Percent correctly classified vowel tokens for the quadratic discriminant analyses for the NSD vowel tokens ($N=600$) and the SSD vowel tokens ($N=600$). The data were classified into 1 of the 15 vowel categories, or into 1 of 9 monophthongal vowel categories /a ɑ ɛ i ɪ ɔ u y ʏ/ ("9 vowels"). The parameters were sampled at one point in the vowel duration: "$t5$" (50% of the vowel duration), or at two points "$t3+t7$" (25% and 75%). For NSD, all percentages higher than 62.9% are significantly higher than the baseline (59.0%), for SSD all percentages higher than 58.2% are significantly higher than the baseline (54.2%).

Parameters	Categories	Sample	NSD	SSD
$F1$, $F2$	15 vowels	$t5$	59.0	54.2
$F0$, $F1$, $F2$	15 vowels	$t5$	65.5	67.5
$F1$, $F2$, $F3$	15 vowels	$t5$	68.0	66.3
$F0$, $F1$, $F2$, $F3$	15 vowels	$t5$	71.5	71.5
$F1$, $F2$, $F3$, dur	15 vowels	$t3+t7$	90.8	90.0
$F1$, $F2$	9 vowels	$t5$	83.6	76.4
$F1$, $F2$, $F3$	9 vowels	$t5$	86.1	83.3
$F1$, $F2$, $F3$, dur	15 vowels	$t5$	90.0	85.3
$F0$, $F1$, $F2$, $F3$, dur	15 vowels	$t5$	89.3	88.8

carried out using $F1$, $F2$, and $F3$ and duration as predictor variables, using all 15 vowels at $t5$, to evaluate the role of duration. For this parameter set, a considerable improvement was found as compared to including only $F1$, $F2$, and $F3$ as parameters. Furthermore, when $F0$ is added as well as $F1$, $F2$, $F3$, and duration, some improvement is again found, but this time only for SSD (85.3% to 88.8%); for NSD the percentage is lower (90.0% to 89.3%). It is unclear why the percentage for NSD is lower, but since it only is a small difference (0.7%) not much weight should be attributed to it. The improvement for SSD is attributable to including $F0$. It appears that $F0$ plays a greater role in distinguishing vowels in SSD than in NSD, whenever $F0$ is added to a parameter set, the improvement is largest for SSD. Duration, on the other hand, seems to play a greater role in distinguishing vowels for NSD; improvements for NSD are largest for NSD whenever duration is added to a parameter set. Finally, when the last analysis is repeated with all parameters at $t3$ and $t7$, the highest percents correctly classified are obtained for both language varieties. Note that the relative improvements are highest for SSD, which is remarkable given the finding that SSD shows less diphthongization of $F1$ and $F2$ for the long mid vowels and the diphthongal vowels.

The results in Table III can be interpreted as follows: first, it appears that the nine monophthongal vowels could be separated reasonably well using only steady-state $F1$ and $F2$, although a higher percent correctly classified vowel tokens was obtained for NSD. Second, the three long mid vowels and the three full diphthongal vowels could be separated fairly well when dynamic properties such as duration and dynamic spectral information were included in the analysis, although again the percent correctly classified vowel tokens was highest for NSD. It could thus be the case that the vowel system for NSD requires more information about dynamic properties (especially for the three long mid vowels and the three diphthongs) in order to be separated acoustically.

V. DISCUSSION

The purpose of this paper was twofold: to give a description of the acoustic characteristics of all 15 Dutch vowel

sounds and to provide an overview of similarities and dissimilarities of the vowel systems of Northern Standard Dutch and Southern Standard Dutch. The results for the first aim were as follows.

First, the 15 vowels of Dutch could be divided into two groups depending on their duration. The group of shorter vowels consists of all full monophthongal vowels /y i u ɑ ɛ ɔ ʏ ɪ/ except for /a/, while the group of long vowels consists of the three diphthongal vowels /ɔu œy ɛɪ/, the three long mid vowels /o ø e/, and /a/. This phonetic pattern in the relative duration is only partially compatible with the phonological characteristics of Dutch vowels as described in Booij (1995), where the set of long vowels includes, besides /o ø e a/, also /i y u/. Our results also deviate slightly from the description provided by Koopmans-van Beinum (1980) and Rietveld *et al.* (in press). Koopmans-van Beinum and Rietveld *et al.* suggest that Dutch vowels can be divided into three groups: short, /ɑ ɛ ɪ ɔ ʏ/, half-long, /i y u/, and long, /a e ø o/ plus /ɛɪ ɔu œy/. If we pool the short and half-long vowels, the system by Koopmans-van Beinum and Rietveld *et al.* is compatible with our findings. This applies to three of the four groups of talkers. The results for the female SSD talkers conform the pattern found by Koopmans-Van Beinum and Rietveld *et al.* Nevertheless, it remains uncertain whether the pattern we reported for the relative average durations would also occur for data in other consonantal contexts and in spontaneous speech. Further research is required to investigate which patterns would occur in other speech samples.

Second, the nine monophthongal vowels of Dutch, /ɑ a ɛ ɪ i ɔ u ʏ y/, could be separated fairly well based on their steady-state characteristics for their first two formants frequencies alone, while the three long mid vowels and three diphthongal vowels required information about their dynamic characteristics as well. The long mid vowels cannot be described adequately acoustically by their steady-state characteristics alone. We suggest that the three long mid vowels for Dutch should therefore not be treated as monophthongal vowels, but instead as semi-diphthongal vowels, when describing Dutch vowels acoustically. This is especially the case for NSD. The diphthongization of /e ø o/ is most likely the result of a vowel shift. It is not clear whether the long mid vowels showed (early) signs of the process of diphthongization in Pols *et al.*'s and Van Nierop *et al.*'s descriptions three decades earlier. Pols *et al.* excluded the full diphthongs /ɛɪ ɔu œy/, but included the three long mid vowels. However, in their Fig. 3, considerable overlap can be observed in the vowel plot between /e/ and /ɪ/, and between /ø/ and /ʏ/ (denoted in Pols *et al.* by ''œ''), and /ɔ/ and /ø/. Furthermore, Pols *et al.* remark, ''...the vowels in these three pairs have very similar formant frequencies and levels. The main difference between them is duration.'' Nevertheless, it is not possible to establish whether /e ø o/ can be distinguished from /ɪ ʏ ɔ/, respectively, through their dynamic characteristics as well as by their duration, because Pols *et al.* and Van Nierop *et al.* do not present the formant frequencies at different time slices.

Regarding the second aim of this study, the difference between the two language varieties, the results were as follows. A first difference between the NSD and SSD was found for the duration of the three full diphthongs /ɛɪ ɔu œy/; the durations for the SSD talkers were longer for all three diphthongs, especially for the female talkers. The comparisons indicated further that NSD and SSD differ little in the steady-state characteristics of the nine monophthongal vowels; the main difference between both varieties was found for the three long mid vowels and the three full diphthongs. For the three diphthongs, more diphthongization (i.e., larger excursions) of the first formant was found for the female NSD talkers than for the female SSD talkers. No such effect was found for the male talkers. Overall, vowels produced by all NSD talkers show more diphthongization than vowels produced by SSD talkers. The differences between NSD and SSD were largest for the three long mid vowels. The results for the duration and dynamic spectral characteristic may be related: the longer durations for the diphthongal vowels for SSD may be used to compensate for the smaller amount of diphthongization for the SSD diphthongal vowels. Further research on other speech samples is necessary to evaluate the hypothesis that NSD and SSD use different quantity distinctions.

A comparison of our data with the data described in Pols *et al.* and Van Nierop *et al.* showed differences in the average frequencies of the nine monophthongal vowels. Most vowels show higher values for $F1$ for Pols *et al.*'s and Van Nierop *et al.*'s data. In addition, Pols *et al.* and Van Nierop *et al.* show lower values for $F2$ for some vowels (especially for /a ɑ ɔ/). We suggested three possible causes for the observed differences: a change in the pronunciation of the vowels, differences between the two data sets, such as the surrounding consonants (/sVs/ vs. /hVt/), or uncontrolled regional variation in Pols *et al.*'s and Van Nierop *et al.*'s data.

The present study is limited in that it describes the acoustic characteristics of the vowels of Dutch in read speech and in a fixed consonantal context only. It would be interesting to compare the acoustic characteristics of the vowels of the present study with other vowel tokens produced by the same talkers in different consonantal context and speaking styles. During the sociolinguistic interview in which our vowel data was obtained, recordings were also made of (guided) spontaneous conversations. We plan to extend our study in the near future and analyze the acoustic characteristics of vowel in these spontaneous speech samples as well. Finally, it was not feasible to provide a complete overview of all the variation patterns in all eight regional varieties in our data set in the present study. However, the individual measurements for all 160 speakers can be obtained by contacting the first or second author.

ACKNOWLEDGMENTS

We are grateful to two anonymous reviewers for their useful comments. We wish to thank James Hillenbrand for his help with Figs. 5–10. We are greatly indebted to Terry Nearey for writing the program and user interface that allowed us to perform our formant measurements. This research was supported by the Netherlands Organization for

Research (NWO) through the Flemish Netherlands Committee (VNC) under Project No. 205-41-069 (PI Roeland van Hout).

[1]Pols, Trom and Plomp (1973) used the vowel set of the 50 male talkers described in Klein, Plomp, and Pols (1970).

[2]Klein, Plomp, and Pols' (1970) talkers are described as ''The 50 speakers were young male adults with a pronunciation representative of correct Dutch.''

[3]No recordings were made for the so-called ''marginal'' vowels of Dutch, i.e., the vowels /i: y: ɛ: œ:/ and the nasalized versions of the vowels /ɛ: ɑ: ɔ:/, which all occur only in loan words in Dutch (Gussenhoven, 1999).

[4]It was not feasible to include measurements of *F*4, because these could not be obtained in the program used for the formant estimation.

Adank, P. (**2003**). ''Vowel Normalization: A perceptual-acoustic study of Dutch vowels,'' Ph.D. thesis, University of Nijmegen.

Bell, A. (**1983**). ''Broadcast news as language standard,'' Int. J. Soc. Lang. **40**, 29–42.

Boersma, P. (**1993**). ''Accurate short-term analysis of fundamental frequency and the harmonics-to-noise ratio of a sampled sound,'' in *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, Vol. 17, pp. 97–110.

Boersma, P. (**2001**). ''Praat, a system for doing phonetics by computer,'' Glot Int. **5(9/10)**, 341–345.

Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon, Oxford).

Deterding, D. (**1997**). ''The formants of monophthong vowels in Standard Southern British English Pronunciation,'' J. Int. Phon. Assoc. **27**, 47–55.

Di Benedetto, M.-G. (**1989a**). ''Vowel representation: Some observations on temporal and spectral properties of the first formant frequency,'' J. Acoust. Soc. Am. **86**, 55–66.

Di Benedetto, M.-G. (**1989b**). ''Frequency and time variations of the first formant: Properties relevant to the perceptions of vowel height,'' J. Acoust. Soc. Am. **86**, 67–77.

Disner, S. (**1980**). ''Evaluation of vowel normalization procedures,'' J. Acoust. Soc. Am. **67**, 253–261.

Gussenhoven, C. (**1999**). ''Illustrations of the IPA: Dutch,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge), pp. 74–77.

Hagiwara, R. (**1997**). ''Dialect variation and formant frequency: The American English vowels revisited,'' J. Acoust. Soc. Am. **102**, 655–658.

Hillenbrand, J., and Gayvert, R. T. (**1993**). ''Identification of steady-state vowels synthesized from the Peterson and Barney measurements,'' J. Acoust. Soc. Am. **94**, 668–674.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic analysis of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). ''Vowel spectra, vowel spaces and vowel identification,'' J. Acoust. Soc. Am. **48**, 999–1009.

Koopmans-van Beinum, F. (**1980**). ''Vowel contrast reduction,'' Ph.D. thesis, University of Amsterdam.

Markel, J. D., and Gray A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in speech perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M., Assmann, P., and Hillenbrand, J. (**2002**). ''Evaluation of a strategy for automatic formant tracking,'' J. Acoust. Soc. Am. **112**, 2323.

Payne, A. (**1980**). ''Factors controlling the acquisition of the Philadelphia dialect by out-of-state children,'' in *Locating Language in Time and Space*, edited by W. Labov (Academic, New York).

Pols, L. C. W., Pols, L. C. W., and Plomp, H. R. C. (**1973**). ''Frequency analysis of Dutch vowels from 50 male speakers,'' J. Acoust. Soc. Am. **53**, 1093–1101.

Rietveld, T., Kerkhoff, J., and Gussenhoven, C. (**in press**). ''Word prosodic structure and vowel duration in Dutch,'' to appear in J. Phonetics.

Scovel, T. (**1988**). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech* (Newbury House, Rowley).

Strange, W. (**1989**). ''Dynamic specification of coarticulated vowels spoken in sentence context,'' J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., Jenkins, J. J., and Johnson, T. L. (**1983**). ''Dynamic specification of coarticulated vowels,'' J. Acoust. Soc. Am. **74**, 695–705.

Syrdal, A. K., and Gopal, H. S. (**1986**). ''A perceptual model of vowel recognition based on the auditory representation of American English vowels,'' J. Acoust. Soc. Am. **79**, 1086–1100.

Van de Velde, H., van Hout, R., and Gerritsen, M. (**1997**). ''Watching Dutch Change,'' J. Sociolinguistics **1**, 361–391.

Van Nierop, D. J. P. J., Pols, L. W. C., and Plomp, R. (**1973**). ''Frequency analysis of Dutch vowels from 25 female speakers,'' Acustica **29**, 110–118.

Whalen, D. H., and Levitt, A. G. (**1995**). ''The universality of intrinsic F0 of vowels,'' J. Phonetics **23**, 349–366.

The acoustic analysis of tone differentiation as a means for assessing tone production in speakers of Cantonese

Johanna G. Barry[a)]
Department of Otolaryngology and the Department of Linguistics and Applied Linguistics, University of Melbourne, Parkville, 3010, Australia

Peter J. Blamey
Department of Otolaryngology, University of Melbourne, Parkville, 3010, Australia

(Received 7 October 2003; revised 31 May 2004; accepted 14 June 2004)

This paper reports on a methodology for acoustically analyzing tone production in Cantonese. F0 offset versus F0 onset are plotted for a series of tokens for each of the six tones in the language. These are grouped according to tone type into a set of six ellipses. Qualitative visual observations regarding the degree of differentiation of the ellipses within the tonal space are summarized numerically using two indices, referred to here as Index 1 and Index 2. Index 1 is a ratio of the area of the speaker's tonal space and the average of the areas of the ellipses of the three target tones making up the tonal space. Index 2 is a ratio of the average distance between all six tonal ellipses and the average of the sum of the two axes for all six tone ellipses. Using this methodology, tonal differentiation is compared for three groups of speakers; normally hearing adults; normally hearing children aged from 4–6 years; and, prelinguistically deafened cochlear implant users aged from 4–11 years. A potential conundrum regarding how tone production abilities can outstrip tone perception abilities is explained using the data from the acoustic analyses. It is suggested that young children of the age range tested are still learning to normalize for pitch level differences in tone production. Acoustic analysis of the data thus supports results from tone perception studies and suggests that the methodology is suitable for use in studies investigating tone production in both clinical and research contexts. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779272]

PACS numbers: 43.70.Jt, 43.70.Ep, 43.66.Ts [AL] Pages: 1739–1748

I. INTRODUCTION

Studies of the acquisition of lexical tone by young children have used auditory transcription analysis of tonal productions, either in free speech samples (e.g., Tse, 1978; Clumeck, 1980; Tse, 1991) or in words elicited as tokens in citation form using picture- or object-naming tasks (e.g., Li and Thompson, 1977; So and Dodd, 1995). No matter how the speech samples are elicited, these studies agree that the first tones emerge very early in the course of speech acquisition, at about the same time as the first recognizable vowels are observed. Two tones are acquired at once, probably reflecting the fact that tone is a relative not an absolute phenomenon. Finally, children are observed to achieve mastery of the full toneme inventory long before they master consonant phoneme inventories.

Auditory analysis requires the segmentation and appraisal of utterances produced by a young child with direct reference to phonologically significant entities in the mature linguistic system. Given the fact that linguistic tone develops through exploitation of pitch—a phonetic consequence of voiced speech—and that it is a relative not an absolute entity, it is interesting to note that in all materials investigating tone acquisition in young infants, there has been little discussion about the features in an infant's phonetic voicing patterns which changed to result in an utterance becoming recognizably tonal.

Rose (1989) noted that tone is a product of an interaction between F0 and other acoustic parameters in speech. It is possible that part of pitch patterning in a young infant's speech becoming amenable to tone labeling is a reflection of the development of increased control of prosodic features such as syllable duration which is coupled with the acquisition of some readily identifiable vowel phones.

A related issue concerns the assignment of an F0 contour to a tone category. This issue has been ignored in tone acquisition studies possibly because the rapid rate and accuracy of acquisition renders it irrelevant or too difficult to pursue. Yet as Kent and Murray (1982) note such phonological style analyses of a young infant's articulations necessarily assume that it is appropriate to categorize the vocalizations of an infant into segments with a phonemically relevant referent where in fact there is no guarantee that these actually do correspond to real segments in the infant's developing linguistic system. They observe that the emergence of early phonemic contrasts results, in part, from the fact that certain sound patterns occur more frequently in speech than others. Thus a vowel heard in a given utterance may reflect an accidental co-occurrence of the correct position of the articulators with a vowel function at that point in time and may not be an indication that the infant has acquired that phone in his/her developing phonemic inventory (Keating, 1980). On this basis, Kent and Murray argue the need for acoustic in-

[a)]Current address: Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3RQ United Kingdom; Electronic mail: johanna.barry@psy.ox.ac.uk

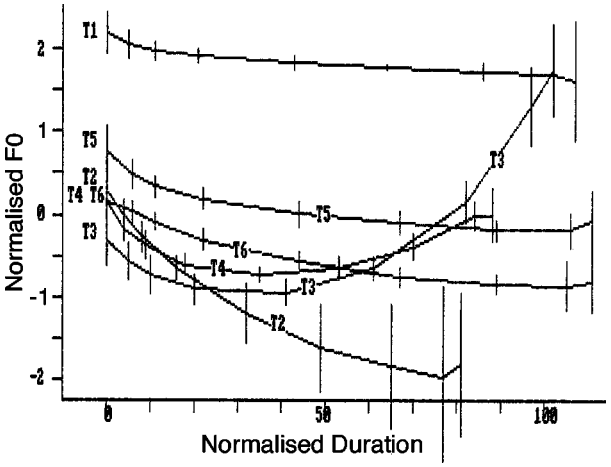


FIG. 1. The six tones in Cantonese normalized for duration and pitch across 10 speakers (five male and five female) (Rose 2000, p. 202). Vertical bars show one standard deviation above and below the mean. T1 is a high level tone (55); T2, a low-falling tone (21); T3 is a high-rise tone (25); T4, a low-rise tone (23); T5 a mid-level tone (33); and, T6 a low-level tone (22) (reproduced with permission from ASSTA and Rose).

vestigations into speech development to verify the perceptual biases of the observer.

There are six tonemes in Cantonese which are labeled; Tone 55, a high-level tone; Tone 25, a high-rise; Tone 33, a mid-level tone; Tone 21, a low-fall tone; Tone 23, a low-rise tone; and Tone 22, a low-level tone. As illustrated in Fig. 1, there are considerable perceptual similarities between some of the tonal contrasts in Cantonese and with the exception of Tone 55, Cantonese tones are primarily distinguished by differences in F0 offset.

Little is known about tone acquisition or indeed tone production in profoundly deaf speakers and still less is known about the effect of using a cochlear implant on the development of tone. Available reports, however, suggest that despite their impaired access to auditory input and the perceptual similarity of the some of the tones in Cantonese, profoundly hearing-impaired Cantonese speakers are able to successfully acquire complete or near complete toneme inventories (Dodd and So, 1994). Dodd and So hypothesize that the children in their study acquired tone successfully because (a) tone has a high functional load in Cantonese; (b) tone production is emphasized during speech habilitation; and (c) tones occur in the low frequency end of the speech spectrum where profoundly hearing-impaired speakers have most residual hearing.

The research reported here describes the development of a methodology for the investigation of tone production by Cantonese speakers using acoustic measurements. The aim of the research was to supplement findings based on more traditional auditory transcription analyses.

From a clinical perspective, an acoustic analysis of the data is attractive because it provides a means for tracking and quantifying changes in development of tone production in response to habilitation or maturational effects. It also avoids issues associated with the use of purely phonological analyses for describing developing linguistic systems. Finally, in the case of Cantonese where some tones are perceptually similar, it provides supplementary support for conclusions based on auditory analyses of the data.

A. Analyzing tone production acoustically

To study tone production and tone development acoustically, a methodology is required which (a) is sufficiently robust to permit the observation and comparison of changes in F0 usage in children with widely varying speech production skills; (b) is readily subjected to a range of statistical analyses; and (c) can be compared with results from auditory analyses of the same sets of data.

In one of the few acoustic studies of tone/intonation production, Kent and Murray (1982) investigated intonation patterns in infants acquiring English. F0 contours were given descriptive labels such as flat, or rising-falling and the frequency with which these patterns were observed in the infant's vocalizations was compared at 3, 6, and 9 months.

Among tonal languages, Gandour *et al.* (1988) analyzed tone productions by one electrolaryngeal and two esophageal speakers of Thai, both acoustically and perceptually, and compared the data with that recorded for five normal speakers of Thai. Like the Kent and Murray study of intonation patterns in infants, the acoustic analyses were highly descriptive in form and based on comparative descriptions of the contours produced by the different speakers.

Wong and Diehl (1999) described the investigation of tone production in speakers affected with Parkinson's disease. They provided little detail about their methodology, but reported that these speakers had a restricted tonal space which made correct labeling of their tones more difficult for normal Cantonese speakers. The concept of tonal space will be referred to again in the context of the present study. It is defined as ''...the set of articulatory and auditory dimensions by which the speaker is constrained in production and perception'' (Abramson, 1986, p. 105). It is interesting to note that Abramson, like Wong and Diehl, also noted an important role for tonal space in the correct perception of tones.

Rose (1987, 1993) developed a more quantitative approach to the acoustic analysis of tone production. In this approach, values of F0 are measured at fixed points along a tone contour. They are expressed as a multiple dispersion from a mean value of F0 and plotted against similarly normalized values of durational differences (Fig. 1). The aim in this analysis is to abstract away from between-speaker differences, while retaining the linguistically significant features in a language's tonal phonology. A more precise description of the linguistic–tonetic form of the tonemes in a language's toneme inventory can thus be achieved. For this analysis, it is important that normalization procedures are performed on transcriptionally equivalent words only, i.e., on units which constitute the same tonal targets for all speakers. If not, there is a risk that linguistically significant differences between speakers will be normalized out, leading to spurious conclusions.

Rose's approach might be suitable for comparing the tonal productions of post-linguistically deafened cochlear implant users with those of normally hearing speakers. However, it cannot be used with prelinguistically deafened cochlear implant users because not enough is known of their

tone perception or tone production abilities to be categorically sure that they (a) would produce the same variety of linguistic tone as their hearing-impaired peers or (b) would produce tonal patterns in response to the same tonal targets as their normally hearing peers.

B. An alternative approach to the acoustic analysis of tone production

Gandour (1978) has shown that five dimensions defining acoustically measurable differences between tones account for a listener's perceptual judgments about tone. These are labeled interpretatively as (a) average pitch, (b) direction, (c) length, (d) extreme endpoint, and (e) slope. A plot of F0 offset versus F0 onset captures all dimensions except length, for each tone type. In such a plot, level tones would be expected to fall midway between the two axes at different points corresponding to different values of average F0. Rising tones would be expected to cluster closer to the *y* axis, while falling tones would cluster closer to the *x* axis.

Ellipses around each tone type can be calculated by determining the distribution of points around a mean for each tone. This is done by first calculating the directions for the major and minor axes for each ellipse using a principal components analysis of the points for the corresponding tone. The data points are projected on to each axis and the standard deviations of the projections calculated. The ellipse radii are then set equal to twice the standard deviation to encompass approximately 95% of the projections on to each axis. Tonal ellipses calculated in this way visually summarize the locations in the plot where each of the six tonemes clusters. These plots are based on categories of perceptual relevance to the listener. They are predicted to be well differentiated where a speaker uses consistent patterns of pitch for each tone. Furthermore, the degree of differentiation should correlate with results based on auditory analyses of the same data. The higher the number of correct tones produced by a speaker, the greater the degree of differentiation of the ellipses in the tone plot.

C. Indexing degrees of differentiation between tones

Apart from describing tone differentiation with respect to visual comparison of the spread and degree of overlap between tonal ellipses, two indices were also developed which provide a numerical measure of the degree of tonal differentiation demonstrated by a speaker.

1. Index 1—Measuring tone differentiation within the tonal space

By acoustic measurement, the tones 55, 25, and 21 are the most differentiated of the six tones in Cantonese. Lines joining the center points of these three tonal ellipses form a triangle. This triangle effectively spans the mean range of F0 used by a speaker and defines the extent of the speaker's tonal space. The area of each ellipse describes the degree of variation of F0 used for each target tone. Tonal differentiation across the tonal space is thus a function of the area of the tonal space (*At*) and the spread of F0 usage for the three tones making up that area (i.e., the average area $Ae_{1,2,4}$, of the ellipses for tones 55, 25, and 21, where 1 = Tone 55, 2 = Tone 25, and 4 = Tone 21),

$$\text{Index } 1 = \frac{At}{Ae_{1,2,4}}. \tag{1}$$

Where $At/Ae_{1,2,4} > 2$, the likelihood of overlap between these ellipses is low. For ratios $At/Ae_{1,2,4} \leqslant 2$, tone ellipse overlap is likely between at least two of the ellipses; the lower the ratio, the greater the degree of overlap between the three ellipses.

2. Index 2—Measuring differentiation among tonemes

Like the ellipse area, the lengths of the two axes around which the ellipse is drawn describe the degree of variation in pitch used for each tone. The distances between the six tone ellipse centers determine the degree of difference in the average values of F0 for each tone. Differentiation within the tonal space can thus be expressed as the ratio of the average of the lengths of the two axes for the six tones (Ave Ax_{1+2}) and the average distance of the centers of the six tone ellipses from each other (Ave Dist.).

$$\text{Index } 2 = \frac{\text{Ave Dist.}}{\text{Ave Ax}_{1+2}}. \tag{2}$$

Index 2 has several advantages over Index 1. It relies on measures for all tones rather than three. It is sensitive to differences in pitch height and contour individually. For example, in the case where a child controls pitch height but not pitch contour, tones 55, 25, and 21 would all lie on the diagonal of the plot and the area of the triangle defining the tonal area would be zero. Since Ave Ax_{1+2} is determined from calculated standard deviation units around the mean of the distribution of points in the F0 onset/offset space, Index 2 is analogous to the signal detection measure d' (*d* prime).

D. Summarizing the approach to acoustic analysis of tone

The focus of our approach to the acoustic study of tone production involves observing the degree of differentiation between tone types produced by an individual speaker. The greater the differentiation among the ellipses, the more likely it is that a speaker has a consistent tone target, i.e., has acquired a toneme. Because this approach is at the level of the individual, there is no requirement for a speaker to have the same tonal target as any other speaker, making it suitable for use with young prelinguistically deafened implant users. In this paper, we shall illustrate how this approach can be used to describe differences in tone production in three groups of speakers: (a) normally hearing adults; (b) young normally hearing children aged between 4–6 years who are reported to have mastered tonal production at this age; and, (c) young prelinguistically deafened implant users.

II. METHOD

A. Subjects

This study included five normally hearing adults: (two male, three female); eight normally hearing children (three male, six female); and, 16 prelinguistically profoundly hearing-impaired implant users (seven male, nine female). All participants were native speakers of Cantonese. The age range for each group of speakers was 23–40 years (adults); 3 years 8 months to 6 years (normally hearing children); and 4 years 2 months to 11 years 3 months (implant users).

All implant users received their implant at the Prince of Wales Hospital (PWH) in Shatin, Hong Kong. They were fitted with either a Nucleus CI-22M (six subjects) or CI-24M (10 subjects) electrode manufactured by Cochlear Limited. At the time of testing, seven children were using the SPEAK speech processing strategy in a SPECTRA™ speech processor (McDermott *et al.*, 1992). The remaining nine children were using the ACE speech processing strategy in a SPRINT™ speech processor (Vandali *et al.*, 2000). The pulse rate for the SPEAK users was 250 Hz. The pulse rate used by the ACE users varied according to mapping. Seven children used a pulse rate of 900 Hz, one child used a pulse rate of 720 Hz and one child used a pulse rate of 1200 Hz. The pure tone average (i.e., average hearing loss in dB HL at 500, 1000, and 2000 Hz) for the group as measured prior to implant in the better ear was 101dB HL (s.d. 11 dB HL). The age at implant ranged from 2 years 3 months to 6 years 9 months. The number of months experience using the implant ranged from 15 months to 73 months.

All implant users received habilitation within an oral–aural scheme. Subsequent to receiving the implant, the children underwent a 2-year weekly program of habilitation at PWH. At the time of testing, 10 children attended schools dedicated to the education of the deaf. Three children attended an integrated kindergarten and were withdrawn everyday for extra habilitation. The remaining three children attended normal primary schools and received extra habilitation from the local audiological service.

B. Recording procedures

1. Materials

The words used in the test were collected as part of a battery of tests investigating tone perception and tone production skills in Cantonese-speaking children. Testing was carried out in a soundproof room at PWH. For the speech production task, the children were required to produce tones in citation form. These were elicited by means of a picture-naming task involving 15 presentations of each tone type on a range of syllables. There were 90 items in total. The words in the test were chosen because they had been shown to be frequently used by Cantonese-speaking children aged three and older (Lee, 2003). They were presented to the children as black and white line drawings arranged two to a page. Where children could not spontaneously name a picture, the principal investigator would try to elicit the word by asking a question, e.g., ''Is this a boy? No it's a ....'' Some items in the test involved tones presented on closed syllables ending in [−p, −t, or −k]. These tones are referred to as clipped or stopped. They are characterized by a shorter duration and were excluded from further acoustic analysis because of differences in form.

2. Recording and pitch extraction

Data from the children were first recorded on to audiotape using either a Portable Cassette Recorder Sony TM5000EV with an internal microphone or an AIWA Millennium Collection JXM2000 Walkman recorder and microphone. When using the latter, the microphone was fitted to the child's collar during recording. Speech samples were then digitized using ''Speech Analyzer'' (1999) a freeware software for the Microsoft Windows™ operating system developed as part of the acoustic speech analysis project at the Summer Institute of Linguistics. The speech for the normally hearing adults was recorded directly into ''Speech Analyzer'' using an electret condenser 600 ohms microphone. A 16-bit monochannel, 22.05 kHz sampling rate was used for digitizing all speech samples. Utterances were then phonetically labeled by auditory analysis to segment the initial consonant from the following rhyme. Phonetic labels and pitch values (sampled at 1 ms intervals) were exported using Speech Analyzer's ''Timetable'' format and downloaded into a program called ''Pitch.'' This program extracted values of F0 at onset and offset in a format which was for suitable further analysis.

3. Determining the onset and offset of F0

''Pitch'' identified the approximate boundaries for each syllable rhyme based on changes in amplitude. These boundaries were checked manually and adjusted to align with points of onset and offset where post-consonantal fluctuations or fluctuations due to vocal fold closure were negligible. In general, F0 onset was readily identifiable and corresponded to the point where the vowel amplitude reached a post-consonantal maximum. Identification of the pitch offset was more complex, particularly for the tones produced by the implant users. Typically, it corresponded with the point where the vowel amplitude had declined to 50% of the maximum. This criterion could not be applied to syllables ending on a nasal. Word-final nasals tend to have a low amplitude relative to the preceding vowel. The F0 offset marker was manually adjusted to a point near the end of the syllable which was representative of the observed trend in the preceding contour.

Test–retest reliability of this approach for marking tone boundaries was performed on data from four speakers (one adult, one normally hearing child, and two implant users). The time separation between the two analyses was more than 2 years. Intra-Class Correlation values (ICC) using a one-way random effects model were calculated for the six mean tone onsets and offsets and the 12 axis lengths. ICCs calculated for F0 onset and offset were above 0.90 for all tones, indicating high test–retest reliability. Similarly ICCs calculated for all tone axis lengths except for Tone 25 were above 0.9. The ICCs for the Tone 25 axes were 0.80 and 0.78. Though lower than the other values calculated, they still indicate acceptable test–retest reliability.

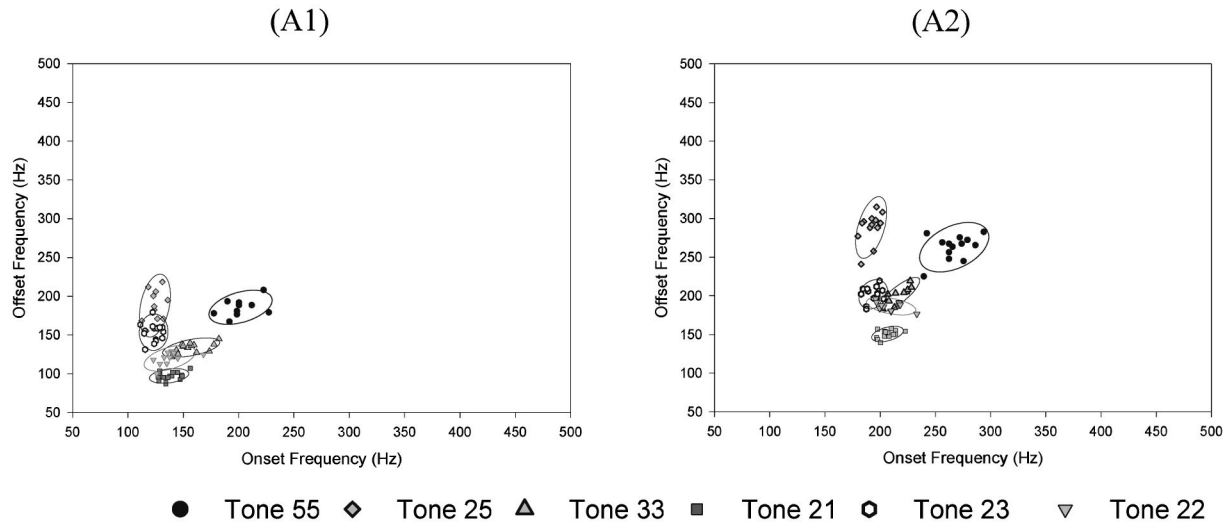


FIG. 2. Tone ellipse plots for two adult speakers. A1 is a male normally hearing adult. A2 is a female normally hearing adult. The plots with the most differentiated ellipses for the group are presented in each case. These plots demonstrate the triangular shape of the tonal space in Cantonese as well as the tightly defined tonal ellipses typical of adult speakers. Some crowding of the three tones (Tone 33, 23, and 22) is also apparent in the center of the tonal space.

C. Data analysis

Three forms of analysis were undertaken to compare tone production in the three groups of speakers. First, as described in the Introduction (Sec. I B), F0 offset versus F0 onset values for all speech tokens were plotted to encompass approximately 95% of the tones tokens making up each tone type. Second, the following parameters were calculated: (a) the lengths of the axes of the tonal ellipses; (b) the areas of the tonal ellipses; and (c) the distances between the center points of each of the tonal ellipses. These parameters were used to calculate Index 1 and Index 2 [Eqs. (1) and (2) above] which in turn function to numerically summarize the degree of tonal differentiation observed in an individual's tonal plot. Third, an auditory transcription analysis of the speech production data for each of the three groups of speakers was undertaken. These data were summarized as tone confusion matrices. Percent correct tones (PCT) was also calculated for each child.

D. Auditory transcription analysis

As illustrated in Fig. 1, there is considerable crowding in the Cantonese tonal system since five of the six tones in Cantonese have similar onsets. Further, tones 25 and 23, and tones 33 and 22 have similar contours and are primarily distinguished by their different offsets. The acoustic similarities among these tones mean that the Cantonese tonal system represents a considerable challenge for reliable transcription. Moreover, the first author who performed the auditory analyses of the data is not a native-speaker of Cantonese. To ensure consistent auditory transcription of the data, protocols for tone labeling were established with reference to linguistic-tonetic description of Cantonese tones developed by Rose (2000). For example, to be labeled a high rise Tone 25, a tone contour had to be uttered with a sharply rising contour and the F0 offset had to be closer to Tone 55 than Tone 33. Similarly, to be labeled as a low-fall Tone 21, a tone had to have a clearly falling contour shape ending with an audibly lower offset than a Tone 22. It is more difficult to label tones in citation form than in sentence context (Wong and Diehl, 2003) and to aid the transcriber in normalizing to a speaker's typical pitch range, all tones produced by a single speaker were transcribed in a block.

For implant users who had good control of F0 across the syllable, but did not use consistently differentiated pitch levels across a series of tone productions, it was often very difficult to identify a pitch range for the speaker or to label specific tones within it. Nonetheless an auditory analysis of F0 was attempted for such speakers despite some doubt about whether it was appropriate to do so. Implant users with poor control of prosodic features such as syllable duration often used pitch patterns that varied unpredictably across the syllable. No attempt was made to perform an auditory transcription of the F0 contours produced by these speakers.

Inter-rater reliability measures were calculated for the first author and a native speaker trained in phonetic transcription (Shriberg and Lof, 1991). Tone transcriptions were compared for one normally hearing child and two implant users. The children varied in intelligibility from highly intelligible to fairly unintelligible. An agreement of 81.8% was observed for the normally hearing child, 80% for the implant user with a high level of intelligibility and 67.6% agreement for the implant user with a low level of intelligibility. These are within inter-transcriber measures reported for transcription of segmental phonemic contrasts in English (Shriberg and Lof, 1991) but less than was reported for transcription of Cantonese tonal contrasts by So and Dodd (1995). They observed 100% agreement between two native-speakers of Cantonese transcribing tones produced by normally hearing Cantonese-speaking children. This may be an overestimation of inter-transcriber reliability, given that the word lists were small and familiar to both transcribers.

III. RESULTS

Tone productions for each speaker were plotted. The plots with the most clearly differentiated tonal ellipses are shown for each group of normally hearing speakers (Fig. 2,

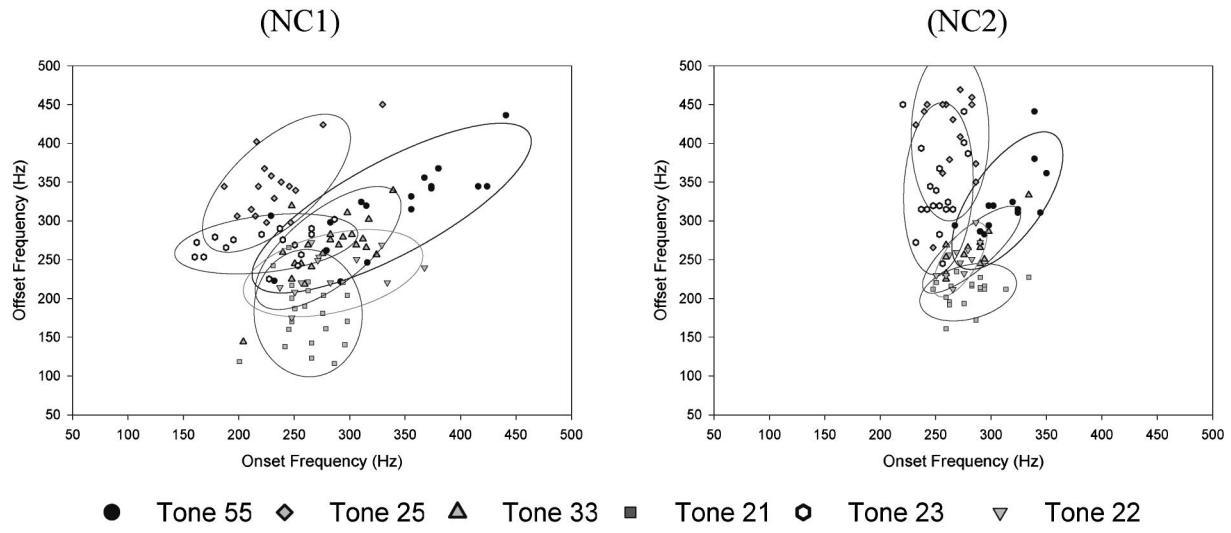


FIG. 3. Tone ellipse plots for two normally hearing children. There was considerable variation apparent in the tone plots from this group and the plots shown here are from children who demonstrated the greatest tonal differentiation. They illustrate the characteristic features of plots from this group, namely: large tonal ellipses located in a large tonal space and reduced tonal differentiation relative to the tonal plots from the adult speakers.

adults; Fig. 3, children).

Tone plots for two male implant users are shown in Fig. 4. The plot for CI1 was typical of this group of speakers. CI2 by contrast demonstrated relatively good tonal differentiation. Only three other implant users in the present study exhibited similar levels of differentiation.

Shapiro–Wilk tests for normality were performed on Index 1 and Index 2; tone ellipse area; *At*, the tonal space; the average of the sum of the two axis lengths for each tone; and, the 15 intertone distances. Only the ellipse area for Tone 33 and four of the 15 intertones distances demonstrated near normal distribution. Statistical analysis of the data thus required the use of the nonparametric Kruskal–Wallis test. Box plots (Figs. 5 and 6) summarize medians and variation around the medians across the three groups for Index 1 and Index 2 (Fig. 5); and, for the largest and smallest tone ellipse areas and the tonal space *At* (Fig. 6).

A. Statistical analysis of the data

1. Statistical comparison of results

Group of speaker and tone area for each tone type were entered into a Kruskal–Wallis nonparametric analysis. Significant differences in median tone area for the three groups of speakers were found for all tones. These differences ranged in significance from [χ^2 (2) = 14.028; $p < 0.001$, Tone 25] to [χ^2 (2) 8.354; $p < 0.05$, Tone 33]. As illustrated in Fig. 6 adult speakers had the smallest median ellipse areas, normally hearing children had the largest median ellipse areas.

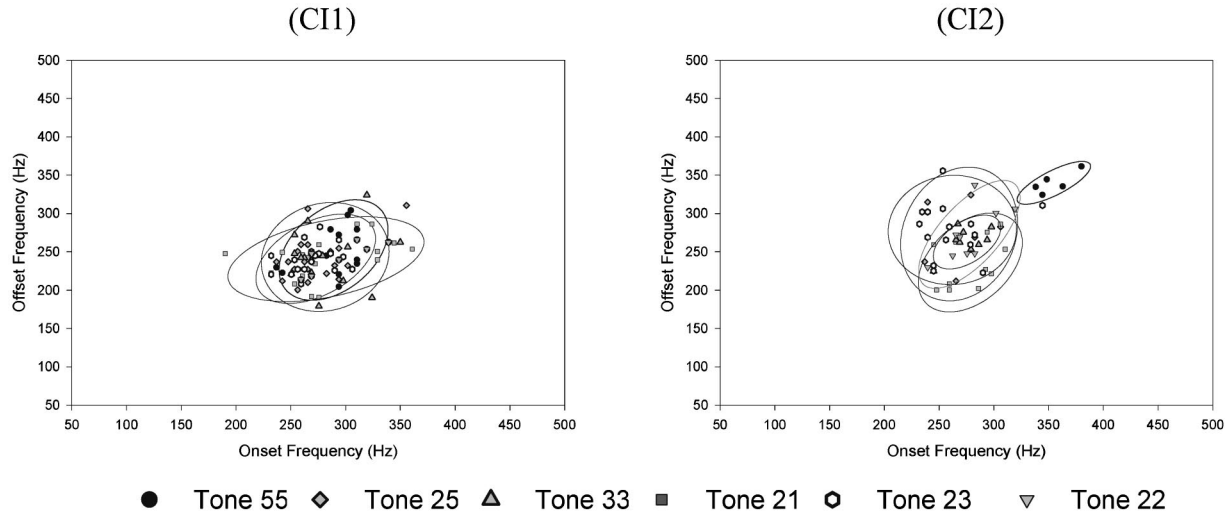


FIG. 4. Tone ellipse plots for two implant users. The plot shown for CI1 is typical of the majority of implant users in the study. The neat centrally located tonal ellipses reflect similar values of F0 onset and offset for most tone tokens regardless of type. The tonal ellipses all cluster in one location in the plot indicating little tonal differentiation. CI2 was one of four implant users who demonstrated reasonable tonal differentiation. The tonal ellipses are more dispersed than they are for CI1. The most clearly differentiated tone is tone 55 which appears as a distinct ellipse separated from the other five tonal ellipses. This suggests a clear opposition between the features high and nonhigh.

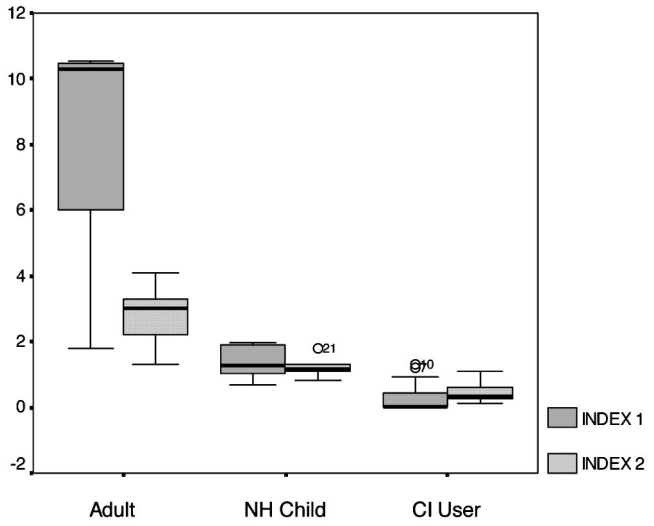


FIG. 5. Box plots for Index 1 and Index 2. The box indicates the interquartile range of values obtained with the median (indicated by the solid horizontal line). The range of measurement is shown by whiskers extending from the box except for points more than 1.5 (indicated by "○") or 3 box lengths (indicated by "*") from the upper or lower edge of the box.

No comparison of the lengths of the ellipse axes was performed because these values were used to calculate the ellipse areas and a similar significant relationship would be expected.

2. Differences in tone ellipse separation for the three groups of speakers

Subject group and intertone distances were entered into a Kruskal–Wallis nonparametric analysis. A significant difference in 14 of the 15 median tone distances for the three groups of speaker was observed. These differences ranged from [χ^2 (2)=18.592; $p<0.0001$ for the median intertone distance between Tones 25 and 21] to [χ^2 (2) 7.865; p <0.05 for the distance between Tones 23 and 22]. No significant difference in the median tone distance between Tone 33 and Tone 22 was observed.

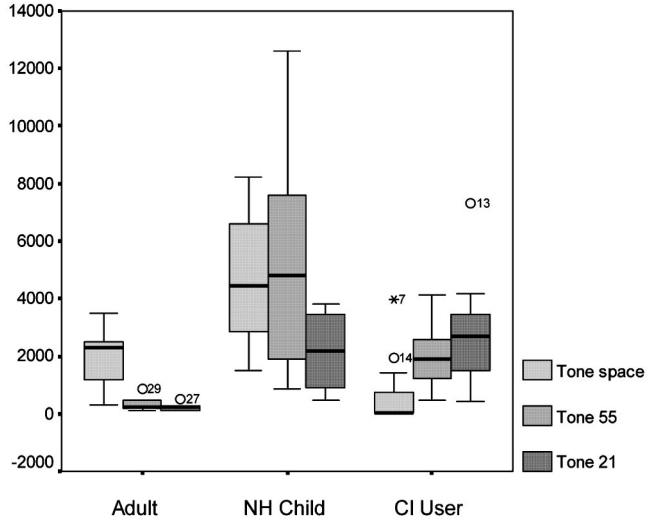


FIG. 6. Box plots summarizing the variation in tonal space and the range of variation in tone ellipse areas for the three groups of speakers. To indicate variation in tone ellipse area, box plots are shown for Tone 55 which was typically one of the larger tonal ellipses together with box plots for Tone 21 which was one of the smaller ellipses among the three groups of speaker.

TABLE I. Tone confusion matrix from an auditory analysis of tones produced by normally hearing adults. Confusions are shown as a proportion of total target productions for each tone.

	Actual					
Target	55	25	33	21	23	22
55	0.97		0.02	0.01		
25		0.88		0.05	0.02	
33			0.94	0.02	0.01	0.03
21				0.90		0.10
23		0.05	0.01		0.94	
22			0.07	0.04	0.01	0.88

The implant users tended to have the smallest intertonal distances with medians ranging from 10.1 Hz (Tone 55–Tone 23) to 32.0 Hz (Tone 55–Tone 23). The range of median intertone distances for the normally hearing children from 147.2 Hz (Tone 25–Tone 21) to 16.9 Hz (Tone 33–Tone 22) as compared with 83.5 Hz and 16.63 Hz, respectively, for the adult speakers.

Tonal space, *At*, and speaker group were entered into a nonparametric Kruskal–Wallis analysis. A significant difference between the median tonal spaces for the three groups of speakers was observed [χ^2 (2) 16.891, $p<0.001$]. The median tonal spaces were 40.23 Hz^2 (implant users); 4452.81 Hz^2 (normally hearing children); and 1965.22 Hz^2 (normally hearing adults).

3. Indexing differentiation

Values of Index 1 and speaker group were entered into a Kruskal–Wallis nonparametric analysis. A significant difference in median values for Index 1 was observed [(χ^2 (2) 18.746 $p<0.001$]. The median values for Index 1 were 0.015 (cochlear implant users); 1.27 (normally hearing children); and 10.28 (normally hearing adults).

Index 2 and speaker group were entered into a Kruskal–Wallis analysis. A significant difference in Index 2 among the three groups of speakers was observed [χ^2 (2) 20.866 $p<0.001$], medians: 0.32 (implant users); 1.17 (normally hearing children); and 3.01 (adults). The order of magnitude for both indices was CI users <NH children <NH adults.

B. Auditory transcription analysis of tone production

The average PCT (percent correct tones) observed for the normally hearing children by auditory analysis was 78%. The observed range of PCT produced was 60%–93%. A Pearson's product moment correlation analysis on PCT and the area of the tonal space *At* showed a significant positive correlation between *At* and PCT ($r=0.843, p<0.01$).

A confusion matrix was prepared for the tones produced by all the normally hearing adults (Table I) and children (Table II). Relative to the children, fewer confusions are apparent in the adult data though the trends are similar with most confusions being apparent between tones 33 and 22, and tones 25 and 23. In all, where the children were required to produce target tones 33 and 22, 20% were confusable,

TABLE II. Tone confusion matrix from an auditory analysis of tones produced by normally hearing children. Confusions are shown as a proportion of total target productions for each tone.

	Actual					
Target	55	25	33	21	23	22
55	0.89		0.11			
25	0.02	0.70	0.01	0.07	0.19	0.01
33	0.04		0.78	0.01		0.17
21			0.02	0.80	0.03	0.15
23	0.01	0.14	0.01	0.01	0.69	0.14
22			0.22	0.04		0.74

16% of tones 2 and 5 were confusable, and 10% of tones 4 and 6 were confusable. This compares with 5%, 6%, and 6%, respectively, in the adult data.

The average PCT produced by the implant users was 38% correct with a range of 21–59 % correct. A confusion matrix for the data is provided in Table III. Contour tones had the lowest percent correct scores and typically they were perceived as either Tone 33 or Tone 22.

IV. DISCUSSION

This study compared tone productions by children using a cochlear implant with two groups of normally hearing speaker. The aim was to demonstrate the application of a methodology for the acoustic analysis of tone production. In this methodology, F0 offset was plotted against F0 onset for a series of tone tokens and the degree of differentiation between ellipses encompassing each tone target was observed qualitatively and then summarized numerically. Since F0 onset and offset were the only points plotted, the emphasis in the methodology was on observing pitch level differences between tone types. Some contour effects were also reflected by overall position of individual ellipses in the tone plot.

This approach is quite logical for the analysis of Cantonese, since pitch level has been suggested to be perceptually more salient to the Cantonese-speaker than pitch contour (Gandour, 1983). Given the relative perceptual importance of pitch level to speakers of Cantonese and the fact that pitch level differences between some tones are quite small (Fig. 1) one would predict that the range of onset and offset values for each of the six citation form tones would necessarily be narrow to maintain pitch level differences, i.e., all tokens for a particular tone target would congregate in tightly defined ellipses covering restricted areas of the tonal plot.

TABLE III. Summary of tone confusions from an auditory analysis of tones produced by children using a cochlear implant. Confusions are shown as a proportion of correct tones per total number of target tones.

	Actual					
Target	55	25	33	21	23	22
55	0.39	0.03	0.40	0.02	0.02	0.14
25	0.10	0.19	0.36	0.07	0.13	0.15
33	0.06	0.01	0.64	0.05	0.04	0.20
21	0.04	0.03	0.29	0.23	0.06	0.36
23	0.02	0.09	0.33	0.04	0.22	0.28
22	0.04	0.0	0.29	0.06	0.03	0.58

This prediction was confirmed in part by the tone production data for the adult speakers. Among this group of speakers, greatest overlap was observed between tones 33, 23, and 22. Some overlap was also observed between tones 25 and 23.

The plots obtained for the normally hearing children and the implant users were quite different to those obtained for the group of normally hearing adults. First, the areas of the tone ellipses for both groups of children were larger than they were for the older speakers. Second, the median size of the tonal space *At* for each of the three groups of speakers were significantly different from each other.

The tonal space for the implant users was smaller than it was for either group of normally hearing speaker. Combined with the relatively diffuse tonal ellipses, this resulted in little or no clear differentiation among the tones produced by these children. By contrast median value of *At* for the normally hearing children was over twice as large as median *At* observed for the normally hearing adults (4452.81 Hz^2 versus 1965.22 Hz^2). This suggests that among normally hearing speakers, there is a direct relationship between the spread of pitch used for each tone type and the size of the tonal space in which the tonal ellipses are located. The young children in this study had diffuse tonal ellipses but maintained a reasonable degree of differentiation by virtue of their broad tonal spaces.

Given the similarities in F0 among some tones in Cantonese, it is interesting to note how the normally hearing children had significantly larger tonal ellipse areas than the implant users. This suggests a greater spread of pitch usage for each tone type. The normally hearing children are reported to master tone production early and because of having access to all available auditory information about tone, one might have predicted that their tonal ellipses would have been smaller than those of the implant users.

Testing conditions for the two groups of children were kept constant, however the normally hearing children were younger than many of the implant users and were less used to such tasks. These factors may partially explain some of the variation observed between the two groups of children. It is further possible that the younger normally hearing children are still developing control of vocal fold use during speech further explaining some of the differences in the spread of their tonal ellipses relative to the other groups of speakers. Nonetheless, the degree of difference in the size of the tonal spaces and tonal ellipses for the three groups of speakers suggests that there are real differences in tone production abilities among the three groups. These results further suggest that it is not enough to evaluate differentiation among individual tones through examination of the spread of F0 usage. Measures of tone ellipse separation within the tonal space and degree of ellipse overlap are also required.

Index 1 and Index 2 were developed to meet this requirement and to summarize the interplay between the ellipse area and tonal space in determining the degree of tone differentiation. The three groups of speakers were clearly identifiable based on their observed values of Index 1 and Index 2. The medians of both indices were significantly greater for the adults than for the normally hearing children.

In the case of the implant users, the median for Index 1 was close to zero and for Index 2 was barely above noise. Both values indicate little to no differentiation in the implant users' production of tone.

A. Auditory transcription analysis of the data

Normally hearing children averaged 78% correct tone production, though there was considerable variation. Among the implant users, there was an average PCT of 38% (range 21%–59%). This is barely above chance level. Two points can be made with respect to these data.

First, Wong and Diehl (1999) found that the size of the tonal space of the speaker seems to affect the relative ease with which the different tones produced by a Cantonese speaker can be perceived. This observation also holds true for the participants in this study. In the case of the young normally hearing children a significant positive correlation was observed between the size of the speaker's tonal space and the percentage correct tones produced by the speaker ($p<0.01$).

Second, these observations are significantly different to results from previous studies investigating tone development in children acquiring Cantonese. Typically, it is reported that Cantonese-speaking children acquire a complete toneme inventory by two years of age and make very few tonal errors thereafter (So and Dodd, 1995). Similarly children with impaired hearing have been reported to be able to acquire most if not all tones in Cantonese (Dodd and So, 1994). Could the differences in the results presented here reflect the fact that the auditory analyses were performed by a non-native speaker of Cantonese? This does not seem to be an adequate explanation, given that the auditory analyses are supported by acoustic analysis of the same data.

B. A conundrum: Tone production is mastered before tone perception

The literature suggests that tone production is mastered early by both hearing and deaf children learning Cantonese. By contrast, adultlike tone perception skills do not develop until around the age of 10 years (Ching, 1984). Ching concluded that the younger children in her study were generally able to identify gross pitch contour differences but they still had to develop the capacity to normalize for pitch level differences across tones.

In the case of profoundly hearing-impaired children, Ching (1988) reported that they perform at chance levels on tone perception tests except when listening to larynx tones presented as sinusoids. ''High'' was the best perceived tonal feature and the children showed a basic opposition in their tone identification between high and nonhigh. Further, Dodd and So (1994) also noted that tone production skills were better than tone perception skills in the hearing-impaired children in their study.

Despite some discussion about the exact relationship between speech perception and the development of spoken language, it is generally agreed that perception must precede production. How is it that the acquisition of good tone production abilities in normally hearing children seems to outstrip the acquisition of accurate tone perception abilities? Or, in the case of children with impaired hearing, how is it that despite a reduced ability to accurately perceive tonal contrasts, these children reportedly acquire and accurately produce most tonal contrasts in Cantonese?

The acoustic analysis of the tone production data indicated significant differences in tone production abilities between profoundly hearing-impaired prelinguistically deafened speakers, normally hearing children and normally hearing adults. The primary tonal confusions by auditory analysis among the normally hearing children were between tones 33 and 22, tones 23 and 25, and tones 21 and 22. This was also reflected in individual plots of F0 offset versus F0 onset for each of the children such that there was considerable overlap for tones 33 and 22 and tones 23 and 25 and also some overlap in the ellipses produced for tones 21 and 22 (refer to Fig. 3 for two examples). Thus, while it is true that tone confusion results are dependent on the skills of the transcriber to correctly identify and label the tones produced, acoustic analysis of the same data supports the results from auditory analyses.

Among the implant users, only four children demonstrated any meaningful tonal differentiation (CI2 in Fig. 4 is an example of one of these children). However, auditory analysis showed that though most of the implant users produced a range of F0 contours that could be labeled as corresponding to a particular toneme in Cantonese, in fact, they did not occur with sufficient frequency in the phonologically correct context to be judged as having been acquired. Auditory analyses averaged across the group of implant users showed that the likelihood of correct production of a particular tone type on the correct syllable was barely above chance at 38%. Once again these results match reported tone perception results for implant users. Ciocca *et al.* (2002) using a simple, forced-choice tone identification test involving minimal pair alternatives found that most participants performed below chance on most of the 15 tone contrasts in Cantonese. Further, like the hearing-impaired children in Ching's study (1988), they noted that implant users were more successful at identifying tones with larger separations in average pitch levels. This effect is evident in the plot for CI2 (Fig. 4) where the Tone 55 ellipse is quite distinct from the other tonal ellipses in the plot.

Overall the results from the acoustic analysis support results from auditory analyses. Furthermore, results from both methodologies reflect observations from studies investigating tone perception abilities in normally hearing children and children using a cochlear implant. Moreover the large tonal spaces noted for the young normally hearing children in this study may provide interesting evidence from the analysis of speech production materials to support Ching's hypothesis (1984) that Cantonese-speaking children of this age are still acquiring skills for normalization of pitch level differences among tones.

Taken together, these results suggest a close link between a child's tone perception abilities and the subsequent development of tone in spoken language. From these data, the possibility that mastery of tone production outstrips the development of adultlike tone perception abilities is not tenable and the potential conundrum disappears.

V. CONCLUDING REMARKS

A key aim for this study was to develop a means for analyzing tone production acoustically to support or enhance auditory analyses of speech production materials. F0 offset versus F0 onset was plotted to examine general tendencies in F0 usage for different tone types. Two indices were also developed for numerically summarizing the degree of differentiation between tone ellipses. Using this approach, it was shown that normally hearing adult speakers have small ellipses located in a relatively small tonal space. Young normally hearing children have large tonal ellipses together with a large tonal space. These children are reported to have mastered tone production by the age of two yet the acoustic data suggest that their production skills are still maturing. An interesting aspect of these results is that they support conclusions from a variety of studies investigating Cantonese tone perception.

There are considerable differences between the auditory analyses reported here and results from previous studies investigating tone production by young speakers of Cantonese. These may reflect the fact that the analyses were performed by a non-native speaker of Cantonese. However, acoustic analyses of the data provide effective back-up for the auditory analyses performed and raise some interesting questions about the nature of tone perception. For one thing, an important role for linguistic experience when labeling tones is suggested by the fact that young children who have such broadly spread overlapping tonal ellipses are judged to have mastered tone production by native speakers. This raises a further question regarding the degree to which tone perception is categorical.

In summary, this approach to the acoustic analysis of tone supports what is known about tone perception abilities in young speakers and enhances understanding about tone production based on auditory analyses. It is proposed that the procedure has application both in cross-sectional studies to observe differences between groups of speakers and in longitudinal studies to observe developmental changes in tone production. It may also have application for aiding in tone habilitation. A prototype habilitation tool employing this methodology has since been developed which can be downloaded. (URL http://www.psych.ox.ac.uk/oscci/johanna%20barry/v2/index.htm)

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of the children, their families and the staff of the Prince of Wales Hospital Shatin New Territories Hong Kong. In particular, we thank Dr. Kathy Lee who provided the list of words used in data collection and Grace Chan who phonetically transcribed samples for obtaining inter-rater transcription measures. Dr. Johanna Barry was funded by an Australian Postgraduate Award and a Victoria Fellowship. Further support was also provided from an Australian National Health and Medical Research Council (NHMRC) project Grant No. 970257 and the Bionic Ear Institute, Melbourne. Dr. Peter Blamey was supported by an Australian NHMRC Senior Research Fellowship (No. 208998). The research was conducted with the approval of the Human Research and Ethics Committee of the Royal Victorian Eye and Ear Hospital (No. 99/376H) and the Clinical Research Ethics Committee at the Chinese University of Hong Kong (No. CRE 9271).

Abramson, A. S. (**1986**). ''The Thai tonal space,'' Status Report of Speech Research (No. 85), Haskins Laboratories, pp. 105–114.

Ching, T. (**1984**). ''Lexical tone pattern learning in Cantonese children,'' Lang. Learn. Commun. **3**, 243–414.

Ching, T. (**1988**). ''Lexical tone perception by Cantonese deaf children,'' in *Cognitive Aspects of the Chinese Language Vol. 1*, edited by I. M. Liu, M. J. Chen, and H. C. Chen (Asian Research Service, Hong Kong), pp. 93–102.

Ciocca, V., Francis, A., Aisha, R., and Wong, L. (**2002**). ''The perception of Cantonese lexical tones by early-deafened cochlear implantees,'' J. Acoust. Soc. Am. **111**, 2250–2256.

Clumeck, H. (**1980**). ''The acquisition of tone,'' in *Child Phonology: Production Vol. 1*, edited by G. H. Yeni-Konshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 257–275.

Dodd, B. J., and So, L. K. H. (**1994**). ''The phonological abilities of Cantonese-speaking children with hearing loss,'' J. Speech Hear. Res. **37**, 671–679.

Gandour, J. (**1983**). ''Tone perception in Far Eastern languages,'' J. Phonetics **11**, 149–175.

Gandour, J., Weinberg, B., Petty, S. H., and Dardarananda, R. (**1988**). ''Tone in Thai alaryngeal speech,'' J. Speech Hear Disord. **53**, 23–29.

Gandour, J. T. (**1978**). ''The perception of tone,'' in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 41–76.

Keating, P. (**1980**). ''Patterns of fundamental register and vocal registers,'' in *Infant Communication: Cry and Early Speech*, edited by T. Murry and J. Murry (College-Hill, Houston, Texas), pp. 209–233.

Kent, R., and Murray, A. D. (**1982**). ''Acoustic features of infant vocalic utterances at 3, 6, and 9 months,'' J. Acoust. Soc. Am. **72**, 353–365.

Lee, K. (**2003**). ''The construction and validation of a speech perception test for Cantonese-speaking children,'' Ph.D. thesis, Chinese University of Hong Kong.

Li, C. N., and Thompson, S. A. (**1977**). ''The acquisition of tone in Mandarin-speaking children,'' J. Child Lang **4**, 185–199.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). ''A new portable sound processor for the University of Melbourne/Nucleus Limited multi-electrode cochlear implant,'' J. Acoust. Soc. Am. **91**, 3367–3371.

Rose, P. (**1987**). ''Considerations in the normalization of the fundamental frequency of linguistic tone,'' Speech Commun. **6**, 343–351.

Rose, P. (**1989**). ''On the non-equivalence of fundamental frequency and linguistic tone,'' in *Prosodic Analysis and Asian Linguistics to Honour R. K. Sprigg. Pacific Linguistics C-104*, edited by D. Bradley (Dept. of Linguistics, Research School of Pacific and Asian Studies, Australian National University, Canberra), pp. 55–82.

Rose, P. (**1993**). ''A linguistic-phonetic acoustic analysis of Shanghai tones,'' Aust. J. Ling. **13**, 185–220.

Rose, P. (**2000**). ''Hong Kong Cantonese citation tone acoustics: A linguistic tonetic study,'' Proceedings of the 8th Australian International Conference on Speech Science and Technology, Canberra, pp. 198–203.

Shriberg, L. D., and Lof, G. L. (**1991**). ''Reliability studies in broad and narrow phonetic transcription,'' Clin. Ling. Phonetics **5**, 225–279.

So, L. K. H., and Dodd, B. J. (**1995**). ''The acquisition of phonology by Cantonese-speaking children,'' J. Child Lang **22**, 473–495.

Speech tools support (**1999**). *Speech Analyzer*, Ver. 1.5, Test version 10.6 (Summer Institute of Linguistics Dallas, Texas).

Tse, A. C.-Y. (**1991**). ''The acquisition process of Cantonese phonology—a case study,'' M. Philosophy, University of Hong Kong, Hong Kong.

Tse, J. K.-P. (**1978**). ''Tone acquisition in Cantonese: A longitudinal case study,'' J. Child Lang **5**, 191–204.

Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). ''Speech perception as a function of electrical stimulation rate: Using the Nucleus-24 cochlear implant system,'' Ear Hear. **21**, 608–624.

Wong, P., and Diehl, R. L. (**2003**). ''Perceptual normalization for inter-and intra-talker variation in Cantonese level tones,'' J. Speech Hear. Res. **46**, 413–421.

Wong, P., and Diehl, R. L. (**1999**). ''How Parkinson's disease affects tone language and [what it] tells us about tone comprehension,'' http://www.acoustics.org/press/137th/wong.html

Integration across frequency bands for consonant identification

Diane Ronan,[a] Ann K. Dix,[b] Phalguni Shah,[c] and Louis D. Braida[d]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 4 November 2003; revised 3 June 2004; accepted 7 June 2004)

A comparison of the predictions of models of integration to data on the reception of consonants filtered into a variety of frequency bands is reported. New data on the consonant identification are presented. Three experiments were conducted testing the following bands: experiment I, 0–2100 Hz and 2100–4500 Hz; experiment II, 0–700 Hz combined with 700–1400, 1400–2100, 2100–2800, and 2800–4500 Hz; experiment III, all combinations of 700–1400, 1400–2100, 2100–2800, and 2800–4500 Hz. The predictions of four models, Fletcher's [*Speech and Hearing in Communication* (Van Nostrand, New York, 1950)] independent errors model, Massaro's fuzzy logical model of perception [Proc. Int. Congress of Phonetic Sciences, Stockholm, Vol. 3, pp. 106–113 (1987)], and Braida's pre-labelling and post-labelling models of integration [Q. J. Exp. Psychol. A **43**, 647–677 (1991)], were compared in terms of their ability to predict combined-band scores. At least two models were capable of predicting performance for each combined-band condition. For experiment I, all models were able to make satisfactory predictions. For experiment II, a variant of the pre-labelling model was able to make satisfactory predictions. For experiment III, no model was able to make satisfactory predictions, but the fuzzy logical model of perception and a variant of the pre-labelling model made relatively good predictions. Thus the ability of the models to predict performance depended more on whether the condition included the lowest frequency band than on the adjacency or frequency separation. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777858]

PACS numbers: 43.71.Es [PFA] Pages: 1749–1762

I. INTRODUCTION

The cues used to identify speech segments are distributed broadly over a frequency range of more than five octaves rather than confined to a small number of narrow spectral regions. Auditory processing by contrast, exhibits a consistent tonotopic organization at a variety of levels in the nervous system. One consequence of this tonotopic organization is that narrow band stimuli primarily influence the perception of signals in nearby spectral regions. Although the physiological processing of speech sounds is increasingly being studied, many details are poorly understood. However, based on the proximity of spectral processing regions, any tonotopically organized system may be expected to deal with sounds that contain spectrally adjacent cues differently from sounds that contain spectrally distinct cues.

The most relevant perceptual studies of the spectral distribution of speech cues have employed filtered speech stimuli. The reception of speech processed by lowpass and highpass filtering has been studied extensively. Together with studies of the masking of speech by additive noise, these studies form the basis of current standardized methods of estimating the effects of linear filtering and additive noise distortion, such as the articulation index (AI, ANSI, 1969) the speech transmission index (STI, Steeneken and Houtgast, 1980) or the speech intelligibility index (SII, ANSI, 1997).

These formulations assume that the intelligibility score in a combined-band case is related nonlinearly to the scores in the separate bands that comprise it. Specifically the combined-band score is a monotonic function of an index formed as a simple linear sum of the indices that represent, nonlinearly, the scores in the separate frequency bands. The additivity assumption has been called into question, however. For example, Grant and Braida (1991) found that for combinations of adjacent bands (70–700 Hz combined with 700–1100 Hz, and 1800–2800 Hz combined with 2800–3500 Hz), the sum of the AI values exceeded the observed AI value by 18%. Recently, Steeneken and Houtgast (1999) have argued that ''redundancy'' is responsible for the sum of the indices of adjacent bands overstating their contribution to the combined index and have proposed a correction to the AI calculation procedure in the case of octave bands of speech.

By contrast the reception of speech processed by relatively narrow bandpass filters has received relatively little systematic study. Rather the effects of bandpass filtering are usually inferred from the results of high- and lowpass filtering by differencing techniques. The inadequacy of the AI approach to the problem of integrating speech cues across frequency has long been recognized. When more than one narrow band of speech is presented simultaneously, reception can improve substantially beyond the level predicted by the articulation index. For example, Kryter (1960) argued that the standard AI calculation underestimates the intelligibility that results when separated narrow bands are presented simultaneously. He showed three separated 500-Hz-wide

[a]Current address: Oregon Hearing Research Center, Portland, OR 97239. Electronic mail: ronand@ohsu.edu
[b]Current address: SAR Health Sciences, Boston University, Boston, MA 02115. Electronic mail: dix@bu.edu
[c]Current address: 5913 Ricky Ridge Trail, Orefield, PA 18609. Electronic mail: pshah@mit.edu
[d]Author to whom correspondence should be addressed. Current address: 36-747 M.I.T, Cambridge, MA 02139. Electronic mail: ldbraida@mit.edu

bandpass filters centered at 500, 1500, and 2500 Hz yield intelligibility scores for phonetically balanced words that exceed those for single bandpass filters of equivalent total bandwidth (1500 Hz) even though the single band systems have higher predicted AI values. Grant and Braida (1991) found that for combinations of nonadjacent bands (70–700 combined with 1800–2800 Hz or 2800–3500 Hz, and 700–1100 Hz combined with 2800–3500 Hz) the sum of the AI values was less than the observed AI value by 41%. Warren *et al.* (1995) reported that very high intelligibility can be achieved from combinations of bands that are widely separated in center frequency. He showed that the intelligibility of sentences processed by one-third octave filters centered at 370 and 6000 Hz was much larger, when the bands were combined, than expected based on the intelligibilities of the individual bands. Although Warren *et al.* did not interpret their results in terms of AI predictions, they report improvements relative to the component band conditions that are ''superadditive,'' such as those occasionally seen in cases of audiovisual integration. Recently, Müsch and Buus (2001) have proposed a model of speech intelligibility that accounts for a much larger proportion of this difference than the standard SII model.

Although many theories of speech perception assume that speech cues are readily integrated across frequency once nonuniform auditory spectral resolution is accounted for, much auditory research has focused on limitations on the ability to integrate acoustic cues for simple sounds across frequency. Lim *et al.* (1977) reported that when the loudnesses of two sequentially presented tones are compared in a ''roving-level'' paradigm, resolution decreases as the frequency difference between the tones increases. Green (1988), although demonstrating that listeners can employ details of spectral shape remote from a signal frequency to improve detection performance, also documented limitations on this ability. Yost and Sheft (1990) reported that the ability to detect amplitude modulations of one tone can be reduced when other tones with similar modulations but different frequencies are presented. Both Sheft and Yost (1992) and Takeuchi and Braida (1993) reported that listeners have considerable difficulty comparing the envelope modulations of noise bands having different center frequencies. Hall and Grose (1990) reported that while listeners are able to benefit from the presentation of a second narrow band noise with an envelope similar to that of a masker noise band when detecting a tone located in the masker, the improvement is markedly reduced when the second noise is displaced from the first by more than one critical band.

The integration of speech cues across frequency is also relevant to the design of signal processing schemes for the hearing-impaired and the deaf. Frequency selectivity is often reduced in listeners with sensorineural impairments. Although this is often most evident in measures of ''spread of masking,'' it also implies that speech cues are not segregated normally into frequency bands, so that the cue-integration problem for the hearing impaired may be different than for listeners with normal hearing. Several recently developed processing schemes treat speech separately in several frequency bands. For example, multiband amplitude compression systems (e.g., Villchur, 1973; Lippmann *et al.*, 1981; De Gennaro *et al.*, 1986; Bustamante and Braida, 1987; Rosengard *et al.*, 2004) reduce the range of level variation in 4–16 bands to match the processed signal to the residual dynamic range of the impaired listener. Although the failure of these schemes to provide substantial benefit relative to linear amplification has been attributed to short term alterations of the overall spectral envelope, the extent to which it results from within-band distortions as opposed to across-band distortions has not been studied.

In addition to studying the perception of filtered speech signals, the envelopes of filtered bands of speech have also received attention. They have been found to provide a basis for the design of processing schemes intended for acoustic aids and cochlear implants (e.g., Breeuwer and Plomp, 1984; Grant *et al.*, 1993; Kasturi *et al.*, 2002; Shannon *et al.*, 1998, 2001). However, listeners do not always benefit from the presentation of multiple envelopes, particularly when the envelopes are presented as amplitude modulations of closely spaced sinusoidal carriers. The extent to which this represents a failure to integrate the cues for the different band envelopes as opposed to simpler types of perceptual interference, such as masking, has not been determined.

Finally, recent research on audiovisual speech reception (e.g., Massaro, 1987; Braida, 1991) has led to the development of models of perceptual integration. These models have been formulated abstractly, so that they are potentially capable of describing the integration of auditory cues across spectral bands. Some of the models proposed are of the ''optimal processor'' type in that the performance predicted in the combined-band condition is the highest possible consistent with the performances observed in the isolated single-band conditions assuming that simultaneous presentation neither creates interband cues nor results in interband interference effects, such as masking. Interband cues might arise from comparison of levels across bands. Although the existence of interband cues is often hypothesized, direct quantitative evidence is lacking. Several studies have documented conditions in which interband interference is to be expected (e.g., Franklin 1969, 1975, 1979; Goshorn and Studebaker, 1984), and the ANSI procedure for calculating the articulation index recognizes the possibility of spread of masking from one band of speech to another. Dichotic presentation of the component bands provides an obvious means for controlling peripheral interference effects, although it is unclear whether speech cues presented contralaterally are integrated as well as those presented ipsilaterally (see Loizou *et al.*, 2003).

In the remainder of this paper we describe a set of experiments intended to evaluate across-band integration for consonant sounds (Sec. II), and report the results in terms of identification scores (Sec. III). Models of across-band integration are described in Sec. IV and applied to the experimental results in Sec. V. Finally, we discuss the implication of these findings and suggest future research in Sec. VI.

II. EXPERIMENTS

We report the results of three identification experiments using filtered consonant-vowel-consonant (CVC) syllables.

A. Materials

The CVC materials (Lippmann *et al.*, 1981) were recorded by one male and one female talker in the form /ə/-CVC, where /ə/ is the unstressed schwa. The vowels were /i, ɑ, u, ɪ ɛ, ʊ/ and the consonants were /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, dʒ, tʃ, h, ʒ/. The mean durations of the syllables spoken by each of the two talkers were 634 and 574 ms.

The test items were divided into two groups, one for testing and one for practice. Testing was restricted to syllables made up of the 12 consonants /p, t, k, b, d, g, θ, s, ʃ, v, ð, and z/. The practice items had excluded consonants (dʒ, tʃ, ʒ, h, and f) in either the initial or final positions. There were 864 items in the set of full set of syllables and 486 items in the set of 12 test syllables. Materials were lowpass filtered at 9 kHz then converted to 12-bit digital samples at a sample rate of 20 kHz.

The initial and final consonants for a CVC token were independently drawn with probability 1/12 from the set of 12 consonants, allowing for duplications of CVCs (one male and one female token) and omissions. Tests were conducted in groups of 48 items.

B. Listeners

The listeners in all experiments were young adults with clinically normal hearing (thresholds better than 20 dB HL at audiometric test frequencies from 200 to 8000 Hz). In experiment I there were five listeners (3 F and 2 M), in experiment II there were four listeners (2 F and 2 M), while in experiment III there were five listeners (3 F and 2 M). No listeners participated in more than one experiment.

C. Filtering

Before filtering the CVC materials were downsampled to a bandwidth of 5000 Hz. All filtering was performed by linear-phase FIR filters. Filter lengths varied from 896 to 1216 points. Transition region widths were 50 Hz, with out-of-band attenuations of 80 dB. After filtering, speech-shaped noise (at +35 dB SNR) was added to the filtered signal to obscure the speech in the stop bands. Filter bands were (A) 0–700, (B) 700–1400, (C) 1400–2100, (D) 2100–2800, and (E) 2800–4500 Hz. These bands are fairly similar to those studied by Wang *et al.* (1978) except that the band from 1400 to 2800 Hz was divided in two (bands C and D) to avoid ceiling effects on performance. The bands produced roughly equal levels of performance (±7 percentage points) when tested in isolation.

In addition we tested various combinations of these bands. In experiment I, we tested (F), composed of bands A, B, and C and also (G), composed of D and E. In experiment II, we tested A in combination, separately, with bands B, C, D, and E. In experiment III, we tested all combinations of bands B, C, D, and E.

D. Conditions

Experiment I was intended to investigate the ability of listeners to integrate cues for consonants from two, relatively wide, bands of speech. Further, the two bands were presented either to the same ear (homolateral presentation), or to different ears (contralateral presentation). This experiment was part of a larger study of the benefits that might be obtained from dichotic presentation. In experiments II and III, diotic presentation was used exclusively. In experiment I the CVC syllables were filtered into component bands F and G. One listener received F in the left ear, the other four received it in the right ear. The combined bands (F+G) were then presented both monaurally, in which both bands F and G were presented to one ear, and dichotically, in which F was presented to one ear and G to the opposite ear. Each consonant was presented to each listener 48 times in each of the single-band conditions and 24 times in each of the combined-band conditions.

Experiment II was intended to investigate the ability of listeners to integrate cues for consonants from two, relatively narrow, bands of speech, a low frequency band and various higher frequency bands. Each of the five component bands (A, B, C, D, and E) and combined bands consisting of bands A+B, A+C, A+D, and A+E were presented diotically, to each of four listeners. Each consonant was presented to each listener roughly 92 times in each of the four single-band conditions and roughly 32 times in each of the combined-band conditions.

Experiment III was intended to investigate the ability of listeners to integrate cues for consonants from two to four bands of speech that excluded the low-frequency band. In experiment III the CVC syllables were filtered into component bands B, C, D, and E. Each of these four component bands and combined bands consisting of all 11 possible combinations of the component bands were then presented diotically to each of five listeners. Each consonant was presented to each listener roughly 200 times in each of the four single-band conditions and roughly 72 times in each of the combined-band conditions.

In all experiments, more data was collected in the single-band conditions than in the combined-band conditions to facilitate the application of models to the data (Sec. IV). Most of these models require detailed data on the confusions made by individual observers in identifying specific consonants in the single-band conditions. We collected less data in the combined-band conditions because we were only interested the overall consonant identification score for these conditions.

E. Test procedures

The processed materials were presented over TDH-49 headphones at an unfiltered listening level of 75 dB SPL. To minimize the contributions of out-of-band speech components, a continuous speech-shaped background noise was presented continuously at a level 40 dB below that of the unfiltered speech.

Before testing, for each processing condition listeners practiced identifying syllables with correct-response feedback provided. Training proceeded until scores appeared to reach a plateau. The materials used during these practice sessions were never presented during the test sessions. Listeners were tested without correct answer feedback.

TABLE I. Identification scores for initial (I) and final (F) consonants in experiment I for various bands and combinations of bands. Presentation conditions include homolateral (H, single-band sounds were presented to one ear), contralateral (C, the high-band was presented to the opposite ear), monaural (M, both bands were presented to the homolateral ear), and dichotic (D, the low-band sound was presented to the homolateral ear and the high-band sound was presented to the contralateral ear). Average scores (across four listeners) as well as the minimum and maximum scores achieved by the listeners are reported.

Consonant	Band(s)	Presentation condition	Minimum score	Average score	Maximum score
I	F	H	76.9	80.4	85.8
I	G	H	57.4	68.3	80.0
I	G	C	60.6	66.9	81.4
I	F+G	M	85.8	90.9	93.8
I	F+G	D	83.0	87.6	93.4
F	F	H	67.7	75.0	81.3
F	G	H	51.6	68.0	76.0
F	G	C	48.8	65.1	79.7
F	F+G	M	84.0	89.4	92.0
F	F+G	D	82.6	87.9	93.4

F. Data analysis

The data from the experiments were analyzed by constructing confusion matrices for each listener and each filtering and presentation condition, averaged over testing run and vowel. Data were analyzed separately for initial and final consonants, by averaging over the presentations of each of these consonant groups. These matrices were then analyzed to determine an overall percentage of correct response score S, computed from the entries in the confusion matrix, $\{N_{ij}\}$.

$$S = 100 \frac{\Sigma_{i=1}^{M} N_{ii}}{\Sigma_{i=1}^{M} \Sigma_{j=1}^{M} N_{ij}}. \quad (1)$$

We present the identification scores averaged over listeners, and the ranges of identification scores. We discuss the results of an analysis of variance on the arc-sine transformed identification scores with listener, band(s), and consonant type (initial/final) as factors. We consider results to be significant at the 0.05 level.

III. RESULTS

A. Experiment I

The results of experiment I (Table I) indicate that the highest average scores (typically 88%–91% correct) were obtained for the wideband condition (F+G, 0–4500 Hz), followed by the low-frequency (F, 0–2100 Hz) band (75%–80% correct), and the high-frequency (G, 2100–4500 Hz) band (65%–68% correct). With the possible exception of the low-frequency band, average scores for the various bands were nearly the same for initial and final consonants. Average scores for the monaural presentation condition were two to three points higher than average scores for the dichotic presentation condition.

The results of the ANOVA indicate that listeners and bands-modality were statistically significant and contributed substantially (74% for bands-modality, and 15% listeners) to the total variance. Of the remaining factors, all except for the interaction between bands and consonant types were significant, but none contributed more than 4% to the total variance.

TABLE II. Identification scores for initial (I) and final (F) consonants in experiment II for various bands and combinations of bands. Average scores (across four listeners) as well as the minimum and maximum scores achieved by the listeners are reported.

Consonant	Band(s)	Minimum score	Average score	Maximum score
I	A	43.8	47.2	52.0
I	B	43.9	47.1	52.5
I	C	42.2	48.3	53.3
I	D	39.2	40.7	43.0
I	E	44.5	49.2	51.7
I	A+B	52.2	58.9	61.6
I	A+C	58.0	64.5	71.8
I	A+D	58.7	63.0	66.9
I	A+E	58.8	68.1	74.9
F	A	31.2	34.3	39.7
F	B	40.7	42.5	43.4
F	C	48.0	49.2	50.3
F	D	41.6	44.2	46.9
F	E	43.4	47.1	53.7
F	A+B	48.6	53.9	58.2
F	A+C	56.2	60.4	63.1
F	A+D	54.6	61.7	67.7
F	A+E	56.3	64.8	74.9

B. Experiment II

The results of experiment II (Table II) indicate that the highest average scores were obtained for the combined-band conditions (54%–68% correct), followed by the single-band (B,C,D,E) conditions (41%–49% correct), and the low-frequency (0–700 Hz) band condition (34%–47% correct). With the exception of the low-frequency band, where scores for initial consonants were higher by roughly 13 points, average scores for the various bands were roughly the same for initial and final consonants.

The results of the ANOVA indicate that bands and listeners were statistically significant factors and contributed substantially (77% for bands and 3% for listeners) to the total variance. Of the remaining factors, all except for the interaction between listeners and types were significant, but none contributed more than 7% to the total variance.

C. Experiment III

The results of experiment III (Table III) indicate that average scores generally fell into four groups. Scores for single bands (45%–57% correct) were generally lowest, while those for pairs of bands adjacent in frequency (67%–70% correct) were somewhat higher but below those for non-adjacent pairs (73%–78% correct). The highest scores (78%–90% correct) were obtained when three or four bands were combined. There was also little difference between scores for the initial and final consonant type: 3.3 percentage points averaged across listeners and bands.

The results of the ANOVA indicate that bands and listeners were statistically significant factors and contributed

TABLE III. Identification scores for initial (I) and final (F) consonants in experiment III for various bands and combinations of bands. Average scores (across four listeners) as well as the minimum and maximum scores achieved by the listeners are reported.

Consonant	Band(s)	Minimum score	Average score	Maximum score
I	B	42.4	52.1	57.5
I	C	46.7	52.7	56.6
I	D	40.0	45.3	55.3
I	E	48.2	52.8	59.1
I	B+C	65.8	70.7	75.0
I	C+D	64.9	68.9	74.4
I	D+E	61.1	66.6	76.0
I	B+D	72.2	77.4	85.0
I	B+E	72.6	77.7	86.9
I	C+E	70.1	74.4	80.4
I	B+C+D	79.7	85.3	89.4
I	C+D+E	76.2	80.8	84.5
I	B+C+E	84.7	88.3	93.2
I	B+D+E	82.2	87.2	93.5
I	B+C+D+E	86.3	90.0	93.9
F	B	39.7	47.1	55.6
F	C	52.7	55.4	56.9
F	D	41.5	47.3	56.8
F	E	50.3	57.1	63.0
F	B+C	61.8	67.0	71.2
F	C+D	62.1	67.0	74.1
F	D+E	57.4	62.7	72.5
F	B+D	67.0	72.6	82.2
F	B+E	70.3	74.7	83.1
F	C+E	71.4	74.6	81.1
F	B+C+D	76.7	80.8	88.6
F	C+D+E	74.1	78.2	86.7
F	B+C+E	78.4	83.9	90.4
F	B+D+E	80.0	84.3	90.2
F	B+C+D+E	83.0	86.6	92.3

substantially (88% for bands and 7% for listeners) to the total variance. Of the remaining factors, all except for the interaction between listeners and bands were significant, but none contributed more than 1% to the total variance.

D. Consistency across experiments

Several conditions were tested in more than one experiment. Specifically band G (D+E, 2100–4500 Hz) was tested in both experiments I and III while bands B, C, D, and E were tested in both experiments II and III. We analyzed scores from each of these two sets of experiments using the nonparametric Wilcoxson rank-sum test. The results of this analysis indicated that the (small) differences between experiments I and III were most likely due to chance. However, the differences between scores for final consonants under C, D, and E were very likely not due to chance (at the 0.05 level of significance). This difference is most likely due to the use of different listeners in experiments II and III.

E. Articulation index

The scores observed in all three experiments are plotted as a function of the estimated articulation index (AI) in Fig. 1. In each case we have superimposed a curve that represents a hypothesized relation between AI and the identification score. According to the assumptions that underlie the calculation of the articulation index (AI), the intelligibility score for a given presentation condition is related to the AI for that condition by a function that depends on the talker, the materials, and the listener, but is independent of the condition. Although this function can be determined empirically (e.g., by measuring identification performance systematically as a function of signal-to-noise ratio), we chose to employ the function suggested by Fletcher and Galt (1950),

$$S = A + Be^{-C\times \mathrm{AI}}, \tag{2}$$

where A, B, and C are fitting parameters whose values were derived from the observed scores. The sum $A+B$ was constrained to equal 8.3% $(100/N)$, the number of items to be identified so that $S=100/N$ when $\mathrm{AI}=0$. Effectively two free parameters were used to fit the observed scores. The values of AI plotted and used to estimate the free parameters were computed from the third-octave band intelligibility weights for nonsense syllables given by Pavlovic (1987) for the various bands and combinations of bands. The best fitting values of parameters A and C together with the rms error (E) of the fit of the curves are listed in Fig. 1. The results indicate that the smallest values of E were obtained for experiment III and the largest for experiment I. For experiments I and III, but not for experiment II, the values of parameters A and C, and also of the rms error E, was very similar for initial and final consonants. It should be noted that the values of E that we found were roughly comparable to those found by Steeneken and Houtgast (1999) who found rms errors of 8.8 for female speech and 12.8 for male speech. We were unable to apply Steeneken and Houtgast's (1999) correction to the articulation index values because of differences in bandwidths.

Clearly, although the values of these parameters are reasonably consistent, the predicted curves fit the data only moderately well. In particular, the results for experiment II are puzzling. It is unclear why the performance in the case of band A+B, which had the highest predicted value of AI, is relatively so poor, or why performance on bands D and E, which had the lowest predicted values of AI, is relatively so good. The results for experiment III are both more orderly and better fit by the hypothesized curve than the results for experiments I and II.

IV. INTEGRATION MODELS

The success of calculations based on the articulation index, such as those discussed in Sec. III, suggests that it may be possible to use measures of performance in single-band conditions to predict scores in combined-band conditions for individual listeners without the need for auxilliary experimentally determined parameters, such as the band importance weights in the articulation index computation, or derived fitting constants. In this section we describe five models of the process by which cues are integrated across frequency bands that can make the desired predictions.

All of the models considered make predictions based on observed confusion matrices. Let $P_K(R_j|S_i)$ denote the probability that response R_j is made when the stimulus S_i is presented via band K and $P_K(R|S)$ denote the confusion matrix for band K. In all cases we use the relative frequency of

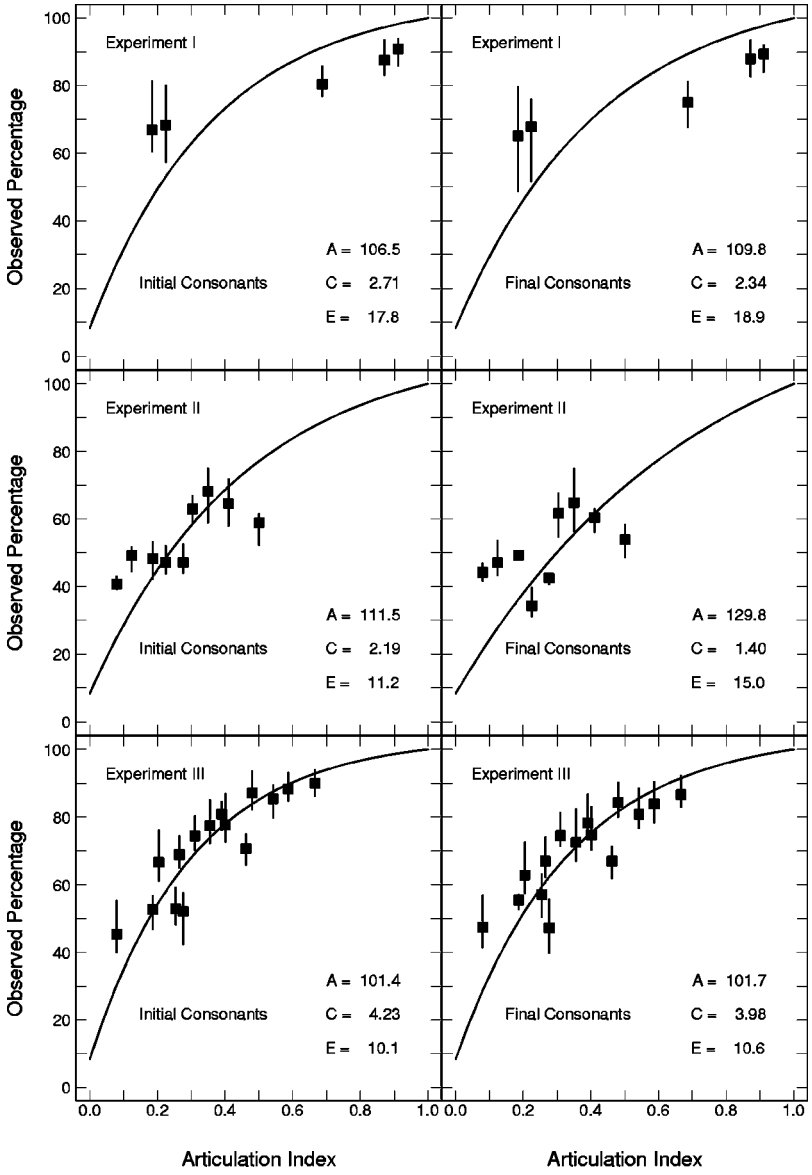


FIG. 1. Plot of scores, averaged across listeners, as a function of the estimated value of the articulation index (AI). Symbols indicate the scores, observed in experiment I (top row), II (middle row), and III (bottom row) for initial (left column) and final (right column) consonants plotted versus the AI for the band(s). Bars indicate the ranges of scores across listeners. For experiment I, data points corresponding to the contralateral and dichotic conditions are indicated by unfilled symbols which have been displaced leftward by 0.02 units, data points corresponding to the homolateral and monaural conditions are indicated by filled symbols which have been displaced rightward by 0.02 units. Each panel reports the values of the fitting parameters *A* and *C*, as well as the rms fitting error, *E*.

observing response R_j when stimulus S_i is presented to estimate $P(R_j|S_i)$. Given the confusion matrices for bands *K* and *L*, the models make predictions for the confusion matrix when these bands are presented simultaneously, $P_{KL}(R|S)$. We estimate the predicted probability of a correct response in the combined-band condition as C_{KL}, where

$$C_{KL}=\frac{1}{N}\sum_{i=1}^{N} P_{KL}(R_i|S_i). \tag{3}$$

A. Multiplicative error probabilities

The simplest band combination rule is one used by Fletcher (1953) in the analysis of ''Articulation Tests:'' a response error is made in the combined-band condition if and only if an error is made in each of the component band conditions. Fletcher appears to have applied this method to predicting the overall probability of error:

$$C_{KL}=1-(1-C_K)\times(1-C_L). \tag{4}$$

We denote this method Flet1. It is also possible to apply the rule separately to each of the *N* consonants to be identified,

$$P_{KL}(R_i|S_i)=1-(1-P_K(R_i|S_i))\times(1-P_L(R_i|S_i)), \tag{5}$$

and then use Eq. (3) to predict C_{KL}. We denote this method Flet2.

B. Post-labelling integration

According to the Post-labelling integration model (Braida, 1991), the listener is assumed to make separate identification judgments about the stimulus based on the cues available in each band, and to combine these judgments to determine the response to the multiband stimulus. In the KL condition, for example, presentation of stimulus S_i generates a pair of labels $(\mathcal{L}_m,\mathcal{L}_n)$ corresponding respectively to bands K and L. The labels correspond to the responses that would be given in single-band experiments. The collection of label pairs is assumed to be divided into mutually exclusive and collectively exhaustive sets. In particular, all labels belonging to the set $\mathcal{U}_j$ elicit response R_j so that

$$P_{KL}(R_j|S_i)=\sum_{m,n\in\mathcal{U}_j} P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i), \tag{6}$$

where $P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i)$ is the probability that stimulus S_i gives rise to the label pair $(\mathcal{L}_m,\mathcal{L}_n)$ and $P_{KL}(R_j|S_i)$ is the probability that response R_j is given to this stimulus when bands K and L are presented together.

To relate multiband identification to single-band identification, we assume that the label produced for each band is independent of the labels in the other bands. Labelling judgments for each band are thus assumed to be made by the same processes that operate in single-band presentation conditions:

$$P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i)=P^*_K(\mathcal{L}_m|S_i)\times P^*_L(\mathcal{L}_n|S_i). \tag{7}$$

In particular, we assume that the probability of assigning labels is described by the confusion matrices observed in the single-band presentation conditions. For example, when band K is presented,

$$P^*_K(\mathcal{L}_m|S_i)=P_K(R_m|S_i), \tag{8}$$

and when band L is presented,

$$P^*_L(\mathcal{L}_n|S_i)=P_L(R_n|S_i), \tag{9}$$

where $P_K(R_m|S_i)$ and $P_L(R_n|S_i)$ determine the frequencies observed in the single-band confusion matrices.

There are many possible rules (or equivalently many possible choices of the $\mathcal{U}_j$ sets) that the listener can use to map the label pair $(\mathcal{L}_m,\mathcal{L}_n)$ into a response. When the *a priori* stimulus presentation probabilities of the stimuli are equal, it is well known (e.g., van Trees, 1968) that no rule leads to greater identification accuracy (probability of responding correctly) than the *maximum likelihood* rule: for each $(\mathcal{L}_m,\mathcal{L}_n)$ the response should be the identity of the stimulus for which $P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i)$ is greatest.

Let $\mathcal{T}_j$ denote the set of (m,n) for which $P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_j)>P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i)$ for all $i\neq j$. Then, combining the above assumptions, the conditional response probabilities for the maximum likelihood rule are specified by

$$P_{KL}(R_j|S_i)=\sum_{m,n\in\mathcal{T}_j} P_K(R_m|S_i)\times P_L(R_n|S_i). \tag{10}$$

If there is a set of $(\mathcal{L}_m,\mathcal{L}_n)$ for which

$$P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_i)=P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_j)>P^*_{KL}(\mathcal{L}_m,\mathcal{L}_n|S_k) \tag{11}$$

for all $k\neq i$, j, then these $(\mathcal{L}_m,\mathcal{L}_n)$ may be mapped arbitrarily (on a deterministic or probabilistic basis) to either R_i or R_j without affecting identification accuracy. If there is no such set, the mapping that leads to greatest identification accuracy is unique. We denote this rule as PostLI.

C. Fuzzy logical model of perception (FLMP)

In the FLMP (Massaro, 1987), the response to each stimulus is determined, in a probabalistic fashion, by the ''feature value'' of that stimulus for each of the possible responses. In single-band presentations the feature value is estimated as the conditional probability, $P_K(R_j|S_i)$. In multiband conditions, the feature value is assumed to be proportional to the product of the feature values for the corresponding unimodal conditions (Cohen and Massaro, 1995). Thus $P_{KL}(R_j|S_i)$ is

$$P_{KL}(R_j|S_i)=\frac{P_K(R_j|S_i)P_L(R_j|S_i)}{\sum_{k=1}^{N}P_K(R_k|S_i)P_L(R_k|S_i)}. \tag{12}$$

[The denominator term in Eq. (12) normalizes the feature values for the N responses to conditional probabilities.] We denote this rule by FLMP.

D. Pre-labelling integration

According to the pre-labelling integration model (Braida, 1991), single-band sensory data are assumed to be represented in continuous valued cues that are combined optimally before labels are assigned. The statistical properties of the cues are inferred from singleband confusion matrices using a type of multidimensional scaling (Braida, 1988). Each presentation of one of the N possible consonants S_i is assumed to generate a D-dimensional vector of cues $\vec{X}=\langle x_1,x_2,\ldots,x_D\rangle$ described by the conditional probability density

$$p(\vec{X}|S_i)=\prod_{k=1}^{D} p(x_k|S_i)=\prod_{k=1}^{D}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x_k-s_{ik})^2}{2}\right). \tag{13}$$

Equivalently, the cue vector $\vec{X}$ may be thought of as displaced from the *stimulus center* $\vec{S}_i=\langle s_{i1},s_{i2},\ldots,s_{iD}\rangle$ by an additive noise vector whose components are independent Gaussian random variables with zero means and common unit variance.

Corresponding to each response there is a *response center* or *prototype* $\vec{R}_i=\langle r_{i1},r_{i2},\ldots,r_{iD}\rangle$. (It is *not* necessarily true that $\vec{R}_i=\vec{S}_i$.) The listener is assumed to respond R_k if and only if the distance from the observed vector of cues $\vec{X}$ to $\vec{R}_k$ is smaller than the distance to any other prototype, i.e., $|\vec{X}-\vec{R}_k|<|\vec{X}-\vec{R}_j|$ for all $j\neq k$. For any two vectors $\vec{Y}$ and $\vec{Z}$, the distance $|\vec{Y}-\vec{Z}|$ is defined in terms of the vector components y_k and z_k as

$$|\vec{Y}-\vec{Z}|=\sqrt{\sum_{k=1}^{D}(y_k-z_k)^2}. \tag{14}$$

In effect, this rule partitions the D-dimensional space of cue vectors into N compact ''response regions'' that are bounded by hyperplanes (see Fig. 2). All the points in each region are closer to the prototype in that region than to any other prototype. The probability that R_j is the response when the stimulus is S_i is the probability that the vector of cues $\vec{X}$ is closer to $\vec{R}_j$ than to any other prototype when S_i is presented. Since the density $p(\vec{X}|S_i)$ weights the relative likelihood of occurrence of cue vectors in these conditions, $P(R_j|S_i)$ is determined by integrating the density with mean $\vec{S}_i$ over the response region containing $\vec{R}_j$.

Thus the decision process in an identification experiment reflects the listener's comparison of cue vectors with prototypes. Errors result from the assumed noisiness of the cue vector and the distribution of error responses reflects the

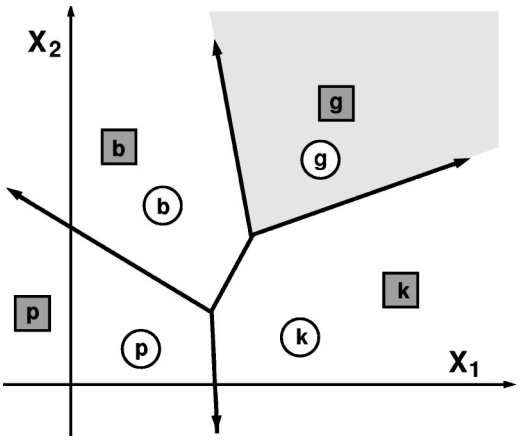


FIG. 2. Two-dimensional cue space for a hypothetical identification experiment with the four consonants /p,b,k,g/. Location of the stimulus centers (circles) relative to response prototypes (squares) is arbitrary in this figure. Response regions are bounded by line segments (extended bold lines), each of which coincides with a segment of the perpendicular bisector of a line connecting two response prototypes. All the points in a given response region are closer to the prototype in that region than to any other prototype. In general $P(\mathcal{R}_j|\mathcal{S}_i)$ is computed by integrating the multidimensional Gaussian density with mean $\vec{S}_i$ over the response region that contains $\vec{R}_j$. Thus the predicted probability of response "g" to stimulus "b" would be obtained by integrating the two-dimensional Gaussian density at stimulus center b over the shaded area which contains response center g.

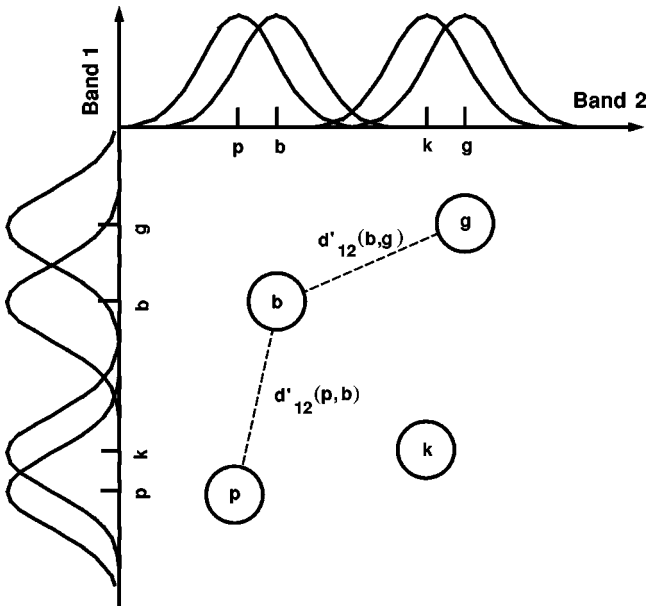


FIG. 3. Relation of single- and combined-band cue spaces for /p,b,k,g/ identification experiments. Confusions in the single-band conditions are each assumed to be adequately described by cues with scalar Gaussian densities having a common standard deviation. Stimulus centers in the combined-band condition are points in the plane. Since the band 1 and band 2 coordinates in the combined-band condition are assumed to be the same as the scalar coordinates in the single-band conditions, distances between points in the plane are just the Pythagorean sums of the single-band distances for the same stimulus pair. In all conditions the distance between a pair of points is the subject's sensitivity for distinguishing between the corresponding consonants.

relative locations of stimulus and response centers. The boundaries of the response regions are analogous to the decision criteria of signal detection theory. The locations of the prototypes, which determine these boundaries, are assumed to be at least partially under the control of the listener reflecting the effects of instructions and payoffs.

In the multiband presentation condition, we assume that cues are combined optimally and that there is no perceptual interference (e.g., masking or distraction) across bands. We model the integration of cues by assuming that the cue densities in the multiband condition are the "Cartesian products" of the densities corresponding to the separate bands (see Fig. 3). This space is constructed from the orthogonal composition of the cue spaces corresponding to each of the stimulus bands. A subset of the coordinates of each point in the product space corresponds to the coordinates of each of the subspaces for these bands. Consider, for example, a multiband condition. If the cue space corresponding to the band K condition has dimension D_K with each stimulus represented by a center $\vec{K}_i=\langle k_{i1},k_{i2},\ldots,k_{iD_K}\rangle$, while that for the band L condition has dimension D_L with each stimulus represented by a center $\vec{L}_i=\langle l_{i1},l_{i2},\ldots,l_{iD_L}\rangle$, then in the combined-band (KL) condition, the cue space has dimension $D_{KL}=D_K+D_L$ and each stimulus center $\vec{S}_i$ has the coordinates

$$\vec{S}_i=\langle s_{i1},s_{i2},\ldots,s_{iD_{KL}}\rangle=\langle k_{i1},\ldots,k_{iD_K},l_{i1},\ldots,l_{iD_L}\rangle. \quad (15)$$

Locations of the stimulus centers in the combined cue space for three hypothetical identification experiments involving the consonants /p,k,b,g/ are shown in Fig. 3. For simplicity, the figure assumes that confusions in the band K and band L conditions can satisfactorily be accounted for in one-dimensional cue spaces. The cue space in the KL combined-band condition is therefore predicted to be two-dimensional. This model for the combined condition predicts that there is a simple relation between a listener's sensitivity in distinguishing S_i from S_j in the multiband condition, $d'_{KL}(i,j)$, and the corresponding single-band sensitivities $d'_K(i,j)$ and $d'_L(i,j)$:

$$d'_{KL}(i,j)=\sqrt{d'_K(i,j)^2+d'_L(i,j)^2}. \quad (16)$$

According to the assumptions stated above, the configurations derived from the single-band confusion matrices determine the predicted stimulus centers for the multiband configurations. Predicting accuracy in the multiband condition requires a specification of the response centers as well. Unfortunately, this model does not predict the locations of the prototypes in the multiband case, since response bias is logically unconstrained by sensitivity. The locations of response centers used in an experiment (relative to fixed stimulus centers) can be estimated iteratively. However, we have chosen to compute accuracy in three special cases:

(A) The response centers in the multiband case

$$\vec{M}_i=\langle m_{i1},m_{i2},\ldots\rangle \quad (17)$$

are the Cartesian products of the response centers in the one band cases $m_{ij}=o_{ij}$. We denote this method PRELI0.

(B) The response centers in the multiband case coincide with the multiband stimulus centers $m_{ij}=s_{ij}$. We denote this method PRELIH.

(C) The response centers in the multiband case are halfway between the response centers in the one-band case and multiband stimulus centers $m_{ij}=(o_{ij}+s_{ij})/2$. We denote this method PRELI1.

Case A represents the case of minimal adjustment to the multiband presentation condition: the response centers are located at the same locations they had in the single-band conditions.

Case B, the coincident center condition, has two special properties. First, when the *a priori* stimulus presentation probabilities are equal, as they usually are in consonant identification experiments, it leads to the maximum value of identification accuracy among all choices of response center locations (van Trees, 1968). This facilitates direct comparisons between the models of pre- and post-labelling integration. Second, because this condition leads to a maximum value of overall accuracy, small deviations in the response center locations from the stimulus center locations produce only second-order changes in overall accuracy (although they produce first-order changes in the conditional response probabilities). Thus if listeners use response center locations that are close to the coincident locations, observed accuracy should be very close to that predicted by the method.

Case C represents an intermediate case of adjustment to the multiband presentation condition: the response centers are halfway between the response centers of case A and those of case B.

Note that when the single-band response centers are located at the positions of the single-band stimulus centers, then all three cases have the same locations for the multiband response centers.

V. MODEL PREDICTIONS

Although the different models of integration discussed in Sec. IV are capable of making predictions for scores in combined-band conditions on the basis of performance in the component bands, they base these predictions on different aspects of component-band performance. To compare the models in a consistent fashion, predictions of scores in the combined-band conditions were computed from the confusion matrices observed in the single-band conditions. These predictions were computed by assuming that the observed confusion matrices for the single-band conditions gave precise estimates of the conditional probabilites [e.g., $P_K(R_j|S_i)$ and $P_L(R_j|S_i)$] which were then used to predict the probability of correct responses, C_{KL}, or equivalently the score $S_{KL}=100C_{KL}$, in the combined-band conditions. In applying this model to the experimental data, we estimate the $P_K(R_j|S_i)$ and $P_L(R_j|S_i)$ directly from the single-band matrices, and use these to compute predicted values of $P_{KL}(R_j|S_i)$ by means of Eq. (12). Massaro and Cohen (1995) have argued that this procedure may tend to underestimate the extent to which the FLMP is capable of fitting the data. Nevertheless, we use this approach because it is most similar to the approaches used to obtain predictions for the other models considered.

Predictions were made separately for each listener, combined-band condition, and initial/final consonants in experiments I, II, and III. In the case of the pre-labelling model, single-band confusions were fit with $D=4$ dimensions using procedures described by Braida (1988) to estimate the locations of stimulus centers.

Although filter band and listener were often statistically significant factors, their interaction accounted for only a small fraction of the total variance. Thus we evaluated the adequacy of the predictions for a given set of listeners in terms of average bias and also in terms of chi-squared, computed over listeners.

The bias was computed as the average difference between predicted score and the observed score for listener *i*,

$$B=\frac{1}{N_L}\sum_{i=1}^{N_L} S_{Pi}-S_{Oi},$$

where N_L is the number of listeners, S_{Oi} is the observed score (in percent of items correct) for listener *i*, and S_{Pi} is the predicted score. Thus positive values of bias correspond to predicted scores that are higher than actual scores, negative values to predicted scores that are lower than actual scores.

The value of chi-squared was computed as the sum

$$X^2=\sum_{i=1}^{N_L}\frac{N_i}{100}\frac{(S_{Pi}-S_{Oi})^2}{S_{Pi}},$$

where N_i is the total number of entries in the confusion matrix for observer *i*. The chi-squared test was applied in order to determine cases for which a given model did not make an accurate prediction of the scores obtained by the listeners. Values of chi-squared which are in excess of a critical value (corresponding to the 0.05 level of significance) indicate that predictions were relatively poor, independent of the prediction bias. This use of the chi-squared test is similar to that employed by Braida (1991) in evaluating the ability of models to account for audiovisual integration.

A. Experiment I

Predictions were made separately for the monaural and dichotic presentation conditions. The results, summarized in terms of average bias and chi-squared (computed across listeners), are shown in Tables IV and V. Overall the Flet1 and PreLI1 prediction methods overestimated observed performance when bands were combined, whereas the PostLI and PreLI0 methods underestimated performance. No methods made predictions that exceeded the chi-squared value corresponding to the 0.05 level of significance. Consequently, we conclude that all methods made relatively satisfactory predictions of the scores.

B. Experiment II

Predictions were made separately for each band combination. The results are shown in Tables VI and VII. For both initial and final consonants, the Flet1, FLMP, and PostLI prediction methods overpredicted observed performance while the Flet2 and PreLI0 methods underpredicted performance when two bands were combined. The bias was generally much less for the PreLIH method.

The PreLI0, and PreLI1 methods made predictions that exceeded the chi-squared value corresponding to the 0.05 level of significance for all band combinations for both initial

TABLE IV. Bias (B) of predicted identification scores for initial (I) and final (F) consonants in experiment I for the combined of bands in the monaural (M) and dichotic (D) presentation conditions of experiment I.

Consonant	Bands	Presentation condition	Model						
			Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	F+G	M	2.70	−2.32	−3.80	−0.81	−5.80	1.12	4.54
I	F+G	D	5.68	0.55	−1.37	2.17	−3.46	3.80	7.44
F	F+G	M	2.46	−1.90	−2.16	−0.64	−5.14	2.62	5.30
F	F+G	D	3.18	−1.70	−1.27	0.46	−4.62	3.46	6.22

and final consonants. By contrast, the FLMP method did not exceed the critical value on six band initial–final combinations, the Flet2 on seven, and PreLIH method on all eight band combinations. We conclude that the FLMP, the Flet2, and the PreLIH methods made relatively satisfactory predictions of the scores. For the Flet1, PreLI0, and PreLI1 methods instances of high chi-squared were generally associated with instances in which the predictions exhibited substantial bias.

C. Experiment III

Predictions were made separately for each band combination. The results are shown in Tables VIII and IX. No completely consistent pattern of average bias was observed for any method, although the Flet2 and PreLI0 methods generally underpredicted performance and the Flet1 and PreLI1 methods generally overpredicted performance.

The chi-squared tests indicate that none of the methods make completely satisfactory predictions. All methods made some predictions that exceeded the chi-squared value corresponding to the 0.05 level of significance, and no method satisfied that test for all band combinations on either initial or final consonants. The Flet2 and PreLI0 methods failed the test on 16 and 18 (out of a possible 22) combinations, while the FLMP and PreLIH method failed the test on 5 and 6 combinations and the other methods failed the test 9–11 times.

It is noteworthy that there is no general pattern that relates the performance of a prediction method on the chi-squared test and the band combination. For the FLMP, for example, the 0.05 criterion was exceeded on the band combination B+C for both initial and final consonants, and exceeded on band combinations C+D and D+E for the final consonants. For the PreLIH, on the other hand, the band combinations for which the chi-squared criterion value was exceeded were different for initial and final consonants. Both the FLMP and the PreLIH methods exceeded the criterion value in the predictions for band combinations C+D and D+E, however. It is noteworthy that the predictions of the FLMP and PreLIH methods for these latter two band combinations were biased: average predicted scores exceeded observed scores by 7–12 points. For the FLMP, instances of high chi-squared were generally associated with instances in which the method substantially (by 6–12 points) overestimated performance. The same also tended to be true for the PreLIH method.

VI. DISCUSSION

A. Experiments

The experiments reported in this paper provide new data relevant to the identification of consonants on the basis of acoustic cues conveyed in distinct spectral bands. The data we have obtained also suggest that there may be substantial individual differences in consonant identification when low-frequency spectral components are involved.

It is natural to question whether combined band performance improves as the separation between frequency bands increases. One might expect that as the separation increased, bands would provide increasingly different cues for consonant identification. The correlation of confusion matrix entries provides a crude measure of the redundancy of adjacent bands. For experiment II, it was found to decrease monotonically from 0.89 for bands A and B to 0.48 for bands A and E. For experiment III, the decline was somewhat less systematic, 0.78 for bands B and C, and 0.58 for bands B and D, but 0.62 for bands B and E. The difference between the combined band scores and the greater of the individual band scores was negatively correlated with the correlation between the individual-band confusion matrices for the final consonants of experiment II ($\rho=-0.71, p<0.05$) but only marginally so for the initial consonants ($\rho=-0.48, p$

TABLE V. Chi-squared values per degree of freedom for predicted identification scores for initial (I) and final (F) consonants in experiment I for the combined bands, in the monaural (M) and dichotic (D) presentation conditions. All entries have been divided by 5, the number of degrees of freedom. Cases for which the chi-squared test was statistically significant at the 0.05 level, of which there are none in this table, are marked with an asterisk.

Consonant	Bands	Presentation condition	Model						
			Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	F+G	M	0.38	0.36	0.73	0.27	1.32	0.08	0.87
I	F+G	D	1.09	0.16	0.32	0.34	0.54	0.53	2.00
F	F+G	M	0.28	0.30	0.43	0.09	1.26	0.33	0.98
F	F+G	D	0.53	0.38	0.27	0.08	0.91	0.49	1.50

TABLE VI. Bias (B) of the predicted identification scores for initial (I) and final (F) consonants in experiment II for various combinations of bands.

		Model						
Consonant	Bands	Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	A+B	12.40	−1.77	−3.45	3.62	−6.20	2.32	10.27
I	A+C	8.02	−3.94	−5.02	6.32	−9.70	0.27	6.42
I	A+D	6.07	−2.99	−3.62	5.81	−9.08	0.70	8.37
I	A+E	4.27	−2.99	−1.37	5.73	−8.55	1.93	8.57
F	A+B	9.64	−1.44	−5.61	4.23	−9.05	0.33	7.70
F	A+C	6.18	−1.67	−9.68	2.46	−11.05	0.20	8.33
F	A+D	2.10	−5.06	−9.31	0.04	−12.32	−0.30	6.58
F	A+E	2.47	−2.25	−4.92	1.72	−10.50	−0.28	6.25

>0.05) and for both the initial and final consonants of experiment III ($\rho=-0.68$, and $\rho=-0.64$, $p<0.05$). Thus the improvement scores associated with combining bands appears to be inversely related to the correlation between the individual-band confusion matrices.

Based on the data in Tables II and III, it would appear that, as the separation between frequency bands increases, performance in the combined band conditions improves. This is approximately true, but the improvements in score that are associated with increasing frequency separation are not unambiguous. Overall performance in the single-band conditions varies. For example, the final-consonant scores in the combined-band conditions of experiment II were significantly correlated with the individual band scores ($\rho=0.68, p<0.05$) but there was no correlation for initial consonant scores. For experiment III, there was no significant correlation for initial or final consonant scores for the combinations of pairs of bands that involved band B, but there was for the pairs that involved band C ($\rho=0.76, p<0.05$).

Reduced performance in combined-band presentation conditions is often seen in elderly listeners and in listeners with disorders of the central auditory system. Matzger (1958) reported that elderly listeners with no diagnosed central pathology are much less capable of dichotic across-band integration than younger listeners. Palva and Jokinen (1975) also report a decline in dichotic integration after age 60. Haas (1982) has pointed out that since nearly the same reduction is seen when diotic presentation is used, the integration problem for elderly listeners may be evident even in monaural situations. Of course, this reduced performance may also reflect poor reception of one or both of the separate bands. The models we are developing allow these factors to be separated because combined-band predictions are based on measured confusions in single-band presentation conditions for the same listener.

B. Models

We have discussed seven model-based methods that are potentially applicable to the problem of predicting scores in combined-band conditions on the basis of descriptions of performance in component-band conditions. Three of these models (the FLMP, and the pre- and post-labelling models) have previously been applied to the problem of predicting audiovisual consonant-identification scores on the basis of descriptions of performance in separate audio and visual identification experiments. Although the other two models (Flet1 and Flet2) were not applied to the audiovisual integration problem, it is unlikely that they would have provided good accounts of audiovisual scores. In related work Braida (1993) showed that the pre-labelling model provides a better account than the post-labelling model (though perhaps no better than Flet1) of data obtained by Miller and Nicely (1955) on the integration of consonants subject to lowpass and highpass filtering. Moreover, he showed that both of these models can be extended to the problem of predicting the effects of varying the signal-to-noise ratio by treating increases in signal-to-noise ratio as equivalent to multiple

TABLE VII. Chi-squared values per degree of freedom for predicted identification scores for initial (I) and final (F) consonants in experiment II for various combinations of bands. All entries have been divided by 4, the number of degrees of freedom. Cases for which the chi squared test was statistically significant at the 0.05 level are marked with an asterisk.

		Model						
Consonant	Bands	Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	A+B	9.06*	1.50	1.57	1.16	3.71*	1.13	6.93*
I	A+C	3.71*	1.67	2.78*	2.59*	7.00*	0.57	2.71*
I	A+D	2.59*	1.05	1.36	2.17	7.87*	0.54	4.30*
I	A+E	2.65*	1.51	2.45*	3.59*	6.60*	1.23	5.21*
F	A+B	6.51*	0.59	4.52*	1.81	8.58*	0.23	4.38*
F	A+C	2.31	0.67	8.20*	0.43	10.34*	0.18	4.05*
F	A+D	0.86	2.70*	7.30*	0.20	12.61*	0.16	2.94*
F	A+E	1.62	0.96	2.20	1.33	8.26*	0.06	2.55*

TABLE VIII. Bias (B) of the predicted consonant identification scores for initial (I) and final (F) consonants in experiment III for various combinations of bands.

		Model						
Consonant	Bands	Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	B+C	6.60	−6.80	−2.17	6.49	−9.06	1.84	7.82
I	C+D	4.95	−7.42	−2.79	1.14	−10.10	1.70	7.90
I	D+E	7.78	−4.42	−1.15	1.66	−7.86	5.34	12.74
I	B+D	−3.54	−15.09	−11.33	−2.07	−18.90	−6.92	−0.38
I	B+E	−0.56	−10.68	−4.73	0.52	−11.78	−0.56	5.72
I	C+E	2.72	−8.12	−2.06	−0.97	−8.54	2.48	8.38
I	B+C+D	2.17	−13.71	−10.98	0.11	−19.72	−7.40	−0.42
I	C+D+E	7.30	−7.67	−6.43	0.24	−14.84	−1.36	6.56
I	B+C+E	1.20	−12.33	−9.09	−1.13	−16.32	−5.88	0.78
I	B+D+E	0.06	−14.36	−14.70	−2.51	−19.64	−7.04	0.24
I	B+C+D+E	3.84	−11.11	−13.92	−0.50	−17.84	−6.28	1.26
F	B+C	9.32	−3.26	−1.77	7.06	−7.52	2.68	8.74
F	C+D	9.56	−0.75	2.95	7.36	−3.66	7.64	12.30
F	D+E	13.93	0.73	8.99	12.08	0.88	11.62	16.62
F	B+D	−0.03	−12.02	−7.05	0.82	−15.48	−4.72	1.54
F	B+E	2.83	−7.28	−0.75	3.35	−9.28	0.04	5.94
F	C+E	5.48	−5.71	6.39	5.34	−2.34	5.70	9.30
F	B+C+D	7.29	−7.26	−6.30	5.10	−16.04	−2.54	3.92
F	C+D+E	12.02	−0.99	5.82	9.81	−3.58	6.22	10.50
F	B+C+E	6.23	−6.42	−2.03	5.15	−10.92	−0.64	4.76
F	B+D+E	2.75	−11.74	−4.42	2.17	−15.64	−4.38	1.58
F	B+C+D+E	7.96	−5.45	−2.20	6.12	−11.78	−0.36	4.78

independent samples of noisy cues. Both the pre- and post-labelling models are capable of accounting for the effects of changing signal-to-noise ratio over more than a 20-dB range.

The models differ markedly in structure. Flet1 is based only on overall scores in component bands, Flet2 is based on scores for each vowel in component bands, while the FLMP and the pre-Labelling and post-labelling models are based on the detailed patterns of confusion made in the component bands. The models correspondingly differ in level of detail available in their predictions (overall scores versus residual

TABLE IX. Chi-squared values per degree of freedom for predicted identification scores for initial (I) and final (F) consonants in experiment III for various combinations of bands. All entries have been divided by 5, the number of degrees of freedom. Cases for which the chi-squared test was statistically significant at the 0.05 level are marked with an asterisk.

		Model						
Consonant	Bands	Flet1	Flet2	PostLI	FLMP	PreLI0	PreLIH	PreLI1
I	B+C	3.00*	4.06*	0.93	2.82*	6.42*	0.50	3.64*
I	C+D	1.82	4.49*	0.87	0.39	7.92*	0.39	3.62*
I	D+E	3.84*	2.08	0.27	0.53	4.88*	1.88	8.72*
I	B+D	0.87	15.62*	8.70*	0.55	26.34*	3.07*	0.22
I	B+E	0.83	7.72*	1.98	0.59	10.01*	0.61	2.52*
I	C+E	0.70	4.22*	0.46	0.53	4.92*	0.71	3.60*
I	B+C+D	0.42	11.36*	7.52*	0.23	26.04*	3.34*	0.28
I	C+D+E	2.77*	3.79*	2.82*	0.39	14.03*	0.53	2.40*
I	B+C+E	0.42	9.03*	5.32*	0.30	16.51*	2.03	0.56
I	B+D+E	0.44	12.75*	14.23*	0.66	25.21*	3.19*	0.92
I	B+C+D+E	0.83	6.86*	12.01*	0.19	19.35*	2.24*	0.57
F	B+C	5.51*	1.50	0.79	3.65*	4.72*	1.24	5.03*
F	C+D	5.47*	0.43	1.11	3.43*	1.64	3.71*	8.46*
F	D+E	10.70*	0.35	5.40*	8.37*	0.43	7.63*	14.65*
F	B+D	0.26	10.78*	3.77*	0.16	18.07*	1.86	0.68
F	B+E	0.58	3.61*	0.44	0.70	5.48*	0.18	2.12
F	C+E	1.74	2.22*	2.21	1.54	0.59	1.64	4.14*
F	B+C+D	3.28*	3.89*	3.34*	1.85	17.20*	0.62	1.16
F	C+D+E	7.18*	0.76	2.08	4.88*	1.16	2.11	5.43*
F	B+C+E	2.47*	2.92*	0.67	1.72	7.02*	0.29	1.51
F	B+D+E	0.84	8.97*	1.35	0.57	15.71*	1.75	0.80
F	B+C+D+E	3.27*	1.90	0.58	2.05	7.88*	0.17	1.33

confusion patterns). However, all the models assume that the cues are statistically independent from band to band, that cues are combined without interference, and that no new cues arise from interband comparisons. Within the confines of these assumptions, the models differ with respect to how they characterize the cues available in component bands and in how they describe the process by which cues from component bands are combined. Since these models make a range of predictions for a given combination of bands, they provide bounds on the range of combined-band scores over which their common assumptions cannot be rejected.

The fact that the cues are assumed to be statistically independent from band to band implies that, in general, the combined-band score exceeds the larger of the individual-band scores. Even when the two single-band confusion matrices are the same, the combined-band score would generally be expected to exceed the single band score, except for special cases, such as when all responses in the single band matrix are correct, or when all responses have an equal (uniform) probability. In all cases when a single band was combined with itself, the predicted score exceeded the single band score. In addition, the predicted combined band score for pairs of bands exceeded the maximum of the single band scores. Braida (1993) provides examples of the difference in predictions made by the pre- and post-labelling models between the cases of completely redundant cues (identical confusion matrices) and complementary cues. In each case, performance is predicted to be greater in the case of complementary cues.

We evaluated the quality of the predictions made by the seven methods in terms of bias and chi-squared. Two of the methods, the FLMP and the PreLIH, seemed to provide the best overall predictions. For experiment I, FLMP predictions were less biased than those of PreLIH. The reverse was generally true for experiment II, while for experiment III biases were roughly equal. When evaluated in terms of chi-squared, both the FLMP and the PreLIH methods made predictions that fared about equally.

In general the FLMP and PreLIH methods made better predictions for experiments I and II, which included the 0–700 Hz band, than for experiment III, which excluded this band. It did not prove to be any more difficult to make satisfactory predictions for bands that were widely separated than for bands that were close in frequency.

Although none of the methods made completely satisfactory predictions, we are not overly discouraged by this result. For every band combination and for both initial and final consonants, at least two of the models made predictions which were relatively unbiased and did not exceed the critical chi-squared value.

By contrast with the results we obtained when applying the models to across-band integration tasks, Braida (1991) reported that the pre-labelling model (the equivalent of PreLI1) made relatively good predictions of scores in tasks involving audiovisual integration. This difference in predictive abilities is difficult to interpret because there are many differences between the tasks considered. Audiovisual predictions were based on group data rather than on responses obtained from individual listeners. However, the integration tasks also differ with respect to the opportunities for perceptual interference (at least at the peripheral level) and also with respect to the structure of perceptual confusions in the component conditions: error patterns with the structure characterized by visemes are rarely seen in the confusion matrices for bandpass-filtered consonants.

We plan to continue to develop and evaluate models for the integration of cues across frequency bands. In future work it will be important to determine whether different models are required depending on whether cues from the 0–700 Hz band of speech are made available to the listener. Unfortunately our experiments do not allow us to separate this question from possible differences in listeners and speech materials. It is also important to understand why the different models had more or less success in making predictions for consonants in initial and final position in CVC syllables. More generally, the models need to be compared in terms of their ability to make predictions of combined-band performance that are more detailed than overall score, such as detailed patterns of errors.

C. Concluding remarks

Although we have investigated a substantial number of band combinations, there are obviously many more that need to be studied if the models described in Sec. IV are to be investigated rigorously. Of particular interest in future work are band combinations consisting of multiple separated bandpass conditions such as those investigated by Kryter (1960) and Warren *et al.* (1995).

The models we have investigated may provide new means for identifying interband interference effects in speech reception. At present such effects are identified operationally by *ad hoc* techniques, such as noting when the addition of a band reduces an intelligibility score, or when greater improvements are obtained for dichotic than diotic band combinations. Such criteria are likely to be too lax. Interference may be operative whenever the listener combines the cues from the component bands in a less than optimal fashion.

ACKNOWLEDGMENTS

This work was supported by the N. I. H. (Grant R01-DC00117). We are grateful for thoughtful discussions with J. B. Allen, P. Duchnowski, A. Oxenham, W. M. Rabinowitz, and C. M. Reed. We are grateful for the comments and criticisms of the Associate Editor and the anonymous reviewers. We are also grateful to M. W. Connell, J. Greenberg, D. S. Lum, and P. Milner for assistance with the software used to process the stimuli and conduct the experiments. Readers interested in obtaining the data matrices obtained as part of the research described should contact the corresponding author.

ANSI (**1969**). *Methods for the Calculation of the Articulation Index, ANSI S3.5-1969* (American National Standards Institute, New York).

ANSI (**1997**). *Methods for the Calculation of the Speech Intelligibility Index, ANSI S3.5-1997* (American National Standards Institute, New York).

Braida, L. D. (**1988**). "Development of a model for multidimensional identification experiments," J. Acoust. Soc. Am. **84**, S142.

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. A **43**(3), 647–677.

Braida, L. D. (**1993**). ''Integration Models of Intelligibility,'' in *Proc. 1993 N.A.S.-C.H.A.B.A. Symp. on Speech Communication Metrics and Human Performance*, Washington, DC, 3–4 June, 1993.
Breeuwer, M., and Plomp, R. (**1984**). ''Speechreading supplemented with frequency-selective sound-pressure information,'' J. Acoust. Soc. Am. **76**, 686–691.
Bustamante, D. K., and Braida, L. D. (**1987**). ''Principal component amplitude compression for the hearing impaired,'' J. Acoust. Soc. Am. **82**, 1227–1242.
Cohen, M. M., and Massaro, D. W. (**1995**). ''Perceiving visual and auditory information in consonant-vowel and vowel syllables,'' in *Levels in Speech Communication: Relations and Interactions: A Tribute to Max Wajskop*, edited by C. Sorin, H. Meloni, and J. Schoentgen (Elsevier Science, Amsterdam).
De Gennaro, S., Braida, L. D., and Durlach, N. I. (**1986**). ''Multichannel Syllabic Compression for Severely Impaired Listeners,'' J. Rehabil. Res. Dev. **23**, 17–24.
Fletcher, H. M. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York).
Fletcher, H. M., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.
Franklin, B. (**1969**). ''The effect on consonant discrimination of combining a low-frequency passband in one ear and a high-frequency passband in the other ear,'' J. Aud Res. **9**, 365–378.
Franklin, B. (**1975**). ''A comparison of the effect on consonant discrimination of combining low- and high-frequency passbands on consonant recognition in the hearing impaired,'' J. Speech Hear. Res. **18**, 719–727.
Franklin, B. (**1979**). ''A comparison of the effect on consonant discrimination of combining low- and a high-frequency passbands in normal, congenital, and adventitious hearing-impaired subjects,'' J. Am. Audiol Soc. **5**, 168–176.
Goshorn, E. L., and Studebaker, G. A. (**1984**). ''Effects of Intensity on Identification of Filtered Nonsense Syllables,'' ASHA-84.
Grant, K. W., and Braida, L. D. (**1991**). ''Evaluating the articulation index for auditory-visual input,'' J. Acoust. Soc. Am. **89**, 2952–2960.
Grant, K. W., Braida, L. D., and Renn, R. J. (**1993**). ''Auditory supplements to speechreading: Combining amplitude envelope cues from different spectral regions of speech,'' J. Acoust. Soc. Am. **95**, 1065–1073.
Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).
Haas, G. (**1982**). ''Impaired Listeners' Recognition of Speech Presented Dichotically through High- and Lowpass Filters,'' Audiology **21**, 433–453.
Hall, J. W., and Grose, J. H. (**1990**). ''Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,'' J. Acoust. Soc. Am. **87**, 269–283.
Kasturi, K., Loizou, P. C., Dorman, M., and Spahr, T. (**2002**). ''The intelligibility of speech with holes in the spectrum,'' J. Acoust. Soc. Am. **112**, 1102–1111.
Kryter, K. D. (**1960**). ''Bandwidth compression through spectrum selection,'' J. Acoust. Soc. Am. **32**, 547–556.
Lim, J. S., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1977**). ''Intensity perception VIII. Loudness comparisons between different types of stimuli,'' J. Acoust. Soc. Am. **62**, 1256–1267.
Lippmann, R. P., Braida, L. D., and Durlach, N. I. (**1981**). ''Study of multichannel amplitude compression for persons with sensorineural hearing loss,'' J. Acoust. Soc. Am. **69**, 524–534.
Loizou, P. C., Mani, A., and Dorman, M. F. (**2003**). ''Dichotic speech recognition in noise using reduced spectral cues,'' J. Acoust. Soc. Am. **114**(1), 475–483.
Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Earlbaum, Hinsdale, NJ).
Massaro, D. W., and Cohen, M. M. (**1995**). ''Modelling the Perception of Bimodal Speech,'' Proc. Int. Congress of Phonetic Sciences, Stockholm Sweden, Vol. 3, pp. 106–113.
Matzker, J. (**1958**). *Ein binauraler Horsynthese-Test zum Nachweis zerebraler Horstungen* (Thieme, Stutgart).
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Müsch, H., and Buus, S. (**2001**). ''Using statistical decision theory to predict speech intelligibility. I. Model structure,'' J. Acoust. Soc. Am. **109**, 2896–2909.
Palva, A. K., and Jokinen, K. (**1975**). ''Role of the binaural test in filtered speech audiometry,'' Acta Oto-Laryngol. **79**, 310–313.
Pavlovic, C. V. (**1987**). ''Derivation of primary parameters and procedures for use in speech intelligibility predictions,'' J. Acoust. Soc. Am. **82**, 413–422.
Rosengard, P. S., Payton, K. L., and Braida, L. D. (**2004**). ''Effect of slow-acting wide dynamic range compression on measures of intelligibility and ratings of speech quality in simulated-loss listeners,'' J. Speech Hearing Res. (in press).
Shannon, R. V., Galvin, III, J. J., and Baskent, D. (**2001**). ''Holes in Hearing,'' J. Assoc. Res. in Otolaryngology **3**, 185–199.
Shannon, R. V., Zeng, F. G., and Wygonski, J. (**1998**). ''Speech recognition with altered spectral distribution of envelope cues,'' J. Acoust. Soc. Am. **104**, 2467–2476.
Sheft, S., and Yost, W. A. (**1992**). ''Spectral transposition of envelope modulation,'' J. Acoust. Soc. Am. **91**, 2333.
Steenecken, H. J. M., and Houtgast, T. (**1980**). ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **69**, 318–326.
Steenecken, H. J. M., and Houtgast, T. (**1999**). ''Mutual dependence of the octave-band weight in predicting speech intelligibility,'' Speech Commun. **28**, 109–123.
Takeuchi, A., and Braida, L. D. (**1993**). ''Effect of frequency transposition on the discrimination of amplitude envelope patterns,'' J. Acoust. Soc. Am. **97**, 453–460.
van Trees, H. L. (**1968**). *Detection, Estimation, and Modulation Theory* (Wiley, New York).
Villchur, E. (**1973**). ''Signal processing to improve intelligibility in perceptive deafness,'' J. Acoust. Soc. Am. **53**, 1646–1657.
Wang, M., Reed, C., and Bilger, R. (**1978**). ''A comparison of the effects of filtering and sensorineural hearing loss on patterns of consonant confusions,'' J. Speech Hear. Res. **21**, 5–36.
Warren, R. M., Reiner, K. M., Blashford, J. J. A., and Brubaker, B. S. (**1995**). ''Spectral redundancy: Intelligibility of sentences heard through narrow spectral slits,'' Percept. Psychophys. **57**, 175–182.
Yost, W. A., and Sheft, S. (**1990**). ''Modulation detection interference: Interference for detecting a change in the depth of amplitude modulation,'' J. Acoust. Soc. Am. **88**, S145.

Auditory discontinuities interact with categorization: Implications for speech perception[a)]

Lori L. Holt[b)]
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto
Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska 68131

Randy L. Diehl
Department of Psychology and Center for Perceptual Systems, University of Texas at Austin, Austin, Texas 78712

(Received 5 September 2003; revised 9 June 2004; accepted 14 June 2004)

Behavioral experiments with infants, adults, and nonhuman animals converge with neurophysiological findings to suggest that there is a discontinuity in auditory processing of stimulus components differing in onset time by about 20 ms. This discontinuity has been implicated as a basis for boundaries between speech categories distinguished by voice onset time (VOT). Here, it is investigated how this discontinuity interacts with the learning of novel perceptual categories. Adult listeners were trained to categorize nonspeech stimuli that mimicked certain temporal properties of VOT stimuli. One group of listeners learned categories with a boundary coincident with the perceptual discontinuity. Another group learned categories defined such that the perceptual discontinuity fell within a category. Listeners in the latter group required significantly more experience to reach criterion categorization performance. Evidence of interactions between the perceptual discontinuity and the learned categories extended to generalization tests as well. It has been hypothesized that languages make use of perceptual discontinuities to promote distinctiveness among sounds within a language inventory. The present data suggest that discontinuities interact with category learning. As such, ''learnability'' may play a predictive role in selection of language sound inventories. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1778838]

PACS numbers: 43.71.Es, 43.71.Pc, 43.66.Ba [PFA] Pages: 1763–1773

I. INTRODUCTION

Considering the large set of physically realizable sounds of the human vocal tract, the regularity in phonetic inventories of surveyed languages is remarkable (Maddieson, 1984). There have been a number of hypotheses about factors that shape preferred sound inventories. A common theme running throughout these proposals is that sounds are favored if they are easy to produce and/or easy to distinguish auditorily from other sounds (Diehl *et al.*, 2003; Liljencrants and Lindblom, 1972; Stevens, 1972, 1989). For example, according to the quantal theory (Stevens, 1972, 1989), there are regions along certain phonetic dimensions where articulatory changes have large acoustic consequences and other regions where comparable articulatory variation results in much more modest acoustic change. These alternating regions of acoustic instability and stability are assumed to create conditions for a kind of optimization of sound inventories. If sound categories are located in the stable regions, they require less articulatory precision and are thus easier to produce reliably than sound categories in unstable regions. These stable sound categories will also tend to be auditorily distinctive because they are separated by regions of acoustic instability.

Of course, communicative benefits arise not from acoustic distinctiveness among sounds but from distinctiveness within the human auditory system. Mathematical simulations that incorporate distance metrics based on realistic auditory representations have been able to predict some of the regularities witnessed in phonetic inventories (Diehl *et al.*, 2003; Lindblom, 1986). Auditory representations differ from acoustic representations because the auditory perceptual space may be ''warped'' in such a way that equal acoustic distinctions need not produce equivalent perceptual changes. Languages may capitalize on regions of perceptual space where sensitivity is enhanced, adopting sounds for which moderate changes in articulatory or acoustic characteristics result in disproportionately large perceptual consequences. For example, there are some regions of acoustic space where sharp peaks in discrimination functions are observed. These ''auditory discontinuities'' [natural boundaries in perceptual space across which discrimination is enhanced (Kuhl and Miller, 1975)][1] would be good candidates for placement of boundaries between phonetic categories because tokens from each category would have enhanced distinctiveness with minimal acoustic change. Enhanced distinctiveness in turn provides more robust communication. Stevens (1989) extended the quantal theory to the relation between acoustic

[a)]Preliminary reports of this work were presented at the 145th Meeting of the Acoustical Society of America, June 2003 and at the 10th Annual Cognitive Neuroscience Society Meeting, March 2003.
[b)]Electronic mail: lholt@andrew.cmu.edu

signals and auditory responses. Certain nonlinearities in the auditory system were assumed to give rise to regions along acoustic dimensions that are auditorily stable and other regions along those same dimensions that are auditorily unstable. Again, it was hypothesized that sound categories tend to be located in the stable regions. An intervening unstable region would then correspond to a kind of auditory discontinuity or natural boundary that enhances the distinctiveness of sound categories.

In addition to enhanced distinctiveness, there is another benefit that may be conferred by utilizing auditory discontinuities as phonetic boundaries. A discontinuity may provide a parsing of the auditory space and delineate an initial ''natural'' boundary between phonetic categories (Kuhl, 1993). For a novice language-learner, these boundaries could facilitate the imposing task of forming phonetic categories from the considerable acoustic variance in speech. Phonetic systems that include boundaries coincident with auditory discontinuities may thus be more *learnable*. If this is the case, learnability may be a characteristic of communication systems that exerts selective pressure on the development of languages (Deacon, 1997; Lotto, 2000). That is, attributes of language systems that promote learning may tend to be favored. This includes aspects of phonetic categories. Evidence from studies of categorization in other domains suggests that some sets of stimuli are more readily categorized than others (Alfonso-Reese *et al.*, 2002; Ashby *et al.*, 1999). Many factors may contribute to the relative ease of categorization including stimulus dimensionality, covariance among stimulus dimensions, nonlinearity in category boundaries, and overlap in input distribution variability.

We propose here that the placement of category boundaries at preexisting auditory discontinuities may facilitate phonetic category formation and that this increase in learnability may be one of the sources of regularities observed in phonetic inventories. A particular example is the regularity in voicing category distributions seen across languages and the relation of this regularity to a well-documented nonmonotonicity in temporal processing.

A. Auditory discontinuities in temporal processing

Although a host of cues differentiate voiced from voiceless consonants (Lisker, 1986), the temporal interval between oral release and onset of voicing (voice onset time, VOT) is especially predictive of the contrast. Across languages, speakers tend to produce three kinds of voicing: *lead* consonants for which voicing precedes the release of the consonant occlusion (thus having a negative VOT value), *short lag* consonants for which voicing closely follows the oral release (with VOT values around 0–20 ms), and *long lag* consonants for which voicing lags well behind the consonant release (with large positive VOT values). Frequency distributions of VOT productions aggregated across languages reflect three modes, with modes centered at −100, +10, and +75 ms VOT (Cho and Ladefoged, 1999; Keating, 1984; Lisker and Abramson, 1964).[2] These characteristic distributions are cross-linguistically regular not just in their modes, but also in more precise statistical patterning. For example, distributions of VOT productions typically have a gap in the region corresponding to small negative VOT values. These values appear very rarely in speech production cross-linguistically. Keating (1984) suggests that this gap may be a general characteristic of languages regardless of the number of voicing categories. The extent, or range, of the distributions appears to be fairly regular across languages as well. Perhaps partly as a consequence of the absence of VOT productions in the small negative VOT range, short-lag distributions tend to have quite constrained variability across VOT, with values clustered from about 0 to 20 ms VOT (Keating, 1984). Therefore, cross-linguistically, boundaries between the voicing distributions generally occur at roughly −20 and +20 ms VOT.

Most languages make use of only two of the three distributions. English, for example, distinguishes short-lag from long-lag consonants, whereas Spanish contrasts lead with short-lag voicing. Of note, no language reported by Lisker and Abramson (1964) contrasts lead with long-lag voicing. This is of interest because theories predicting sound inventories from acoustic factors might expect to observe two-way voicing contrasts positioned at opposite ends of the VOT continuum, thus enhancing the acoustic distinctiveness of the contrast. However, each of the 51 languages surveyed by Keating *et al.* (1983) employed some type of short-lag stop consonant. Thus, rather than dispersing voicing productions acoustically, languages tend to contrast lead or long-lag voicing with the more intermediate (and less acoustically distinct) short-lag distribution.

Provocative early findings in infant speech perception indicated that mechanisms responsible for patterns of voicing perception are functional at a very early age (Eimas *et al.*, 1971; Trehub and Rabinovich, 1972). Moreover, experience appears to play a relatively minor role in shaping voicing discrimination early in development (Lasky *et al.*, 1975; Streeter, 1976). Lasky *et al.* (1975), for example, found that infants from a Spanish-speaking environment discriminate differences in VOT if stimuli straddle either the Spanish (lead versus short-lag) or the English (short-lag versus long-lag) VOT boundary but show little evidence of discrimination otherwise. This pattern contrasts with Spanish adults, who exhibit a single peak in discriminability across VOT at the lead versus short lag boundary (Abramson and Lisker, 1972; Elman *et al.*, 1977; Williams, 1977). Aslin *et al.*, (1981) have found comparable patterns of discriminability for infants from English-speaking homes. Moreover, Kikuyu-reared infants exhibit a similar discrimination pattern along a labial VOT continuum despite the fact that Kikuyu adults do not contrast voicing for labial articulations (Streeter, 1976). Thus, infants may exhibit regions of better-than-chance discriminability along a VOT dimension that are not characteristic of native adult perceivers. Moreover, the regions of best discrimination are relatively consistent across groups of infants and align fairly well with the boundaries at −20 and +20 ms VOT that partition the trio of distributions observed in VOT productions across languages.

Research with nonhuman animals has suggested that general characteristics of mammalian auditory processing may be responsible for these patterns of discrimination. Chinchillas and macaque monkeys without significant expe-

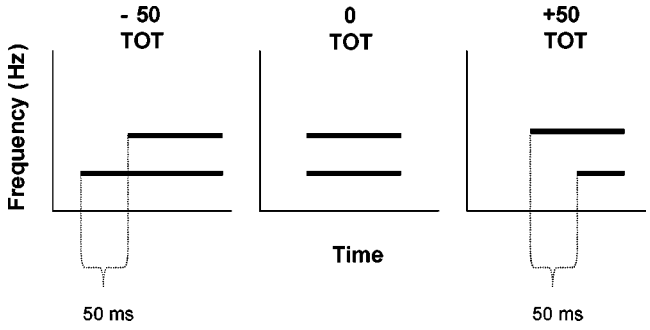


FIG. 1. Schematic spectrograms of TOT stimuli modeled after those of Pisoni (1977). Two tones (500 and 1500 Hz) vary in relative onset to create series of stimuli varying from −60 to +120 ms TOT. Three representative stimuli are shown.

rience with speech discriminate the VOT dimension much like human listeners (Kuhl and Miller, 1975; 1978; Kuhl and Padden, 1982). Kuhl and Miller interpreted these results as indicative of a general psychoacoustic discontinuity in VOT perception. Pisoni (1977) examined the generality of this proposal with human adult listeners by constructing a nonspeech dimension modeling a temporal characteristic of VOT (see also Miller *et al.*, 1976). The nonspeech stimuli consisted of coterminous tones of 500 and 1500 Hz. Along a tone-onset-time (TOT) dimension, tones varied in their relative onset times such that the lower-frequency tone preceded the higher-frequency tone in onset, the two tones were simultaneous, or the lower-frequency tone lagged the higher-frequency tone. A schematic of three such stimuli is presented in Fig. 1. In a series of experiments, Pisoni (1977) found two regions of relatively more accurate TOT discrimination. Listeners were excellent at discriminating stimuli straddling approximately −20 and +20 ms TOT, but were poorer at discriminating stimuli spanning the same acoustic distance in other regions along the TOT dimension. Interestingly, the two regions of enhanced nonspeech TOT discrimination align quite closely with the −20 and +20 ms boundaries that partition the three distributions of VOT productions that are observed across languages (Lisker and Abramson, 1964). Figure 2 illustrates this discrimination pattern.[3] Similar patterns of TOT discrimination performance are obtained for 10-week-old infants (Jusczyk *et al.*, 1980). Electrophysiological studies have revealed several neurophysiological correlates of discrimination peaks for VOT and TOT stimuli (e.g., Simos and Molfese, 1997; Simos *et al.*, 1998; Sinex *et al.*, 1991; Steinschneider *et al.*, 1999).

Overall, data from adult and infant perception of speech and nonspeech, neurophysiological measures, and animal behavioral experiments indicate that the acoustic VOT and TOT dimensions are not identical to their auditory representations. Rather, there is a discontinuity such that stimulus pairs differentiated by equal acoustic change do not necessarily produce equivalent perceptual change; despite the equivalence of acoustic differences, discriminability may be enhanced for one pair relative to the other. The cross-linguistic regularities in voicing distributions suggest that many languages may exploit this discontinuity by drawing voicing categories from either side of the region of enhanced sensitivity. One benefit of this would be greater perceptual

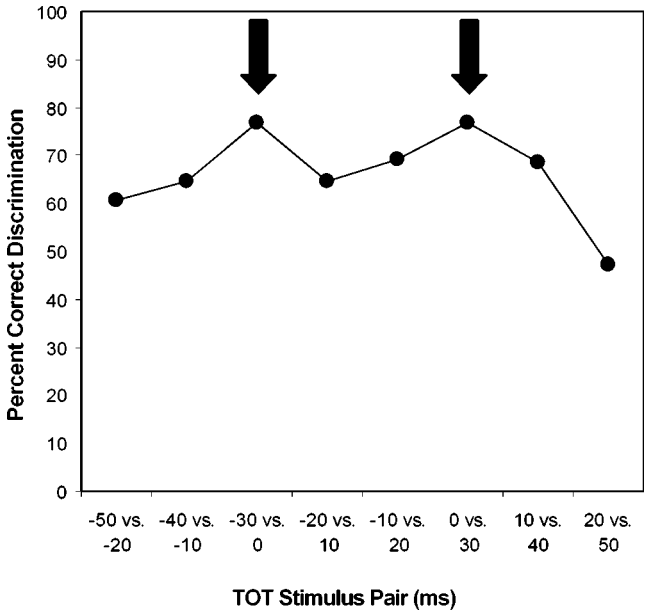


FIG. 2. Discrimination performance of participants in a pilot discrimination test illustrate the auditory discontinuity at approximately −20 ms and +20 ms. Although each stimulus pair spans an equal acoustic distance, discrimination is relatively more accurate when stimulus pairs straddle approximately −20 or +20 ms, as indicated by the arrows.

distinctiveness of tokens with small changes in acoustics (Kuhl and Miller, 1975; Pisoni, 1977). A second possible benefit would be that the discontinuities provide an initial parsing of the VOT space that can facilitate learning voicing categories that are consistent with these natural boundaries. That is, in addition to preferring sound categories that are ''easy to say'' and ''easy to hear,'' languages may favor categories that are ''easy to learn.''

The present experiments were designed to test the hypothesis that operating characteristics of the auditory system constrain the learnability of sound categories. Specifically, they test the prediction that categories defined by onset asynchrony are easier to learn if they are consistent with the auditory discontinuity shown to occur at onset asynchronies near 20 ms. In these experiments, the categories are made up of TOT stimuli similar to those described by Pisoni (1977). The use of nonspeech stimuli provides control over stimulus attributes and participants' history of experience with the novel nonspeech TOT stimuli. As with the original work of Pisoni (1977), the results of nonspeech experiments can illuminate some of the general perceptual and cognitive processes that are involved in online speech perception and that may bias the development of linguistic systems.

II. EXPERIMENT 1

Participants in experiment 1 were trained with explicit feedback to assign TOT stimuli to two categories. One group of listeners learned two categories for which the boundary (as indicated by feedback and the tails of the distributions) was consistent with the reported peak in discriminability at approximately +20 ms. Another group of listeners learned two categories for which the boundary was inconsistent with the peak in discriminability. For this group of listeners, the peak in discrimination at +20 ms occurred within a category.

If learnability is enhanced by aligning category boundaries with perceptual discontinuities, then participants in the consistent condition should learn categories more quickly; that is, they should be faster to reach a criterion level of categorization performance. If learning does not interact with characteristics of the auditory space, then there should be no difference in the time to learn categories in the two conditions.

A. Methods

1. Participants

Twenty-eight listeners participated. One-half of the participants were randomly assigned to the consistent learning condition; the others were assigned to the inconsistent learning condition. All participants reported normal hearing, learned English as a first language, and were Carnegie Mellon University undergraduates or staff. Each participant was paid $15.

2. Stimuli

The stimuli were modeled after those of Pisoni (1977) and were synthesized using Matlab® (Mathworks, Inc.). A TOT dimension was constructed by creating two-tone stimuli with 500- and 1500-Hz tone components. The 1500-Hz tone had a constant duration of 230 ms and was approximately 12 dB lower in amplitude than the 500-Hz tone. Duration of the 500-Hz tone was varied to produce TOT differences. The only deviation from Pisoni's stimulus construction methods was the addition of 5-ms linear amplitude ramps at onset and offset. Stimuli were labeled with negative TOT values when the lower-frequency tone preceded the higher-frequency tone and with positive TOT values when the lower-frequency tone lagged the higher-frequency tone. A TOT value of zero indicated simultaneous onset. The full TOT series consisted of 37 stimuli varying from −60 to +120 ms TOT in 5-ms steps.[4] Figure 1 illustrates representative pseudospectrograms for −50, 0, and +50 ms TOT.

Acoustic presentation was controlled by TDT System II hardware (Tucker-Davis Technologies). Stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented binaurally over linear headphones (Beyer DT-100) at approximately 70 dB SPL(A).

3. Procedure

Participants first received categorization training with feedback to reach a criterion level of performance. Following training, listeners completed an ABX discrimination task. Next, they received a short categorization refresher. Finally, they again categorized stimuli in a generalization test across the entire TOT series without feedback. Each segment of the experiment was completed while the listener sat comfortably in a sound-attenuating booth wearing headphones and holding an electronic response box. Instructions preceded each segment on an LCD display mounted at eye level in front of the participant. Participants were allowed to take short breaks between segments and the entire session lasted less than 2 h.

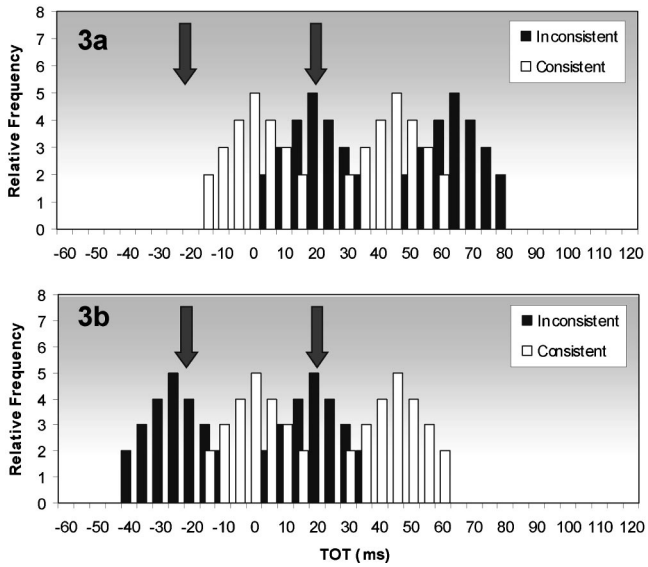


FIG. 3. Stimulus input distributions illustrate the relative frequency of stimulus presentation during training as a function of TOT. (a) Shows stimulus distributions of experiment 1. (b) Shows stimulus distributions for experiment 2. White bars correspond to consistent condition stimuli for which distributions and feedback assignment are coincident with the auditory discontinuity. Black bars show inconsistent condition stimuli where the auditory discontinuity is positioned within a distribution and feedback assignment is inconsistent with the auditory discontinuity.

a. Training. During training, listeners heard a single stimulus over headphones, made a two-alternative forced choice categorization response, and received feedback via a light above the correct response-box button. There were no labels on the response buttons.

Listeners in the consistent condition (for which auditory discontinuity, input distribution boundary, and boundary of feedback assignment were the same) heard stimuli drawn from the distributions illustrated as white bars in Fig. 3(a). The height of the bar illustrates the relative frequency of stimulus presentation for a particular TOT value. The arrows indicate peaks in discrimination observed at −20 and +20 ms TOT (see Fig. 2 pilot-test data). For the consistent condition, stimuli were sampled from two distributions separated along the TOT dimension at approximately +20 ms. Stimuli with TOT values less than +20 ms were assigned to one category and stimuli with TOT greater than +20 ms were assigned to the other category.

The input distributions of the inconsistent condition (for which auditory discontinuity differed from the input distribution boundary and the boundary of feedback assignment) are shown as dark bars in Fig. 3(a). The input distributions were separated along the TOT dimension at approximately +40 ms, a region of the TOT space that has no known nonlinearities in discriminability. The psychoacoustic discontinuity at +20 ms fell at the mode of one of the inconsistent condition distributions.

For each block of trials, listeners heard 46 stimuli that fully sampled the two input distributions. Listeners were instructed that some of the sounds corresponded to the left button whereas others corresponded to the right button. Following each response, the light above the correct button was illuminated as feedback. The mapping between distribution

and response button was counterbalanced across participants. Listeners were encouraged to maximize performance and were told that the length of the experimental session was a function of performance.

After each block of training, listeners' percent correct identification across the 46 categorization trials was calculated. Listeners completed training blocks until reaching a criterion of 90% correct performance (42 of 46 trials correct) or a maximum of 20 blocks (920 trials overall). Categorization of stimuli within a block took less than 5 min to complete.

b. Discrimination. After reaching criterion performance on the training segment or reaching the 20-block maximum, listeners completed an ABX discrimination task. On each trial, a stimulus pair (A and B) separated by five steps (25 ms) along the TOT dimension spanning −25 to +85 ms was presented with a 750-ms interstimulus silent interval. After another 750 ms of silence, a third stimulus (X, identical to either A or to B) was presented. Participants indicated whether the third stimulus was identical to either the first or to the second stimulus by pressing a button on an electronic response box. In all, participants discriminated 23 pairs of stimuli that occurred in four counterbalanced permutations per pair (ABA, ABB, BAA, BAB) for a total of 92 discrimination trials. Participants discriminated this full block of trials two times (making 184 overall discriminations). Collapsed across counterbalanced stimulus orderings, each subject made eight discrimination responses for each pair. Participants received feedback on each trial. Pisoni (1977; Pisoni *et al.*, 1982) has argued that providing feedback in the discrimination test is critical because it forces participants to focus on the intended stimulus dimensions and to maintain the same focus from trial to trial.

c. Categorization refresher. To refamiliarize category mappings after the discrimination task, listeners heard two additional blocks, each fully sampling the input distributions. The procedure was identical to that of training. Listeners responded with a categorization response and received feedback on each trial.

d. Generalization test. Finally, listeners completed a generalization test of categorization. This task provided a test of categorization across the entire TOT range including stimuli not experienced during training. From these data, identification boundaries estimated using probit analysis and generalization of the learning were assessed. Stimuli spanned −60 to +120 ms TOT. Each stimulus was equally probable and was presented three times. Listeners were instructed to categorize the stimuli as in previous segments of the experiment. There was no feedback.

B. Results

1. Training

The primary hypothesis of experiment 1 was that the auditory discontinuity at approximately +20 ms TOT would interact with category learning. Specifically, it was predicted that category learning should be facilitated in the consistent condition, for which the category boundary coincided with the auditory discontinuity. As can be seen in Fig. 4, the data support this prediction. Listeners in the inconsistent condition required significantly more blocks of categorization training (M=8.21) to learn the TOT categories to a criterion level than did consistent-condition listeners (M=3.36), $t(13)=2.28$, $p<0.02$.

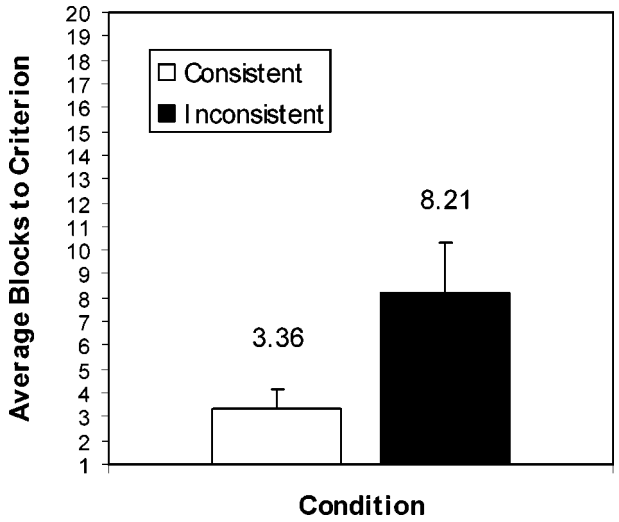


FIG. 4. Mean number of blocks required to reach criterion categorization performance during training for consistent and inconsistent condition listeners in experiment 1. Error bars represent standard error of the mean.

2. Generalization

Listeners labeled TOT stimuli varying from −60 to +120 ms TOT in the generalization segment of the experiment, thus responding to stimuli present in categorization training and to novel stimuli. These categorization results (presented as a function of an arbitrary category "A" indicating the input distribution with the lowest mean value along the TOT dimension) are shown with input distributions from training in Fig. 5. A 2×37 (condition × TOT stimulus) ANOVA revealed a significant main effect of condition, $F(1,36)=18.77$, $p<0.001$. Inconsistent condition participants categorized more of the TOT series stimuli as members of the "A" category than consistent condition listeners. There was also a significant main effect of TOT Stimulus, $F(1,36)=81.57$, $p<0.001$, indicating that categorization varied as a function of the TOT stimulus series as would be expected if listeners learned to accurately categorize the stimuli. There was also a significant condition × TOT stimulus interaction, $F(1,36)=2.28$, $p<0.001$, indicating that the identification functions differed in their shape across conditions. This pattern most likely arises from consistent condition listeners' categorization of TOT stimuli on the far negative end of the stimulus series (see Fig. 5). Consistent condition listeners imperfectly generalized to stimuli at this end of the series, instead rather consistently mimicking the frequency distribution of the trained categories in their generalization responses. Identification boundaries, as estimated from probit boundary analyses computed for individual subjects, support these analyses. There was a significant boundary shift across conditions, $t(13)=5.42$, $p<0.001$ (33.17 ms TOT for consistent listeners versus 51.62 ms TOT for inconsistent listeners). There was no significant difference in the

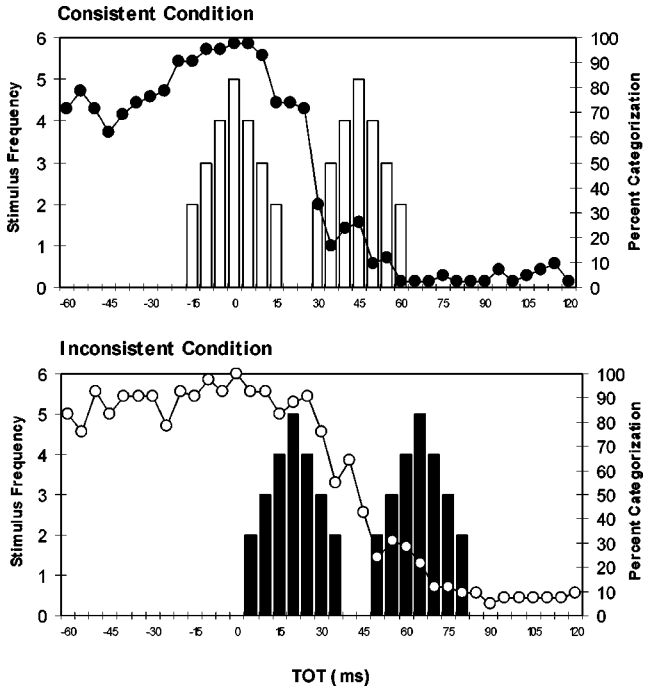


FIG. 5. Categorization responses (presented on the right-most *y*-axis as a function of an arbitrary category "A" indicating the input distribution with the lowest mean value along the TOT dimension) to stimuli presented in the generalization task are shown for consistent and inconsistent condition listeners in experiment 1. The histogram bars denote the position of the categories learned during training along the TOT dimension as a function of the relative frequency of stimulus presentation plotted on the left-most *y*-axis.

probit slopes, $t(13)=0.63$, $p=0.54$; the steepness of the identification functions was statistically equivalent across conditions.

3. Discrimination

Listeners also discriminated TOT stimuli in an ABX task. Whereas categorization and generalization analyses demonstrated significant effects across groups, there was little evidence that patterns of discrimination shifted with categorization training. A 2×23 (condition $\times$ discrimination pair) ANOVA revealed that consistent group listeners were somewhat better at all discriminations although not significantly so, $F(1,22)=3.22$, $p<0.09$. Even this slight trend is surprising, considering that consistent condition listeners had significantly less experience listening to the stimuli because they were faster to reach criterion categorization performance. Pairs were not equally discriminable, $F(1,22)=4.80$, $p<0.0001$, but the patterns did not relate cleanly to categorization experience. There was no condition $\times$ discrimination pair interaction, $F(1,22)=1.28$, n.s. Overall, discrimination does not appear to be a reliable index of listeners' categorization competence in the present paradigm.

4. Summary

The main prediction of the learnability hypothesis was supported by the results of experiment 1. Participants trained with a category boundary that was consistent with the auditory discontinuity learned to categorize the sounds faster than participants in the inconsistent condition. The data also make it clear that the presence of a single discontinuity does not determine possible learnable categories. Participants in the inconsistent condition *did* learn the categories on which they were trained. The generalization test shows a shift in identification boundaries and no obvious change in the identification function related to the discontinuity region. Thus, characteristics of general auditory representations appear to affect the ease with which categories are learned, but the general learning system is capable of overcoming the initial parsing of the perceptual space.

III. EXPERIMENT 2

In experiment 1, the two conditions differed in the range of TOT values that were categorized as well as the consistency of the category boundary with the auditory discontinuity. Although the overall range of TOT was equivalent across conditions, the inconsistent distributions included more large TOT values than did consistent distributions. It is possible that the perceptual space spanned by the acoustic TOT stimuli is compressively nonlinear (Fechner, 1860), rendering perceived stimulus differences of the inconsistent condition less salient and thereby increasing the difficulty of the inconsistent categorization task. This, too, would be an interaction of general operating characteristics of the auditory system with categorization, but not the sort under investigation in experiment 1.

Experiment 2 controls for this possibility. For this experiment, the two input distributions of the experiment 1 inconsistent condition were modified such that the input distribution with a mode at +65 ms TOT [see Fig. 3(a)] was reflected about +20 ms TOT to create a new input distribution with negative TOT values. The other three input distributions of experiment 1 remained unchanged, as shown in Fig. 3(b). Note that, in experiment 2, both of the inconsistent-condition input distributions spanned a region in auditory space defined by high discriminability. The peak in discriminability at approximately +20 ms fell within one input distribution and the peak in discriminability at about −20 ms fell within the opposite distribution. As a result, experiment 2 provides an opportunity to observe whether categories inconsistent with both of these auditory discontinuities are even more difficult for listeners to learn.

A. Methods

1. Participants

Participants were recruited from the Carnegie Mellon University undergraduate psychology participant pool. Each participant earned a credit toward a psychology research requirement and $5 for participation. All participants had normal hearing and learned English as the first language. Fourteen listeners participated in each condition for a total of 28 participants.

2. Stimuli

The TOT series stimuli were identical to those of experiment 1.

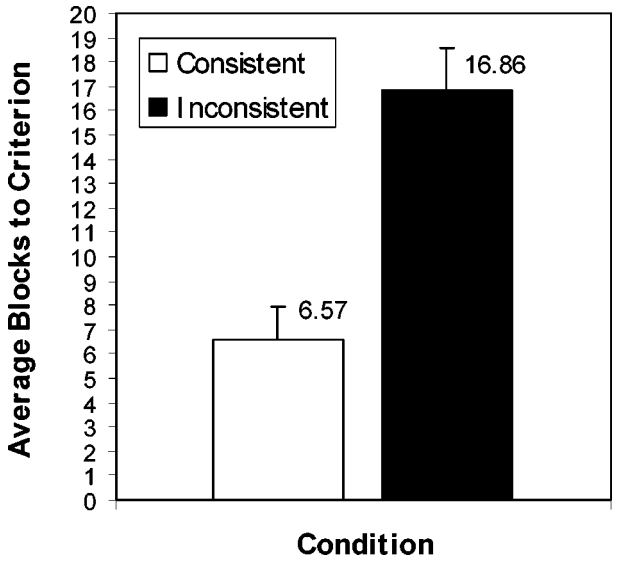


FIG. 6. Mean number of blocks required to reach criterion categorization performance during training for consistent and inconsistent condition listeners in experiment 2. Error bars represent standard error of the mean.

3. Procedure

The procedure was nearly identical to that of experiment 1. The only difference between experiments 1 and 2 was the position of one of the inconsistent condition distributions along the TOT dimension. Figure 3(b) illustrates the input distributions for experiment 2. Three of the four input distributions are the same as those of experiment 1. The TOT input distribution from experiment 1 centered on +65-ms TOT was replaced with an inconsistent condition distribution centered at −25-ms TOT, creating a boundary at approximately −5-ms TOT.

As a consequence of the reflected input distribution, inconsistent condition listeners in experiment 2 were required to categorize input distributions that crossed both −20 and +20 ms discontinuities. As a result, the inconsistent condition learning task of experiment 2 may have been even more difficult than that of experiment 1, for which only one input distribution crossed a discontinuity. Stimulus distributions for the consistent condition were equivalent to those of experiment 1 and thus provide a replication.

Tasks and experimental apparatus for this experiment were identical to those of experiment 1.

B. Results

1. Training

Figure 6 illustrates the mean number of blocks to criterion categorization performance across conditions. The results of experiment 1 do not appear to be merely a result of compressive nonlinearity in perception of large TOT values. Listeners in the inconsistent condition required significantly more blocks of categorization training (M=16.86) to learn the TOT categories to a criterion level than did consistent-condition listeners (M=6.57), $t(13)=4.71$, $p<0.0001$. Thus, the results of experiment 1 are replicated in experiment 2.

The consistent condition was identical to that of experiment 1. As would be expected, there was no significant difference in the number of blocks listeners required to reach criterion in the consistent condition across experiments, $t(13)=1.908$, $p=0.79$. However, it took listeners significantly longer to learn the inconsistent categories of experiment 2 than those of experiment 1 [16.86 vs 8.21 blocks, $t(13)=2.99$, $p<0.01$]. Note that one of the distributions for the inconsistent condition was identical in both experiments and the distance between distributions was constant. Despite these similarities, participants in experiment 2 performed more poorly in the categorization training task.

In fact, listeners in experiment 2 inconsistent condition had difficulty reaching 90% accuracy within the arbitrarily set maximum 20 blocks (920 trials) of training. In experiment 1, categorization performance in the last block of training was equivalent across conditions, $t(13)=1.84$, $p=0.09$, suggesting that listeners ultimately learned their respective categories equally well. However, in experiment 2, consistent condition listeners were significantly more accurate in the last block of training than were inconsistent condition listeners, $t(13)=3.866$, $p=0.002$. This is primarily due to the high proportion of inconsistent condition listeners ($N=10$, 71%) that did not reach 90% criterion within 20 training blocks. Categorization accuracy in the last block was only 72.05% for this subset of inconsistent condition listeners. Nevertheless, the task of inconsistent listeners was not impossible. One listener achieved criterion categorization performance in a single block of trials.[5]

It should also be noted, however, that there was a difference in the listener samples between experiments 1 and 2 such that across conditions, listeners in experiment 2 were slower (M=11.71 blocks) to reach criterion than experiment 1 listeners (M=5.78 blocks), $t(27)=3.44$, $p=0.002$, and the performance of experiment 1 Inconsistent condition listeners was not statistically distinguishable from the consistent condition listeners of experiment 2, $t(13)=0.583$, $p=0.57$. This latter piece of evidence should be considered in light of the fact that the arbitrarily 20-block maximum for the training session differentially affected inconsistent condition listeners, creating the possibility of a ceiling effect for the inconsistent condition and artificially reducing the hypothesized effect between the inconsistent and consistent conditions. Thus, interpretation of between-experiment comparisons should be interpreted with some caution. Nevertheless, whatever the source of listener differences between the experiments, each group's performance supports the primary hypothesis that the operating characteristics of the auditory system interact with learning to make inconsistent condition categories more difficult to learn than consistent condition categories.

2. Generalization

In the generalization segment of the experiment, listeners labeled stimuli varying from −60 to +120 ms TOT with no feedback. Average categorization functions (expressed in terms of an arbitrary category "A") for consistent and inconsistent condition listeners are shown in Fig. 7. A 2×37 (condition × TOT stimulus) ANOVA of these results revealed a significant main effect of TOT Stimulus, $F(1,36)=83.04$, $p<0.0001$, corresponding to the sigmoid shape of

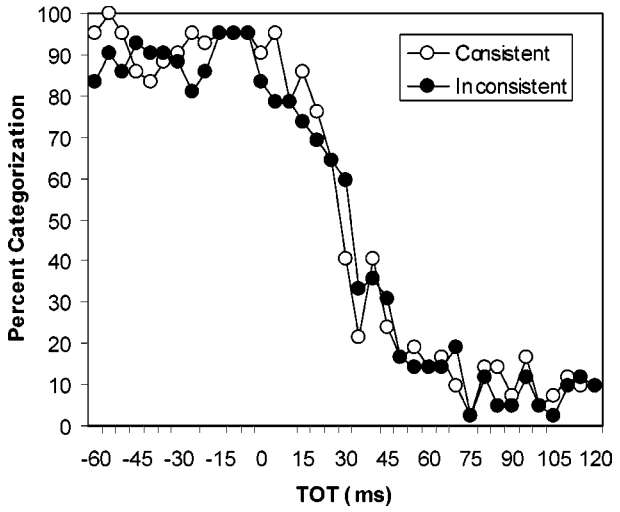


FIG. 7. Categorization responses during generalization task for consistent and inconsistent condition listeners in experiment 2.

the categorization function and indicating quite consistent categorization by listeners. Contrary to the results of experiment 1, there was no significant influence of condition upon generalization responses, $F(1,36)=1.06$, $p=0.3$. As indicated in Fig. 7, there was no evidence of a category boundary shift. There was no significant difference in probit boundaries, $t(12)=0.272$, $p=0.79$ ($M_{Con}=34.4$, $M_{Incon}=37.2$), and no difference in the slopes of the functions, $t(12)=1.84$, $p=0.09$.

3. Discrimination

In a 2×23 (condition × stimulus pair) ANOVA, there was a main effect of stimulus pair, $F(1,22)=5.283$, $p<0.0001$, indicating differences in discrimination accuracy as a function of stimulus pair. There was no condition main effect ($F<1$) and no condition by stimulus pair interaction ($F<1$).

4. Summary

The effect observed in experiment 1 was replicated in experiment 2. As in experiment 1, consistent condition listeners were faster to reach criterion categorization performance than listeners in the inconsistent condition. This result supports the original contention that it is the placement of the category boundary in relation to the auditory discontinuity and not the range over which TOT stimuli are categorized that makes the inconsistent distributions more difficult to learn.

It is of potential interest that experiment 2 inconsistent condition listeners were significantly slower to learn the categories than were experiment 1 inconsistent condition listeners. This appears to be due to the placement of experiment 2 inconsistent condition distributions on auditory discontinuities at both −20 and +20 ms TOT. The difficulty of learning the categories in the inconsistent condition of experiment 2 may be responsible for the absence of a category boundary shift relative to the consistent condition in the generalization test.

IV. DISCUSSION

The present findings indicate that auditory categories are not equally learnable. Basic auditory sensitivities interact with auditory categorization, causing some distributions of sounds to be more easily categorized than others. Specifically, categories separated by regions of high discriminability are more easily learned than those that span such regions. Nevertheless, despite the advantage that they confer to learnability, psychoacoustic sensitivities do not necessarily dictate categorization. Listeners in the inconsistent conditions required more experience, but did ultimately learn to categorize the sounds with high accuracy in experiment 1.

These results may have implications for learning phonetic categories. TOT stimuli were originally created to mimic the onset asynchrony between lower and higher frequency energy that corresponds to VOT in speech (Pisoni, 1977). If this analogy holds, then considering the present results it may be hypothesized that the placement of phonetic category distributions on either side of the discontinuity around 20 ms may confer enhanced learnability of speech categories distinguished by VOT. If linguistic systems are selected to be learnable, then phonetic inventories would be expected to include categories separated by the +20 and −20 ms VOT boundaries. This pattern is, in fact, widely attested among the world's languages (Keating, 1984; Lisker and Abramson, 1964). For example, the English labial voicing distinction (/b/ vs /p/) is realized with distributions of VOT values on either side of the 20-ms boundary. The fact that inconsistent listeners eventually reached criterion performance suggests that contrasts that do not follow this pattern are not unachievable, but may be more difficult to learn.

Previous attempts to account for regularities in phonetic inventories have focused on the communicative benefits arising from ease of production and ease of discrimination (e.g., Diehl *et al.*, 2003; Liljencrants and Lindblom, 1972; Lindblom, 1986; Stevens, 1972). In the case of VOT, the ''easy to say''/''easy to hear'' approach proposes that aligning category boundaries with auditory discontinuities enhances discriminability of tokens from different categories without imposing substantial articulatory costs. The learnability account proposed here should be seen as augmenting, not replacing, such accounts. Phonetic systems should be easy to say, easy to hear, *and* easy to learn. In fact, if assignment of speech sounds to equivalence classes (e.g., phonetic categories) serves a communicative function, then the ease with which categories can be formed and the accuracy with which sounds are assigned to categories may be an essential value on which linguistic systems are selected.

The proposal that auditory discontinuities can provide an initial parsing of the perceptual space for categorization is similar to Kuhl's description of ''natural auditory boundaries'' in her native language magnet theory of phonetic acquisition (Kuhl, 1991, 1993, 2000). According to Kuhl (1993), the initial discriminative abilities of infants reflect the presence of these boundaries. As infants are exposed to native language distributions of speech sounds, these boundaries are modified and sometimes deleted by the process of categorization. The results of the current experiments lend some support to this proposal. The onset asynchrony discon-

tinuity provided a natural boundary for the consistent condition categories but the negative effects of this discontinuity in the inconsistent condition of experiment 1 were ameliorated by learning. However, we must be cautious in generalizing from this one example to all phonetic categories. The voicing contrast is a fairly clear case of alignment between linguistic boundaries and natural auditory boundaries. There is also evidence for natural boundaries between some place-of-articulation contrasts for syllable-initial consonants (Eimas, 1974; Kuhl and Padden, 1983). However, there is no empirical evidence that auditory discontinuities are pervasive and underlie all or even most phonetic distinctions (e.g., Macmillan, 1987; Rosen and Howell, 1987). For example, recent studies have not found strong support for proposed auditory discontinuities in the vowel space (Fahey and Diehl, 1996; Fahey *et al.*, 1996; Hoemeke and Diehl, 1994; Molis *et al.*, 1998).

Inasmuch as auditory discontinuities do not dictate the categories that can be learned (recall that inconsistent condition listeners could learn the present categories with more experience), it is also important to acknowledge that auditory discontinuities themselves are malleable. Auditory discontinuities appear not to be immutable boundaries independent of other stimulus characteristics, but rather are a function of the spectral and temporal makeup of the stimulus and its surrounding context. For example, the boundary for TOT identification shifts with changes in the frequency separation between the two tones (Parker, 1988). Similarly, animal studies of VOT identification demonstrate boundary shifts with increased frequency separation of F1 (at energy onset) and higher formants (Kluender and Lotto, 1994; Kuhl and Miller, 1978). Recent intracranial electrophysiological recording experiments in humans and monkeys provide physiological evidence that the psychophysical discontinuity typically described as the ''20-ms boundary'' is dependent on the spectral characteristics of the stimuli (Steinschneider *et al.*, 2004). Neurons' response properties across a VOT series shift as the F1 frequency of the speech syllables varies. Thus, although we have spoken of a 20-ms auditory discontinuity, it is the case that this boundary, in fact, is dependent upon the spectral characteristics of the acoustic components that define the onset asynchrony. As a result, we would expect the relative impact of the placement of categorization training boundaries to vary with changes in stimulus characteristics.

This more stimulus-bound way of viewing auditory discontinuities aligns with behavioral evidence of context-sensitivity of perception along VOT series (e.g., Miller and Liberman, 1979; Miller *et al.*, 1986; Summerfield, 1981). That is, the present data do not argue that auditory discontinuities determine category boundaries absolutely. Categories with boundaries violating an auditory discontinuity (as in the inconsistent conditions) may be learned with more experience and, likewise, category boundaries may be shifted with adjacent context or changes in stimulus characteristics. What the present data suggest is that learnability of categories is enhanced when the boundaries of category distributions are aligned with the operating characteristics of auditory processing. When the discontinuities shift because of stimulus characteristics, so, too, should category boundaries. This is exactly what is seen for place of articulation and VOT.

Despite converging evidence regarding the 20-ms asynchrony discontinuity, there has been considerable debate over its existence (e.g., Kewley-Port *et al.*, 1988; Pastore, 1988; Pastore and Farrington, 1996). One reason is that the behavioral correlates of the discontinuity are not consistently observed across measurement procedures and changes in spectra and may sometimes be impacted by differences in the relative degree of experience across participants (Watson and Kewley-Port, 1988). However, the present results are difficult to reconcile without suggesting that the auditory system may differentially process stimuli along the TOT dimension. A possible experience-bound alternative interpretation of the present findings is that English-speaking listeners' category boundary for voicing in speech influenced their perception and category learning of nonspeech stimuli along the TOT dimension. Consistent with this proposal, recent results have demonstrated an influence of native-language experience upon nonspeech perception (Bent *et al.*, 2003). However, in experiment 2 inconsistent categories spanned the natural boundaries at both +20 ms (close to the native English voicing boundary) and −20 ms (in a region of space not used contrastively in English voicing). Native-English listeners were significantly poorer at learning the categories in this condition than in the inconsistent condition of experiment 1 that spanned only the +20 ms boundary. Therefore, the possibility that native-language-specific categories along the VOT dimension interact with TOT learning does not account fully for the present findings.

If learnability is a determinant of the structure of phonetic systems, then empirical work on general perceptual categorization is relevant for comprehensive theories of speech perception. Whereas there is a rich tradition of categorization work in vision, there is a dearth of information on the factors that may facilitate the formation of complex auditory categories. Some of these factors may be low dimensionality of the stimulus space, covariance among stimulus dimensions, and linear category boundaries. With respect to the last factor, Diehl *et al.* (2001) report that boundaries for vowel categories in psychoacoustically scaled (e.g., Mel or Bark) formant space appear to be linear. They point to results from experiments in which participants are asked to identify tokens broadly sampled across the formant space (e.g., Molis, 1999). In these studies, the boundaries that provide the best discriminative separation of the identification scores were often (although not exclusively) simple linear functions. Many boundaries were horizontal or vertical lines reflecting unidimensional distinctions between vowel categories, whereas most other boundaries were horizontal lines that reflected distinctions based on formant frequency averaging or difference scores. It is intuitively (and computationally) clear that categories defined by linear boundaries would be easier to learn than categories with complex nonlinear boundaries. However, in order to make strong claims about learnability, the ease of learning nonspeech auditory categories with different boundary configurations must be examined.

Factors affecting phonetic acquisition extend beyond the structure of the categories to the details of the learning environment. Kuhl *et al.* (2003) recently demonstrated that in-

fants exposed to live adult models picked up phonetic information better than infants exposed to recordings of speakers. This result suggests the importance of social interaction in the learning situation, perhaps to direct the attention of the learner to important aspects of the sensory input. In addition, the characteristics of speech directed at children may be tailored to enhance learnability of phonetic categories. Child-directed speech has led to more efficient learning of phonetic categories by a computer model than adult-directed speech (de Boer and Kuhl, 2003).

The degree to which these particular factors of input distributions and learning situations enhance category formation can be tested using methods similar to those of the current study. With nonspeech stimuli, the researcher has control over the training set and the training conditions. Data can be collected from identification, discrimination, and category goodness rating tasks to provide a clear picture of how these factors interact with learning and the structure of the resulting categories. As exemplified by the experiments described here, these results have implications for explaining regularities in phonetic systems specifically and linguistic systems more generally.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (NSF BCS-0078768) to LLH and AJL, by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior to LLH, and by a research grant (5 R01 DC00427) from NIDCD to RLD. The authors thank Christi Adams for her help in conducting the experiments.

[1]A better term for this region may be a ''non-monotonicity.'' The present arguments apply not only to regions of perceptual space that result in a qualitative change in perception or a step-function in responding, but more generally to regions of perceptual space warped in relation to physical acoustic space.

[2]See Cho and Ladefoged (1999) for language-specific nuances that supplement this broad description.

[3]The data depicted in Fig. 2 were collected in a pilot ABX discrimination study using the same TOT stimuli as in the categorization studies described in this report.

[4]Pisoni (1977) investigated a range of TOT spanning −50 to +50 ms.

[5]It is interesting to note that this listener had more than 12 years of musical training and studied composition as a major in the School of Fine Arts.

Abramson, A. S., and Lisker, L. (**1972**). ''Voice-timing perception in Spanish word-initial stops,'' J. Phonetics **1**, 1–8.

Alfonso-Reese, L. A., Ashby, F. G., and Brainard, D. H. (**2002**). ''What makes a categorization task difficult?'' Percept. Psychophys. **64**, 570–583.

Ashby, F. G., Queller, S., and Berretty, P. M. (**1999**). ''On the dominance of unidimensional rules in unsupervised categorization,'' Percept. Psychophys. **61**, 1178–1199.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., and Perey, A. J. (**1981**). ''Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience,'' Child Dev. **52**, 1135–1145.

Bent, T., Bradlow, A. R., and Wright, B. A. (**2003**). ''The influence of linguistic experience on pitch perception in speech and nonspeech sounds,'' J. Acoust. Soc. Am. **113**, 2256.

Cho, T., and Ladefoged, P. (**1999**). ''Variation and universals in VOT: Evidence from 18 languages,'' J. Phonetics **27**, 207–229.

de Boer, B., and Kuhl, P. K. (**2003**). ''Investigating the role of infant-directed speech with a computer model,'' ARLO **4**, 129–134.

Deacon, T. W. (**1997**). *The Symbolic Species: The Co-evolution of Language and the Brain* (Norton, New York).

Diehl, R. L., Lindblom, B., and Creeger, C. P. (**2003**). ''Increasing realism of auditory representations yields further insights into vowel phonetics,'' in *Proceedings of the 15th International Congress of Phonetic Sciences, Vol. 2* (Causal, Adelaide), pp. 1381–1384.

Diehl, R. L., Molis, M. R., and Castleman, W. A. (**2001**). ''Adaptive design of sound systems: Some auditory considerations,'' in *The Role of Perceptual Phenomena in Phonological Theory*, edited by K. Johnson and E. Hume (Academic, San Diego), pp. 123–139.

Eimas, P. D. (**1974**). ''Auditory and linguistic processing of cues for place of articulation by infants,'' Percept. Psychophys. **16**, 513–521.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., and Vigorito, J. (**1971**). ''Speech perception in infants,'' Science **171**, 303–306.

Elman, J. L., Diehl, R. L., and Buchwald, S. E. (**1977**). ''Perceptual switching in bilinguals,'' J. Acoust. Soc. Am. **62**, 971–974.

Fahey, R., and Diehl, R. (**1996**). ''The missing fundamental in vowel height perception,'' Percept. Psychophys. **58**, 725–733.

Fahey, R. P., Diehl, R. L., and Traunmüller, H. (**1996**). ''Perception of back vowels: Effects of varying F1—F0 Bark distance,'' J. Acoust. Soc. Am. **99**, 2350–2357.

Fechner, G. T. (**1860**). *Elemente der Psychophysik* (Breitkopf und Härtel, Leipzig).

Hoemeke, K. A., and Diehl, R. L. (**1994**). ''Perception of vowel height: The role of F1-F0 distance,'' J. Acoust. Soc. Am. **96**, 661–674.

Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (**1980**). ''Discrimination of relative onset time of two-component tones by infants,'' J. Acoust. Soc. Am. **67**, 262–270.

Keating, P. A. (**1984**). ''Phonetic and phonological representation of stop consonant voicing,'' Language **60**, 286–319.

Keating, P., Linker, W. L., and Huffman, M. (**1983**). ''Patterns of allophone distribution for voiced and voiceless stops,'' J. Phonetics **11**, 277–290.

Kewley-Port, D., Watson, C. S., and Foyle, D. C. (**1988**). ''Auditory temporal acuity in relation to category boundaries: Speech and nonspeech stimuli,'' J. Acoust. Soc. Am. **83**, 1133–1145.

Kluender, K. R., and Lotto, A. J. (**1994**). ''Effects of first formant onset frequency on [-voice] judgments result from general auditory processes not specific to humans,'' J. Acoust. Soc. Am. **95**, 1044–1052.

Kuhl, P. K. (**1991**). ''Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not,'' Percept. Psychophys. **50**, 93–107.

Kuhl, P. K. (**1993**). ''Innate predispositions and the effects of experience in speech perception: The Native Language Magnet theory,'' in *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, and J. Morton (Kluwer Academic, Norwell, MA), pp. 259–274.

Kuhl, P. K. (**2000**). ''Language, mind, and the brain: Experience alters perception,'' in *The New Cognitive Neurosciences*, edited by M. S. Gazzaniga (MIT, Cambridge), pp. 99–115.

Kuhl, P. K., and Miller, J. D. (**1975**). ''Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants,'' Science **190**, 69–72.

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.

Kuhl, P. K., and Padden, D. M. (**1982**). ''Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques,'' Percept. Psychophys. **32**, 542–550.

Kuhl, P. K., and Padden, D. M. (**1983**). ''Enhanced discriminability at the phonetic boundaries for the place feature in macaques,'' J. Acoust. Soc. Am. **73**, 1003–1010.

Kuhl, P. K., Tsao, F.-M., and Liu, H.-M. (**2003**). ''Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning,'' Proc. Natl. Acad. Sci. U.S.A. **100**, 9096–9101.

Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (**1975**). ''VOT discrimination by four to six and a half month old infants from Spanish environments,'' J. Exp. Child Psychol. **20**, 215–225.

Liljencrants, J., and Lindblom, B. (**1972**). ''Numerical simulation of vowel quality systems: The role of perceptual contrast,'' Language **48**, 839–862.

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.

Lisker, L. (**1986**). '''Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees,'' Lang Speech **29**, 3–11.

Lisker, L., and Abramson, A. S. (**1964**). ''A cross-linguistic study of voicing in initial stops: Acoustical measurements,'' Word **20**, 384–422.

Lotto, A. J. (**2000**). ''Language acquisition as complex category formation,'' Phonetica **57**, 189–196.
Macmillan, N. A. (**1987**). ''Beyond the categorical/continuous distinction: A psychophysical approach to processing modes,'' in *Categorical Perception: The Groundwork of Cognition*, edited by S. Harnad (Cambridge U. P., New York), pp. 53–85.
Maddieson, I. (**1984**). *Patterns of Sound* (Cambridge U. P., Cambridge).
Miller, J. D., Wier, C. C., Pastore, R. E., Kelly, W. J., and Dooling, R. J. (**1976**). ''Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception,'' J. Acoust. Soc. Am. **60**, 410–417.
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.
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.
Molis, M. R. (**1999**). ''Perception of vowel quality in the F2/F3 plane,'' *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 191–194.
Molis, M. R., Diehl, R. L., and Jacks, A. (**1998**). ''Phonological boundaries and the spectral center of gravity,'' J. Acoust. Soc. Am. **103**, 2981.
Parker, E. M. (**1988**). ''Auditory constraints on the perception of voice-onset time: The influence of lower tone frequency on judgments of tone-onset simultaneity,'' J. Acoust. Soc. Am. **83**, 1597–1607.
Pastore, R. E. (**1988**). ''Burying straw men in imaginary graves: A reply to Kewley-Port, Watson, and Foyle (1988),'' J. Acoust. Soc. Am. **84**, 2262–2266.
Pastore, R. E., and Farrington, S. M. (**1996**). ''Measuring the difference limen for identification of order of onset for complex auditory stimuli,'' Percept. Psychophys. **58**, 510–526.
Pisoni, D. B. (**1977**). ''Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops,'' J. Acoust. Soc. Am. **61**, 1352–1361.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). ''Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants,'' J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.
Rosen, S., and Howell, P. (**1987**). ''Auditory, articulatory and learning explanations of categorical perception in speech,'' in *Categorical Perception*, edited by S. Harnad (Cambridge U. P., Cambridge), pp. 113–160.
Simos, P. G., and Molfese, D. L. (**1997**). ''Electrophysiological responses from a temporal order continuum in the newborn infant,'' Neuropsychologia **35**, 89–98.
Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (**1998**). ''MEG correlates of categorical perception of a voice onset time continuum in humans,'' Cognit. Brain. Res. **7**, 215–219.
Sinex, D., McDonald, L., and Mott, J. (**1991**). ''Neural correlates of nonmonotonic temporal acuity for voice onset time,'' J. Acoust. Soc. Am. **90**, 2441–2449.
Steinschneider, M., Volkov, I., Noh, M., Garell, P., and Howard, M. (**1999**). ''Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex,'' J. Neurophysiol. **82**, 2346–2357.
Steinschneider, M., Fishman, Y. I., Volkov, I. O., and Howard, M. A. (**2004**). ''Spectral modulation of temporal responses in human and monkey primary auditory cortex: Relevance for voice onset time (VOT) encoding,'' in *Abstracts of the Twenty-seventh Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, p. 296.
Stevens, K. N. (**1972**). ''The quantal nature of speech: Evidence from articulatory-acoustic data,'' in *Human Communication: A Unified View*, edited by E. E. J. David and P. B. Denes (McGraw–Hill, New York), pp. 51–66.
Stevens, K. N. (**1989**). ''On the quantal nature of speech,'' J. Phonetics **17**, 3–45.
Streeter, L. A. (**1976**). ''Language perception of 2-month-old infants shows effects of both innate mechanisms and experience,'' Nature (London) **259**, 39–41.
Summerfield, Q. (**1981**). ''On articulatory rate and perceptual constancy in phonetic perception,'' J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.
Trehub, S. E., and Rabinovich, M. S. (**1972**). ''Auditory-linguistic sensitivity in early infancy,'' Dev. Psychol. **6**, 74–77.
Watson, C. S., and Kewley-Port, D. (**1988**). ''Some remarks on Pastore (1988),'' J. Acoust. Soc. Am. **84**, 2266–2270.
Williams, L. (**1977**). ''The perception of stop consonant voicing by Spanish-English bilinguals,'' Percept. Psychophys. **21**, 289–297.

Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition

Mark D. Skowronski[a] and John G. Harris[b]
Computational Neuro-Engineering Laboratory, Electrical and Computer Engineering, University of Florida, Gainesville, Florida 32611

(Received 29 September 2003; revised 4 June 2004; accepted 8 June 2004)

Mel frequency cepstral coefficients (MFCC) are the most widely used speech features in automatic speech recognition systems, primarily because the coefficients fit well with the assumptions used in hidden Markov models and because of the superior noise robustness of MFCC over alternative feature sets such as linear prediction-based coefficients. The authors have recently introduced human factor cepstral coefficients (HFCC), a modification of MFCC that uses the known relationship between center frequency and critical bandwidth from human psychoacoustics to decouple filter bandwidth from filter spacing. In this work, the authors introduce a variation of HFCC called HFCC-E in which filter bandwidth is linearly scaled in order to investigate the effects of wider filter bandwidth on noise robustness. Experimental results show an increase in signal-to-noise ratio of 7 dB over traditional MFCC algorithms when filter bandwidth increases in HFCC-E. An important attribute of both HFCC and HFCC-E is that the algorithms only differ from MFCC in the filter bank coefficients: increased noise robustness using wider filters is achieved with no additional computational cost. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1777872]

PACS numbers: 43.72.Ar, 43.72.Ne, 43.72.Dv [DDO] Pages: 1774–1780

I. INTRODUCTION

Automatic speech recognition (ASR) has not found general widespread use, primarily due to the degraded performance of ASR in noisy environments. The first step in ASR is to form a compact representation of the speech signal, emphasizing phonetic information over variations due to speaker, channel, and background noise sources. The mel frequency cepstral coefficient (MFCC) front end is currently the most common, primarily because MFCC is well-suited for the assumptions of uncorrelated features used in Hidden Markov Models (HMM) to estimate the state distributions and also because of the superior noise robustness of MFCC as compared to linear prediction-based feature extractors.[1] In the work of Jankowski *et al.*, the authors suggest that the filter banks for MFCC, as well as the Seneff and EIH auditory models, are the primary elements of the algorithms which provide noise robustness, while others have shown that the linear prediction algorithm does not accurately model the zeros that additive white noise creates in the speech spectrum.[2] The bandwidth of the triangular filters in MFCC is determined by the *spacing* of filter center frequencies which is a function of sampling rate and the number of filters. That is, if the number of filters in the filter bank increases, the bandwidth of each filter decreases unintentionally.

The authors have previously introduced a novel modification to MFCC called human factor cepstral coefficients (HFCC). In HFCC, filter bandwidth is decoupled from filter spacing and is determined by the *known* relationship between center frequency and critical bandwidth of the human auditory system.[3] Using HFCC, the authors have demonstrated improved robustness over MFCC in various noise environments with filter bandwidth determined by Moore and Glasberg's critical band equivalent rectangular bandwidth (ERB) expression.[4] While HFCC uses a particular expression for filter bandwidth, the important point to note is that HFCC is part of a framework in which filter bandwidth is a free design parameter, opening up an unexplored tangent of research in ASR.

The ability to control bandwidth independent of filter spacing in filter bank design is important for two reasons: (1) it eliminates errors in bandwidth caused by extreme choices in the number of filters or filter bank frequency range, and (2) it allows optimization of bandwidth beyond Moore and Glasberg's ERB expression. Though Moore and Glasberg's ERB expression describes the critical bands of the human auditory system, HFCC is ultimately a preprocessor for an artificial classifier. Different bandwidth expressions may better condition the speech features for use in artificial classifiers. The current work considers a linear scale factor of ERB, called the E-factor, used in HFCC. Since HFCC refers to the algorithm using Moore and Glasberg's original ERB function with unity scaling factor, HFCC with E-factor is referred to as HFCC-E.

Several researchers have investigated feature extraction algorithms for automatic speech recognition that employ various filter bank designs. Hermansky's perceptual linear predictive (PLP) analysis uses a bark-scaled filter bank and cube root compression before estimating linear prediction coefficients.[5] PLP analysis uses engineering approximations to psychophysical laws, such as equal loudness preemphasis and the intensity-to-loudness power law, and admittedly al-

[a]Electronic mail: markskow@cnel.ufl.edu
[b]Electronic mail: harris@cnel.ufl.edu

 © 2004 Acoustical Society of America

lows for variations in deriving the auditory spectrum. Chan *et al.* used multiresolution analysis to design a mel-spaced wavelet filter bank.[6] The filter bank design allowed them to include Weiner filtering in the feature extraction process, and improved noise robust speech recognition was reported.

Other researchers have modified MFCC in order to improve noise robust performance. Tchorz and Kollmeier developed an auditory model-based preprocessor, similar to MFCC.[7] The static log compression in MFCC was replaced with adaptive compression. In an isolated digits experiment, the authors showed improved robustness in different noise environments compared to their model with a static log compressor, though clean-speech recognition accuracy using the adaptive compression was lower. Strope and Alwan directly modified the MFCC algorithm by adding an adaptive masker before the discrete cosine transform (DCT).[8] The masker, along with a peak-picking algorithm, improved robustness to additive noise shaped to match the long-term average spectrum of speech in an isolated digits experiment.

Research has also focused on filter bandwidth in MFCC. The 2000 SPINE evaluation of Singh *et al.* reported using filters in MFCC with double the bandwidth while maintaining filter center frequency and number of filters.[9] Improved noise-robust performance was achieved in large-vocabulary experiments, yet MFCC filter bandwidth was not investigated further. Sinha and Umesh showed improvements over baseline MFCC recognition by increasing the number of filters while maintaining the original filter bandwidth, though no explanation for the performance gain was reported.[10] Furthermore, the authors previously demonstrated improved noise robustness in MFCC by changing the *overlap* between adjacent filters.[11]

HFCC, like MFCC or PLP, is not a perceptual model of the human auditory system but rather a biologically inspired feature extraction algorithm. While MFCC and PLP draw from the theories of auditory transduction and perception laid down by psychoacousticians and neurobiologists, the algorithms concomitantly transform the speech into a representation suitable for the artificial classifiers currently used for speech recognition (hidden Markov models, dynamic time warping, or neural networks). The discrete cosine transform, an integral part of MFCC, is not biological. However, the DCT decorrelates the log energy outputs from the filter bank, strengthening the assumption of diagonal covariance matrices typically used for the Gaussian distribution of features in each state of an HMM. In this paper, the decoupled filter bandwidth in HFCC is exploited by increasing the width of the filters beyond their biologically inspired values. The experiments that follow show that HFCC-E better conditions the cepstral features for HMM classification in noisy environments.

The rest of the paper is organized as follows: the MFCC algorithm is presented, followed by a description of HFCC. Next, details for the ASR experiments using MFCC (the original as well as two modern variations) and HFCC with the isolated English digits from the TI-46 corpus are elaborated. A discussion of the ASR results and the effects wider filters have on noise robust performance concludes the paper.

II. MEL FREQUENCY CEPSTRAL COEFFICIENTS

The original MFCC algorithm introduced by Davis and Mermelstein (DM)[12] combined perceptually spaced filters with the DCT (shown to be similar to principal eigenvectors of Dutch vowels)[13] in a mainstream speech processing publication. The algorithm can be summarized as follows: a time signal is windowed, Fourier transformed to the frequency domain, and scaled by a bank of triangular filters, equally spaced on a linear-log frequency axis, and the sum of magnitude coefficients scaled by each filter is log-compressed and transformed via the DCT to cepstral coefficients. The filter bank is comprised of triangular filters, which are a coarse approximation to the shape of the critical band bandpass response of the human auditory system. The base of each triangle is determined by the center frequencies of the adjacent filters; that is, filter bandwidth in MFCC is determined by the frequency range of the filter bank as well as the number of filters in the bank.

However, coupling bandwidth to other filter bank design parameters creates two problems. As ASR experimenters adapt the algorithm to their own desired frequency range (typically set by the sampling rate of the speech corpus under study), they *unintentionally* change filter bandwidth. MFCC came to the forefront of filter bank-based speech feature extractors because of the perceptually motivated filter spacing, yet the well-known relationship between frequency and critical bandwidth of the human auditory system is not incorporated. This shortcoming is rectified in HFCC. The second problem with coupling bandwidth to the number of filters and to frequency range is that filter bandwidth is not subject to optimization with respect to experimental recognition accuracy.

The following section details the filter bank design in HFCC.

III. HUMAN FACTOR CEPSTRAL COEFFICIENT FILTER BANK

Bandwidth in HFCC is a design parameter independent of filter spacing. To determine the bandwidth, HFCC employs Moore and Glasberg's approximation of critical bandwidth, measured in equivalent rectangular bandwidth (ERB):

$$\mathrm{ERB} = 6.23 f_c^2 + 93.39 f_c + 28.52 \ \ \mathrm{Hz}, \tag{1}$$

where center frequency f_c is in kHz with the curve fit valid between 0.1 and 6.5 kHz.[14] The ERB of a bandpass filter is the width of a rectangle whose height is the maximum of the filter magnitude response and whose area is the same as the filter response. ERB is used as an alternative to 3 dB points to describe filter bandwidth. Filter center frequencies are equally spaced in mel frequency using Fant's expression relating mel frequency $\hat{f}$ to linear frequency f,[15]

$$\hat{f} = 2595 \log_{10}\left(1 + \frac{f}{700}\right). \tag{2}$$

Complete details of the HFCC filter bank construction can be found in the Appendix.

The authors have shown previously that, using Moore and Glasberg's ERB curve to define filter bandwidth, HFCC

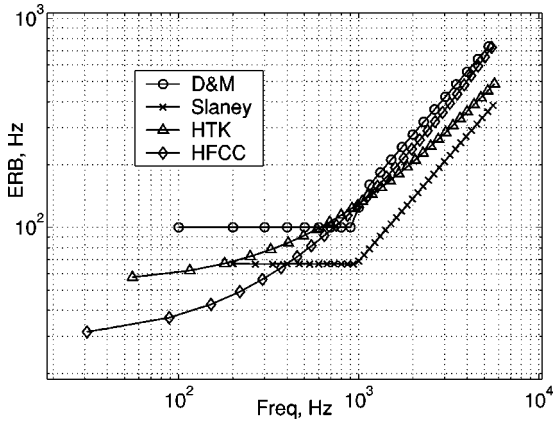


FIG. 1. ERB vs center frequency for HFCC and various MFCC implementations. Each marker represents the center frequency of a filter.

shows improved noise robustness in automatic speech recognition experiments over MFCC.[3] The current work exploits the decoupled filter bandwidth in HFCC by investigating the effects of linearly scaling the ERB bandwidth defined in Eq. (1) on noise robustness.

IV. EXPERIMENTS AND RESULTS

Filter bandwidth in HFCC-E is scaled by multiplying the ERB expression in Eq. (1) by a constant that called the ERB scale factor, or E-factor. The performance characteristics of HFCC-E with various E-factors are determined using ASR experiments in noisy environments. As a standard, three variations of MFCC common in current ASR systems are included. The next section briefly describes the three versions of MFCC before detailing the experiments.

A. Variations of MFCC

The description of the filter bank in DM's original paper has been interpreted differently by researchers using the algorithm. In the experiments, two popular versions commonly used in ASR are included along with DM's original algorithm for direct comparison with HFCC-E. The first variation is from Slaney's Auditory Toolbox for Matlab.[16] The function MFCC.m from the toolbox uses twice as many filters as in DM with a smaller spacing between center frequencies and a wider frequency range. Filters are spaced linearly below 1 kHz and log-spaced above 1 kHz. All triangular filters are equal in area.

The second variation of MFCC is from the Cambridge HMM Toolkit (HTK),[17] which is a popular C library for implementing HMMs for large vocabulary continuous speech recognition. The HTK function has many design parameters (number of filters, frequency range, vocal tract length normalization, the use of signal magnitude or square magnitude)—Young's paper is referenced for parameter values.[17] The HTK version uses Fant's definition of mel frequency, while DM and Slaney use a linear–log approximation. All filters in HTK and DM are the same height. Figure 1 summarizes the bandwidth of filters used in the three variations of MFCC and HFCC (unity E-factor), and Table I summarizes the filter bank parameters for MFCC and HFCC at a 12.5 kHz sampling rate.

TABLE I. Summary of filter bank parameters.

Method	f_{range}, Hz	# filters	Spacing	Amplitude
DM	0–6063	22	lin–log	Equal height
Slaney	133.3–5973	38	lin–log	Equal area
HTK	0–6179	29	mel warp	Equal height
HFCC	0–6250	29	mel warp	Equal height

For all three implementations of MFCC, care is taken *to preserve the original filter bandwidth* while accounting for the sampling rate of 12.5 kHz used in the reported ASR experiments. For example, DM's original sampling rate was 10 kHz.[12] They used 10 filters with center frequencies equally spaced between 100 and 1000 Hz. Above 1 kHz, center frequencies were spaced five per octave (scaling factor of $2^{1/5}\approx 1.149$) for a total of 20 filters. The next few center frequencies above 4 kHz are $4k\times 2^{1/5}\approx 4.595\,\text{kHz}$, $4k\times 2^{2/5}\approx 5.278\,\text{kHz}$, $4k\times 2^{3/5}\approx 6.063\,\text{kHz}$, and $4k\times 2^{4/5}\approx 6.964\,\text{kHz}$. The last frequency is above the Nyquist rate of 6.25 kHz, so the third frequency represents the upper frequency of the last filter. Thus, two filters are added to the DM filter bank, resulting in 22 filters.

B. TI-46 digits experiment

To characterize the performance of HFCC-E, features are extracted from the isolated English digits "zero"–"nine" from the TI-46 speech corpus in noisy environments (white and pink noise from the Noisex92 database).[18] This speech corpus offers an interesting vocabulary while requiring a classifier of only modest complexity. Clean speech recognition is nearperfect, meaning the classifiers used are adequate for the task at hand. The purpose of the experiments in this work is to demonstrate the robustness of the various features to additive noise, and larger speech databases (Aurora, Switchboard) would require further techniques whose performance would obscure the overall recognition score (endpoint detection/word spotting, prosody effects, bigram/trigram modelling, out-of-vocabulary handling, language modelling, conversational speech idiosyncrasies).

Features are extracted from an utterance by analyzing the speech with 20 ms frames at 100 frames/s. Each frame is Hamming windowed, filtered by a first-order pre-emphasis filter ($\alpha=0.95$), and the magnitude spectrum is computed and scaled by the triangular filter bank. The sum of samples scaled by each filter in the bank (output energy) is then log-compressed and transformed via the DCT to cepstral coefficients. Thirteen coefficients per frame are computed. The first cepstral coefficient is replaced by the log of the Parseval energy of the speech frame. Cepstral mean subtraction is applied,[19] and delta coefficients (± 4 frames) are computed and appended to the cepstral coefficients (26 coefficients total).

HMM word models with eight states and diagonal covariance matrices are used to classify the digits. For each of eight trials, the models are trained with clean speech from seven of the eight male speakers from the TI-46 corpus and tested with noisy speech from the remaining male speaker

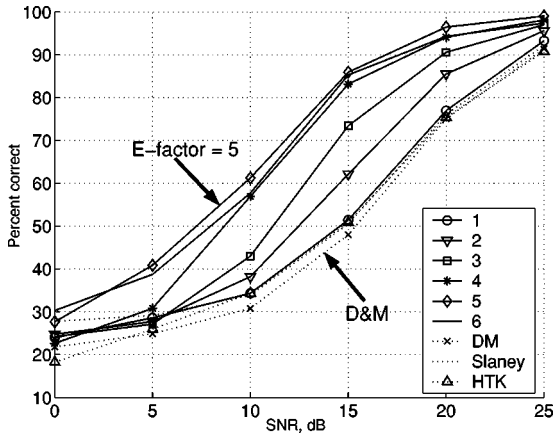


FIG. 2. Recognition accuracy vs white noise global SNR, averaged over eight trials. E-factors range from 1 to 6, with highest recognition at E-factor=5. See Table II for error bars.

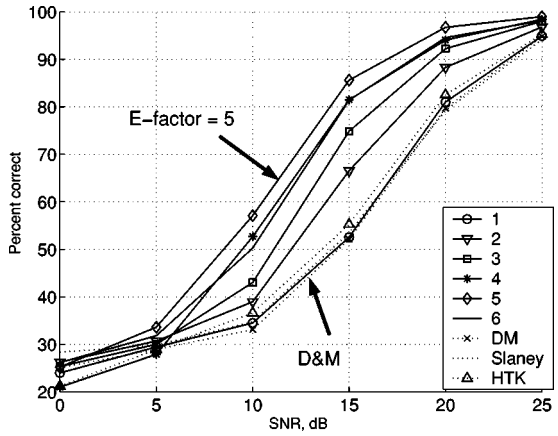


FIG. 3. Recognition accuracy vs pink noise global SNR, averaged over eight trials. E-factors range from 1 to 6, with highest recognition at E-factor=5. See Table III for error bars.

(speaker-independent experiment). The experiment trial is repeated eight times so that each of the eight male speakers is the test speaker. Preliminary experiments showed that male and female speakers are highly separable, so an accurate gender detector is assumed to be available for these experiments.[20] Separate sets of models using male and female speakers are trained and tested on the appropriate set of models after gender has been (ideally) determined. For brevity, only the results from male speakers are reported. Results using female speakers are similar, though the absolute recognition accuracy is lower for all feature extractors.

Figure 2 shows recognition accuracy vs global SNR for white noise and MFCC along with HFCC-E with various E-factors, while Table II shows the same results relative to DM for each trial. The relative results reduce the variation in recognition accuracy due to differences among test speakers. At 15 dB SNR, HFCC-E with an E-factor of 5 shows a recognition accuracy 38 ± 4 percentage points higher than DM. In the transition region between 40% and 80% correct recognition, the curve for DM is about 7 dB to the right of the curve for HFCC-E with and E-factor of 5. That is, HFCC-E with an E-factor of 5 improves recognition by about 7 dB SNR over DM.

Figure 3 and Table III show the same results for pink noise. At 15 dB SNR, HFCC-E with an E-factor of 5 shows a recognition accuracy 33 ± 3 percentage points higher than DM. In the transition region between 40% and 80% correct recognition, the curve for DM is about 6 dB to the right of the curve for HFCC-E with and E-factor of 5.

Figure 4 shows the relative performance of HFCC-E compared to DM at 15 dB SNR for several E-factors and for both white and pink additive noise. Plots for both noise sources peak at an E-factor of 5.

V. DISCUSSION

In both white and pink noise, recognition results at a given SNR increase as the E-factor increases from 1 to 5. Beyond an E-factor of 5, the results decrease. This trade-off can be better understood by considering the effects of filter bandwidth on frame SNR as well as the correlation among the filter outputs.

A. Frame SNR

Frame SNR for the *i*th frame of noisy speech is defined as follows:

$$\text{frame SNR} \equiv 10\log_{10}\left(\frac{E_X(i)}{E_N(i)}\right) \text{ dB}, \tag{3}$$

where $E_X(i)$ and $E_N(i)$ are the sum of output energies for all M filters for speech and noise, respectively, of the *i*th frame,

TABLE II. Relative performance of HFCC-E and MFCC in white noise averaged over eight trials: mean ±95% confidence interval of percentage-point difference relative to DM. E-factor ranges from 1 to 6. Global SNR of 5–20 dB. Largest difference in bold.

	SNR=5	10	15	20 dB
E-factor=1	3.7±1	3.5±1	3.5±3	1.9±2
2	2.8±2	7.4±2	14.1±5	10.4±2
3	2.3±2	12.2±3	25.4±6	15.5±5
4	5.9±4	26.1±7	35.2±6	19.0±4
5	**16.0±4**	**30.4±6**	**38.0±4**	**21.4±6**
6	13.9±2	26.9±4	37.3±3	19.2±6
Slaney	4.3±1	3.1±1	2.5±3	1.0±2
HTK	1.1±3	3.5±2	2.9±4	0.3±2

TABLE III. Relative performance of HFCC-E and MFCC in pink noise averaged over eight trials: mean ±95% confidence interval of percentage-point difference relative to DM. E-factor ranges from 1 to 6. Global SNR of 5–20 dB. Largest difference in bold. Results at 15 dB SNR plotted in Fig. 4.

	SNR=5	10	15	20 dB
E-factor=1	0.5± 0.6	1.4±2	0.4±1	1.3±2
2	1.7± 1	5.9±2	14.4±4	8.6±2
3	1.1± 1	9.9±3	22.8±5	12.6±4
4	−1.1± 2	19.6±7	29.4±5	14.4±4
5	**4.7± 3**	**24.0±7**	**33.5±3**	**17.1±6**
6	2.9± 2	17.2±4	29.3±4	14.9±5
Slaney	0.9± 0.7	1.7±1	1.3±2	−0.4±2
HTK	0± 2	3.6±1	3.1±1	0.5±2

$$E_X(i)=\frac{1}{M}\sum_{m=1}^{M}\sum_{k\in K_m}A_m(k)|X_i(k)|,$$
$$E_N(i)=\frac{1}{M}\sum_{m=1}^{M}\sum_{k\in K_m}A_m(k)|N_i(k)|. \quad (4)$$

In Eq. (4), M is the number of filters in the filter bank, A_m is the triangular function of the mth filter, K_m is the domain of the mth filter, X_i and N_i are the DFT coefficients of the speech and noise signals of the ith frame of noisy speech, respectively, and k is the DFT index.

Frame SNR is the ratio of average speech output to average noise output, where averaging is performed over the M filter outputs for a frame of noisy speech. By contrast, global SNR is defined as the ratio of the signal energy to the noise energy over the duration of the entire utterance. Figure 5 shows the frame SNR of the utterance ''seven'' for several E-factors.

For a fixed global SNR, Fig. 5 shows a steady *increase* in frame SNR in the voiced regions (/EH/, /v/, and /N/) as E-factor increases, while the opposite trend occurs for the unvoiced /s/. Wider filters pass more signal energy as well as noise energy, but the local SNR of each filter only increases as bandwidth increases if the *ratio* of the added signal to added noise is greater than the local SNR of the narrower filter. Since voiced speech is peaky relative to unvoiced speech due to the fundamental harmonics, wider filters tend to pass more harmonics, which tend to have much more signal energy than noise energy. Hence, local SNR of voiced speech increases for increasing E-factor. For the unvoiced /s/, no harmonic peaks are available to increase the added signal to added noise ratio. For the case of the utterance in Fig. 5, the added noise dominates the local SNR calculation.

In summary, wider filters increase frame SNR for voiced frames of speech while they also decrease frame SNR in other parts of the utterance (naturally, since the global SNR is fixed). Thus, wider filters emphasize higher-energy regions of the speech signal over lower-energy regions of speech, which are typically the first regions of speech corrupted by additive noise. Recognition performance in noisy environments relies primarily on high-energy peaks of the speech signal.[8] The noise robustness of HFCC-E is expected to increase as E-factor increases since wider filters increase frame SNR for voiced frames, and voiced frames typically contain high-energy peaks.

B. Pairwise correlation coefficient

The limiting factor concerning wider filters is the fact that neighboring filters overlap more. That is, as the E-factor increases in HFCC, the correlation of filter outputs increases and adjacent filters operate on increasingly similar regions of the speech spectrum. In the limit where each filter spans the

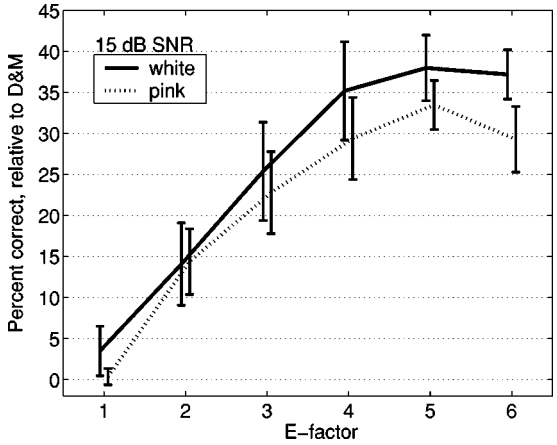


FIG. 4. Relative performance of HFCC-E compared to DM at 15 dB SNR, averaged over eight trials: mean ±95% confidence interval of percentage-point difference relative to DM. E-factor ranges from 1 to 6 for both white and pink additive noise.

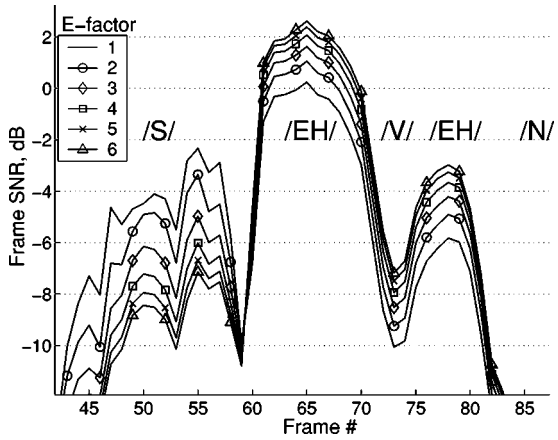


FIG. 5. Frame SNR for a typical utterance of the word ''seven'' for several E-factors. The phonetic labels denote the regions of each phoneme in the utterance. Frame SNR increases in voiced regions as E-factor increases, the opposite trend is true for the unvoiced /s/ phoneme.

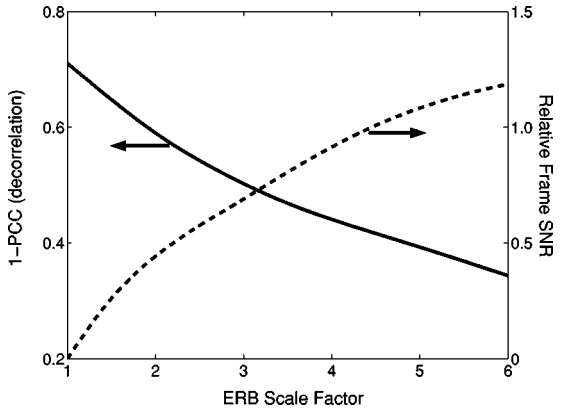


FIG. 6. Left axis: 1-PCC of filter outputs (decorrelation) vs E-factor for clean speech. Right axis: frame SNR, relative to unity E-factor, for the voiced regions in Fig. 5.

entire frequency range of interest, all filters are identical, operate on the entire speech spectrum, and output the same signal (same spectral information). Increased correlation represents a loss of spectral information, and this loss is quantified by calculating the pairwise correlation coefficient (PCC) over all pairs of filter outputs for clean speech.

PCC is defined as the average of the off-diagonal terms of the correlation coefficient matrix R_{pq}:

$$\mathrm{PCC} \equiv \frac{1}{M(M-1)} \sum_{\substack{p=1,\\ p\neq q}}^{M} \sum_{q=1}^{M} R_{pq}, \tag{5}$$

where

$$R_{pq} = \frac{C_{pq}}{\sqrt{C_{pp} C_{qq}}} \tag{6}$$

and C_{pq} is the covariance of the pth and qth filter outputs. For uncorrelated filter outputs, $R_{pq}=0$, while $R_{pq}=1$ when the outputs are identical. Thus, PCC is bounded between [0,1].

The left axis of Fig. 6 shows the expression 1-PCC, indicating that the filters become less *decorrelated* as E-factor increases. Using the same utterance of ''seven'' as in Fig. 5, Fig. 6 depicts the trade-off between the increase in frame SNR and the decrease in decorrelation as E-factor increases in HFCC-E. Both high frame SNR and high decorrelation are desired, and the E-factor is tuned to balance these two objectives for a particular experiment.

VI. CONCLUSIONS

The current work demonstrates the positive benefits of decoupling filter bandwidth from the number of filters used in the HFCC filter bank. HFCC provides a new avenue of research in feature extraction in that filter bandwidth can now be treated as an independent parameter to be optimized. Furthermore, the current work investigates one such possibility by introducing a linear ERB scale factor to increase the bandwidth of filters used in HFCC. The wider filters in HFCC-E provide increased noise robustness over the traditional MFCC algorithm: for white noise, 38 percentage points higher, or 7 dB increased SNR, and for pink noise, the increase is 33 percentage points, or 6 dB increased SNR. HFCC-E allows control over the tradeoff between frame SNR and decorrelation of the filter outputs for a particular application.

Untapped lines of research using HFCC abound: different bandwidth expressions besides ERB, such as approaches due to Patterson[21] or Zwicker,[22] adaptive E-factors, multiple classifiers trained on various E-factors with an intelligent switching mechanism during classification, filters that adaptively move to regions of high signal energy to improve local SNR, vocal tract normalization by scaling filter center frequencies, wavelet-like filters that vary in frequency *and* time to control decorrelation in both domains.

An important practical characteristic of HFCC is its simple elegance: HFCC and HFCC-E are nearly the same as MFCC, except for the coefficients that describe the filter banks. Existing MFCC code can be easily changed to HFCC or HFCC-E by simply replacing the filter bank coefficients. No additional overhead or computation is necessary for implementation.

APPENDIX A: ERB OF THE TRIANGLE FILTER

Equivalent rectangular bandwidth (ERB) is used as an alternative to the 3 dB point when describing the bandwidth of a filter. ERB is defined as the following:

$$\mathrm{ERB} = \frac{\int |H(f)|^2 \, df}{|H(f_c)|^2}, \tag{A1}$$

where $|H(f)|$ is the amplitude of the filter transfer function with peak amplitude at f_c. For HFCC and all MFCC implementations, the triangular filter bank is defined such that $|H(f)|^2$ has a triangular passband. Define f_l, f_c, and f_h as the low, center, and high frequencies of the triangular filter in linear frequency. Without loss of generality, assume $|H(f_c)| = 1$. Then

$$\mathrm{ERB} = \int_{f_l}^{f_c} \frac{f-f_l}{f_c-f_l} df + \int_{f_c}^{f_h} \frac{f-f_h}{f_c-f_h} df = \frac{1}{2}(f_h - f_l). \tag{A2}$$

APPENDIX B: FINDING THE CENTER FREQUENCIES OF THE FIRST AND LAST FILTERS

After specifying the frequency range of the filter bank between $f_{\min}$ and $f_{\max}$, the center frequencies of the first and last filters are determined by these limits. Define f_{l_i}, f_{c_i}, and f_{h_i} as the low, center, and high frequencies for the ith filter in linear frequency, respectively. The triangular filters are equilateral in mel frequency $\hat{f}$. That is,

$$\hat{f}_{c_i} = \tfrac{1}{2}(\hat{f}_{h_i} + \hat{f}_{l_i}).$$

Unwarping to linear frequency using Eq. (2) yields

$$\log\left(1 + \frac{f_{c_i}}{700}\right) = \frac{1}{2}\left(\log\left(1 + \frac{f_{h_i}}{700}\right) + \log\left(1 + \frac{f_{l_i}}{700}\right)\right),$$

$$\log\left(1 + \frac{f_{c_i}}{700}\right) = \frac{1}{2}\log\left[\left(1 + \frac{f_{h_i}}{700}\right)\left(1 + \frac{f_{l_i}}{700}\right)\right], \tag{B1}$$

$$(700 + f_{c_i})^2 = (700 + f_{h_i})(700 + f_{l_i}),$$

where, given $f_{l_1}=f_{\min}$ or $f_{h_N}=f_{\max}$, f_{h_1} or f_{l_N} can be solved for in terms of f_{c_1} or f_{c_N} respectively. Note: all linear frequencies are in Hz. From Eq. (A2) and Eq. (1),

$$af_{c_i}^2+bf_{c_i}+c=\tfrac{1}{2}(f_{h_i}-f_{l_i}) \tag{B2}$$

for f_{c_i} in Hz, where the parameters for Moore and Glasberg's ERB expression are

$$a=6.23\times10^{-6},\quad b=93.39\times10^{-3},\quad \mathrm{c}=28.52. \tag{B3}$$

Given $f_{\min}$ or $f_{\max}$, Eq. (B1) and Eq. (B2) can be written as

$$af_{c_i}^2+bf_{c_i}+c=\hat{a}f_{c_i}^2+\hat{b}f_{c_i}+\hat{c} \tag{B4}$$

or

$$f_{c_i}^2+\overline{b}f_{c_i}+\overline{c}=0, \tag{B5}$$

where

$$\overline{b}=\frac{b-\hat{b}}{a-\hat{a}},\quad \overline{\mathrm{c}}=\frac{\mathrm{c}-\hat{\mathrm{c}}}{\mathrm{a}-\hat{\mathrm{a}}}. \tag{B6}$$

For the first filter ($f_{l_1}=f_{\min}$),

$$\hat{a}=\frac{1}{2}\frac{1}{700+f_{\min}},\quad \hat{b}=\frac{700}{700+f_{\min}},$$
$$\hat{c}=-\frac{f_{\min}}{2}\left(1+\frac{700}{700+f_{\min}}\right), \tag{B7}$$

and, for the last filter ($f_{h_N}=f_{\max}$),

$$\hat{a}=-\frac{1}{2}\frac{1}{700+f_{\max}},\quad \hat{b}=-\frac{700}{700+f_{\max}},$$
$$\hat{c}=\frac{f_{\max}}{2}\left(1+\frac{700}{700+f_{\max}}\right). \tag{B8}$$

Then

$$f_{c_i}=\tfrac{1}{2}(-\overline{b}\pm\sqrt{b^2-4\overline{c}}), \tag{B9}$$

where only a + sign leads to meaningful center frequencies.

APPENDIX C: FINDING THE UPPER AND LOWER FILTER FREQUENCIES GIVEN THE CENTER FREQUENCY

From Eq. (A2),

$$f_{h_i}=f_{l_i}+2\mathrm{ERB}_i, \tag{C1}$$

where ERB_i (and f_{c_i}) have been determined. Equation (B1) yields

$$(700+f_{c_i})^2=(700+f_{l_i}+2\mathrm{ERB}_i)(700+f_{l_i}) \tag{C2}$$

which is quadratic in f_{l_i}. Solving the quadratic for f_{l_i} yields

$$f_{l_i}=-(700+\mathrm{ERB}_i)\pm\sqrt{(700+\mathrm{ERB}_i)^2+f_{c_i}(f_{c_i}+1400)}, \tag{C3}$$

where only the + sign leads to meaningful results. Equation (C1) is then used to find f_{h_i}.

[1] C. R. Jankowski, H. D. H. Vo, and R. P. Lippmann, "A comparison of signal processing front ends for automatic word recognition," IEEE Trans. Speech Audio Process. **3**, 1 (1995).
[2] Y. T. Chan, J. M. M. Lavoie, and J. B. Plan, "A parameter estimation approach to estimation of frequencies of sinusoids," IEEE Trans. Acoust., Speech, Signal Process. **29**, 214–219 (1981).
[3] M. D. Skowronski and J. G. Harris, "Human factor cepstral coefficients," J. Acoust. Soc. Am. **112**, 2279 (2002).
[4] M. D. Skowronski and J. G. Harris, "Improving the filter bank of a classic speech feature extraction algorithm," *International Symposium on Circulatory Systems* (IEEE, Bangkok, Thailand, 2003), Vol. IV, pp. 281–284.
[5] H. Hermansky, "Perceptual linear prediction (PLP) analysis for speech," J. Acoust. Soc. Am. **87**, 1738–1752 (1990).
[6] C. P. Chan, P. C. Ching, and T. Lee, "Noisy speech recognition using de-noised multiresolution analysis acoustic features," J. Acoust. Soc. Am. **110**, 2567–2574 (2001).
[7] J. Tchorz and B. Kollmeier, "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050 (1999).
[8] B. Strope and A. Alwan, "A model of dynamic auditory perception and its application to robust word recognition," IEEE Trans. Speech Audio Process. **5**, 451–464 (1997).
[9] R. Singh, M. L. Seltzer, B. Raj, and R. M. Stern, "Speech in noisy environments: robust automatic segmentation, feature extraction, and hypothesis combination," *International Conference on Acoustics, Speech, and Signal Processing* (IEEE, Salt Lake City, UT, 2001), pp. 273–276.
[10] R. Sinha and S. Umesh, "Non-uniform scaling based speaker normalization," *International Conference on Acoustics, Speech, and Signal Processing* (IEEE, Orlando, FL, 2002), pp. 589–592.
[11] M. D. Skowronski and J. G. Harris, "Increased MFCC filter bandwidth for noise-robust phoneme recognition," *International Conference on Acoustics, Speech, and Signal Processing* (IEEE, Orlando, FL, 2002), pp. 801–804.
[12] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **28**, 357–366 (1980).
[13] L. C. W. Pols, "Spectral analysis and identification of Dutch vowels in monosyllabic words," Ph.D. thesis, Free University, Amsterdam, The Netherlands, 1977.
[14] B. C. J. Moore and B. R. Glasberg, "Suggested formula for calculating auditory-filter bandwidth and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753 (1983).
[15] C. G. M. Fant, "Acoustic description and classification of phonetic units," Ericsson Technics **15**, 1 (1959); reprinted in *Speech Sound and Features*, ISBN 0262060515 (MIT Press, Cambridge, 1973).
[16] M. Slaney, Auditory Toolbox, Version 2, Technical Report No. 1998-010, Interval Research Corporation, 1998.
[17] S. Young, J. Jansen, J. Odell, D. Ollasen, and P. Woodland, *The HTK Book (version 2.0)* (Entropics Cambridge Research Lab, Cambridge, UK, 1995).
[18] A. P. Varga, H. J. M. Steeneken, M. Tomlinson, and D. Jones, "The NOISEX-92 study on the effect of additive noise on automatic speech recognition," Technical report, Speech Research Unit, Defense Research Agency, Malvern, U.K. (unpublished), http://spib.rice.edu/spib/select_noise.html (2004).
[19] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Am. **55**, 1304–1312 (1974).
[20] A. A. Dibazar, J.-S. Liaw, and T. W. Berger, "Automatic gender identification," J. Acoust. Soc. Am. **109**, 2316 (2001).
[21] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654 (1976).
[22] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer Verlag, Berlin, 1991).

Evaluation of formant-like features on an automatic vowel classification task[1]

Febe de Wet[a)]
Department of Language and Speech, University of Nijmegen, Nijmegen, The Netherlands

Katrin Weber[b)]
IDIAP—Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Switzerland and EPFL—Swiss Federal Institute of Technology, Lausanne, Switzerland

Louis Boves[c)] and Bert Cranen[d)]
Department of Language and Speech, University of Nijmegen, Nijmegen, The Netherlands

Samy Bengio[e)]
IDIAP—Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Switzerland

Hervé Bourlard[f)]
IDIAP—Dalle Molle Institute for Perceptual Artificial Intelligence, Martigny, Switzerland and EPFL—Swiss Federal Institute of Technology, Lausanne, Switzerland

(Received 20 December 2002; accepted for publication 23 April 2004)

Numerous attempts have been made to find low-dimensional, formant-related representations of speech signals that are suitable for automatic speech recognition. However, it is often not known how these features behave in comparison with true formants. The purpose of this study was to compare two sets of automatically extracted formant-like features, i.e., robust formants and HMM2 features, to hand-labeled formants. The robust formant features were derived by means of the split Levinson algorithm while the HMM2 features correspond to the frequency segmentation of speech signals obtained by two-dimensional hidden Markov models. Mel-frequency cepstral coefficients (MFCCs) were also included in the investigation as an example of state-of-the-art automatic speech recognition features. The feature sets were compared in terms of their performance on a vowel classification task. The speech data and hand-labeled formants that were used in this study are a subset of the American English vowels database presented in Hillenbrand *et al.* [J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. Classification performance was measured on the original, clean data and in noisy acoustic conditions. When using clean data, the classification performance of the formant-like features compared very well to the performance of the hand-labeled formants in a gender-dependent experiment, but was inferior to the hand-labeled formants in a gender-independent experiment. The results that were obtained in noisy acoustic conditions indicated that the formant-like features used in this study are not inherently noise robust. For clean and noisy data as well as for the gender-dependent and gender-independent experiments the MFCCs achieved the same or superior results as the formant features, but at the price of a much higher feature dimensionality. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781620]

PACS numbers: 43.72.Ne, 43.72.Ar [DOS] Pages: 1781–1792

I. INTRODUCTION

Human speech signals can be described in many different ways (Flanagan, 1972; Rabiner and Schafer, 1978). Some descriptions are directly related to speech production, while others are more suitable for investigating speech perception. Speech production is often modeled as a source signal feeding into a linear all-pole filter. In terms of this model, the phonetically relevant properties of speech signals are the resonance frequencies of the filter, also known as formants. The formant representation of speech signals is attractive because it is parsimonious yet powerful. For instance, it is well known that the frequencies of the first two or three formants are sufficient for the perceptual identification of vowels (Pols *et al.*, 1969; Flanagan, 1972; Minifie *et al.*, 1973). Many attempts have therefore been made to exploit the formant representation in speech synthesis, speech coding and automatic speech recognition (ASR).

A special reason why formants are attractive is their relation, by virtue of their very definition, to spectral maxima. In the presence of additive noise, the lower energy regions of the spectrum will tend to be masked by the noise energy, but the formant regions may stay above the noise level, even if the average signal-to-noise ratio becomes zero or negative (Hunt, 1999). The formant representation may therefore be expected to be robust against additive noise. Automatically extracted formant-like[2] features have shown some potential

[a)]Electronic mail: F.de.wet@let.kun.nl
[b)]Electronic mail: weber@idiap.ch
[c)]Electronic mail: l.boves@let.kun.nl
[d)]Electronic mail: b.cranen@let.kun.nl
[e)]Electronic mail: bengio@idiap.ch
[f)]Electronic mail: bourlard@idiap.ch

for noise robustness in automatic speech recognition, especially when combined with nonparametric spectral features (Garner and Holmes, 1998; de Wet *et al.*, 2000; Weber *et al.*, 2001a).

Despite its apparent advantages, the formant representation is not widely used in speech technology applications. In this area, nonparametric representations of speech signals are most commonly used. Even if the estimate of the spectral envelope is derived from a parametric estimator such as Linear Predictive Coding (LPC) [which can be related to the source-filter model of acoustic speech production (Markel and Gray, 1976)], speech systems avoid an explicit interpretation of the spectral envelope in terms of formants.

Given the explanatory power of the formant representation in speech production and perception research, its absence in speech technology seems awkward. One of the reasons why formants are not widely used in speech technology is that there is no one-to-one relation between the spectral maxima of an arbitrary speech signal and the resonance frequencies of the vocal tract. The exact causes of the many-to-many mapping between spectral maxima and true formants need not concern us here. What is essential is that despite numerous attempts to build accurate and reliable automatic formant extractors (e.g., Flanagan, 1972; Rabiner and Schafer, 1978; Welling and Ney, 1996; Garner and Holmes, 1998; Bazzi *et al.*, 2003), there are still no tools available that can automatically extract true formants from speech reliably. Labeling spectral maxima as formants is often only possible if the phonetic label of the sound is known, because the spectra may contain a varying number of prominent maxima (Garner and Holmes, 1998; Stevens, 1998).

The many-to-many relation between spectral maxima and true formants is not the only reason why speech technology systems avoid formant representations. Not all speech sounds are equally well suited to be described in terms of the resonance frequencies of a linear all-pole filter. Nasals and fricatives, for example, can only be accurately described if antiresonances are specified in addition to the resonances (Ladefoged, 1975; Stevens, 1998). The voice source may also contain spectral peaks and valleys that may affect the spectral peaks in the corresponding speech signals. Thus, even if it were possible to accurately and reliably label spectral maxima as formants, one would still be faced with the fact that many portions of typical speech signals show fewer spectral maxima than the number of vocal tract resonances predicted by acoustic phonetic theory. Most of the search algorithms that are used in ASR are designed to deal with feature vectors of a fixed length. Formant extractors which do not yield a fixed number of spectral peaks labeled as formants for each data frame can therefore not be used in conjunction with standard ASR search algorithms.

If it is difficult, if not impossible, to consistently and reliably extract true formants from arbitrary speech signals, the question arises whether the formant-like parameters that are delivered by one of the existing ''formant'' extraction techniques are as versatile as the true vocal tract resonances. To be useful for current ASR applications, a formant extractor must be guaranteed to deliver an equal number of formant parameters for each speech frame. Moreover, if the parameter values must have at least some relation to vocal tract resonances, they must develop smoothly over time. In this study two formant-like feature representations that fulfill both these basic requirements were investigated: two-dimensional hidden Markov models (HMM2) (Weber *et al.*, 2000) and robust formants (RFs) (Willems, 1986). The details of these techniques will be explained in Secs. II B and II C.

The best way to compare the performance of automatically extracted formant-like features and true formants would be to evaluate their performance in a real ASR system. However, all state-of-the-art ASR systems rely on very large corpora to train probabilistic models in a fully automatic manner. Obtaining corpora that are sufficiently large for ASR purposes is only feasible if no manual intervention is needed in the acoustic analysis of the signals. Due to the lack of tools to compute true formants reliably and accurately, experts are needed to add formant labels to the speech in a training database. This makes it practically impossible to provide sufficiently large training corpora for the development of formant-based processing. Yet, the theoretical attractiveness of the formant representation has motivated several attempts to overcome this hurdle.

One way to circumvent the problem that there are no databases with true formant labels that are sufficiently large to train an ASR system, is to look for another task on which the representations can be compared, and from which one might draw inferences to realistic ASR tasks. Such a task would, of course, require a suitably labeled database. One of the few corpora that does include hand-labeled formants is the *American English Vowels* (AEV) database presented in Hillenbrand *et al.* (1995). The AEV data have been used for experiments with human and automatic vowel classification, a task that is much simpler than continuous speech recognition. However, it is safe to assume that if a formant-like representation fails to approach the same vowel classification performance as the true formants in the AEV database, it is highly unlikely that such a representation could yield the theoretical advantage expected from true formants on a more realistic continuous speech recognition task.

Thus, the goal of the research reported in this paper was to investigate the degree to which formant-like features can approximate the performance of true formants in a vowel classification task, and to interpret the results in terms of the extent to which formant-like features can harness the theoretical advantages of true formants in ASR. More specifically, the aims of the research reported here are

(1) to investigate the degree to which RFs and HMM2 features resemble true formants.
(2) to compare the performance of true formants with RFs and HMM2 features on a vowel classification task. In order to strengthen the link with current research in ASR, a set of nonparametric features, i.e., mel-frequency cepstral coefficients (MFCCs), was also included in the experiments. In addition, two different classification techniques were used: Linear Discriminant Analysis (LDA) and Hidden Markov Models (HMMs). The outcome of these experiments should indicate to what ex-

tent a close relation between acoustic features and vocal tract resonance frequencies is important for automatic vowel classification.

(3) to investigate the claim that formant-like features are inherently robust against additive noise, because they are related to the spectral maxima that will stay above the local spectral level of additive noise.

The rest of this paper is organized as follows: Section II briefly introduces the AEV database, the RF algorithm, and the HMM2 feature extractor. Section III reports on the experimental setup and the results of the vowel classification experiments. The results are followed by a discussion and conclusions in Secs. IV and V, respectively.

II. DATABASE AND FORMANT EXTRACTION

A. Database of American English vowels

The speech material that was used in this study is a subset of the database of American English vowels (AEV) presented in Hillenbrand *et al.* (1995). The AEV database contains recordings of 12 vowels (/i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ʌ, ɝ, e, o/) produced in isolated /h-V-d/ syllables by 45 men, 48 women, and 46 children. Various acoustic measurements were made for each token in the database, including vowel duration, vowel steady-state times,[3] formant tracks, and fundamental frequency tracks.

To obtain the formant tracks, candidate formant peaks were first extracted from the speech data by means of a 14th-order LPC analysis. These values were subsequently edited by trained speech scientists. The formant tracks were only hand-edited between the start and end times of the vowels, i.e., the formants corresponding to the leading /h/ and trailing /d/ of the /h-V-d/ syllables were not manually labeled. Only the formant tracks corresponding to the vowel sections of the /h-V-d/ sections were therefore used in the classification experiments described in Sec. III.

Where irresolvable formant mergers occurred, Hillenbrand *et al.* put zeros into the higher of the two formant slots affected by the merger. In order to use the vowels containing mergers for our classification experiments, we replaced the zeros by the frequency value in the lower formant slot, i.e., two equal values were used. Irresolvable mergers occurred in about 4% of the vowel tokens.

In the Hillenbrand study, F1, F2, and F3 were measured for all the signals. F4 tracks were only measured if they were clearly visible in the peaks of the LPC spectrum. In 15.6% of the utterances, F4 could not be measured. For the purpose of the current investigation, we therefore decided to limit the scope of the hand-labeled formant feature set to the first three formants. In addition, we decided to use an equal number of male and female utterances and not to use the children's data. The latter decision was made because it could not be guaranteed that the two automatic formant extractors could handle children's speech appropriately.

The mean values that were measured for the first three male and female formants were all well below 4 kHz (Hillenbrand *et al.*, 1995). We therefore decided to downsample the original 16 kHz speech data to 8 kHz. Furthermore, the acoustic analyses in our experiments adhered to the same time resolution used by Hillenbrand *et al.* Specifically, all analyses used a frame rate of one frame per 8 ms. This allows a frame-to-frame comparison of the hand-labeled formants with the formant-like features generated by the two automatic extraction techniques. Finally, in keeping with what has become standard practice in ASR, the formant frequencies were mel-scaled before they were used in the classification experiments[4] (Davis and Mermelstein, 1980; Rabiner and Juang, 1993).

B. Robust formant algorithm

The robust formant (RF) algorithm was initially designed for speech coding and synthesis applications (Willems, 1986). The algorithm uses the split Levinson algorithm (SLA) to determine a fixed number of spectral maxima for each speech frame (Delsarte and Genin, 1986). Instead of directly applying a root solving procedure to the LPC polynomial, a so-called singular predictor polynomial is constructed from which the zeros are determined in an iterative procedure. The iterative procedure guarantees that the number of complex conjugate pairs of zeros is always equal to half the LPC order, provided that the order is even. Thus, the algorithm will always return the same number of parameters. Moreover, since the procedure tends to spread the zeros evenly on the unit circle, it enforces a large degree of continuity in the parameter tracks (as a function of time). After the frequency positions of the RF features have been established, their corresponding bandwidths are chosen from a predefined table such that the resulting all-pole filter minimizes the error between the predicted data and the input.

A potential disadvantage of the SLA is that it cannot handle formant mergers in a way that resembles the procedure used in Hillenbrand *et al.* (1995). Because of the tendency of the SLA to distribute poles uniformly along the unit circle, formant mergers are likely to result in one or two ''resonances'' that are shifted away (in frequency) from the true resonances of the vocal tract.

As was mentioned in the previous section, the scope of this study is limited to the frequency range between 0 and 4 kHz and to the values of the first three formants. However, in the AEV database the mean value (taken over all the relevant data) of F4 is 3.536 kHz (σ=135.5) for males and 4.159 kHz (σ=174.7) for females. This implies that, for some of the vowels produced by male speakers, the frequency band between 0 and 4 kHz may contain four vocal tract resonances instead of three. An automatic formant extraction procedure applied to the AEV data should therefore be able to deal with a potential discrepancy between the true number of formants in the signal and the requirement that only the first three formants must be returned. For the RF extractor, the simplest way to cope with this requirement is to use a sixth-order LPC analysis.[5] However, the accuracy of the LPC analysis is bound to suffer if a sixth-order analysis is used to analyze spectra with four maxima, because two complex poles are usually required to model each spectral peak (Stevens, 1998). In these cases an eighth-order LPC seems more appropriate, although it introduces the need to select three RFs from the set of four.

TABLE I. Mean Mahalanobis distance between the hand-labeled formants and the RF features.

Gender	RF3	3RF4
Male	3.5	2.1
Female	1.6	5.3
All	1.9	3.0

Given these constraints, there are a number of possible choices that can be made concerning the calculation of the RFs. We considered two of these: (1) calculate three RF features per frame (RF3); (2) calculate four RF features per frame and use only the first three (3RF4). These two sets of RF features were calculated every 8 ms over 16 ms Hamming windowed segments. We subsequently calculated the Mahalanobis distance between the hand-labeled formants (HLFs) and the RF3 and 3RF4 features, respectively. The Mahalanobis distance between two distributions is defined as (Duda *et al.*, 2001):

$$r^2=(\mathbf{x}-\mu)^t\sum{}^{-1}(\mathbf{x}-\mu). \quad (1)$$

The mean Mahalanobis distance (across all vowels) between the HLFs and the two sets of robust formants are given in Table I. The results in Table I show that the RF features are closer to the HLFs if the order of the analysis corresponds to the inherent signal structure. If there is a mismatch between the number of spectral peaks the algorithm tries to model and the number of spectral maxima that actually occur in the data, the distance between the RFs and HLFs increases. In the rest of this paper we will present results for both gender-dependent and gender-independent data sets. Because the RF3 features yielded the smallest Mahalanobis distance for the mixed data set, these will be used in the gender-independent experiments. In the gender-dependent experiments, the RF3 and 3RF4 features will be used for the female and male data, respectively.

C. The HMM2 feature extractor

HMM2 is a special mixture of hidden Markov models (HMMs), in which the emission probabilities of a conventional, temporal HMM are estimated by a secondary HMM (Weber *et al.*, 2001b). As shown in Fig. 1, one secondary HMM is associated with each state of the temporal HMM. While the conventional HMM works along the temporal dimension of speech and emits a time sequence of feature vectors, the secondary HMM works along the frequency dimension, and emits a frequency sequence of feature vectors, provided that features in the spectral domain are used.

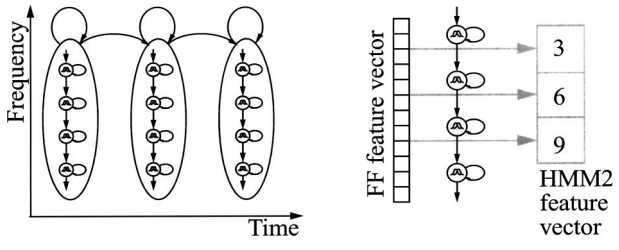


FIG. 1. Left panel: Schematic representation of an HMM2 system in the time/frequency plane. The left–right model is the temporal HMM with a top–down frequency HMM in each of its states. Right panel: Example of a temporal ''FF'' vector (left) as emitted by a frequency HMM. Each of the squares in this feature vector corresponds to a four-dimensional subvector. Gray arrows indicate the frequency positions at which transitions between the different frequency HMM states took place. The corresponding indices form an HMM2 feature vector (right).

In fact, each temporal feature vector can be seen as a sequence of subvectors. The subvectors are typically low-dimensional feature vectors, consisting of, for example, a coefficient, its first- and second-order time derivatives, and an additional frequency index (Weber *et al.*, 2001c). If such a temporal feature vector is to be emitted by a specific temporal HMM state, the associated sequence of frequency subvectors is emitted by the secondary HMM associated with the corresponding temporal HMM state. Therefore, the secondary HMMs (in the following also called frequency HMMs) are used to estimate the temporal HMM state likelihoods. In turn, the frequency HMM state likelihoods are estimated by Gaussian mixture models (GMMs). As a consequence, HMM2 can be seen as a generalization of conventional HMMs, where higher dimensional GMMs are directly used for state emission probability estimation.

Frequency filtered filterbanks (FFs) (Nadeu, 1999) are typically used as features for HMM2, because they are comparatively decorrelated in the spectral domain. In certain ASR tasks, the baseline performance of the FF coefficients has been shown to be comparable to that of other widely used state-of-the-art features such as MFCCs (Nadeu, 1999). For the HMM2 systems that were used in this study, a sequence of 12 FF coefficients was calculated every 8 ms. While a larger number of FF coefficients could possibly be advantageous, this number was chosen in order to make the number of features used for HMM2 comparable to that conventionally used in HMMs. Together with their first- and second-order time derivatives plus an additional frequency index, these FF coefficients form a sequence of 12 four-dimensional subvectors. Each square in the vector labeled ''FF feature vector'' in Fig. 1 therefore represents a four-dimensional subvector.

Speech recognition with HMM2 can be done with the Viterbi algorithm, delivering (as a by-product) the segmentation of the signal in time as well as in frequency. The frequency segmentation of one temporal feature vector reflects its partitioning into frequency bands of similar energy. Supposing that certain frequency HMM states model frequency bands with high energy (i.e., formant-like regions) and others those bands with low energies, the Viterbi frequency segmentation could be interpreted as an alternative way to represent formant-like structures.

For each temporal feature vector, we determined from the Viterbi segmentation at which point in frequency (i.e., between which subvectors) a transition from one frequency HMM state to the next took place. For example, in Fig. 1 the first HMM2 feature vector coefficient is 3, indicating that the transition from the first to the second frequency HMM state occurred before the third subvector. In the case of four frequency HMM states connected in a top–down topology (as

seen in Fig. 1), we therefore obtain three integer indices (corresponding to precise frequency values). In our classification experiments, these indices were used as three-dimensional feature vectors in a conventional HMM.

HMM2 design options

The design of an HMM2 system can vary substantially, depending, for example, on the task and on the data to model. There are a number of design options which determine the performance of an HMM2 system. These include issues like *model topology* (which needs to be considered both in the time and the frequency dimension), the addition of *frequency indices*, different *initialization* possibilities, as well as different (combinations of) segmentation strategies that can be applied for *training and test* purposes. These design options are discussed in detail in Weber (2003).

The models that were used to obtain the results reported on in Sec. III all had a three-state, left–right topology in the time domain and a four-state top–down topology in the frequency domain. Frequency indices were included as additional feature components in the frequency subvectors. The initialization of the gender-independent HMM2 models was based on the assumption of alternating high and low energy frequency HMM states. The gender-dependent models were initialized according to the hand-labeled formant frequencies' segmentation. The HMM2 features that were used for training were obtained by means of forced alignment while those that were used for testing were obtained from a free recognition. Training and testing were done with HTK (Young *et al.*, 1997) and the HMM2 systems were realized as a large, unfolded HMM, which is possible when introducing synchronization constraints (Weber *et al.*, 2001b).

Finally, it should be pointed out that results from a previous study have shown that adding first-order time derivatives does not improve the classification performance of HMM2 features on the AEV database (Weber *et al.*, 2002). In that study, it was argued that this result can be attributed to the nature of the AEV data, exhibiting only very few spectral changes (see Sec. III A 2 for a graphical illustration), in conjunction with the very crude nature of the HMM2 features. Often, the frequency segmentation of one phoneme would be the same for all time steps, resulting in zero-valued time derivatives. In other cases, oscillations between two neighboring segmentations were observed, which gave equally meaningless derivatives.

III. EXPERIMENTS AND RESULTS

In the following, the design, execution, and results of the vowel classification experiments are described. In Sec. III A, the first question posed in Sec. I is addressed, i.e., to what extent the features yielded by the two automatic formant extractors resemble the hand-labeled formants in the AEV database. The design of the classification experiments is subsequently described in Sec. III B. Section III C reports on the Linear Discriminant Analysis (LDA) classification results. The LDA experiments enable us to relate our results to those reported in Hillenbrand *et al.* (1995). The results of the HMM classification experiments are presented in Sec. III D.

TABLE II. Mean Mahalanobis distance between the hand-labeled formants, RFs and HMM2 features.

Gender	RF	HMM2
Male	2.1	8.0
Female	1.6	9.1
All	1.9	5.6

The HMM experiments were conducted in order to determine whether the classification performance of hand-labeled formants with LDA generalizes to the classification performance obtained with the maximum likelihood (ML) procedures that are dominant in the ASR community. Finally, Sec. III E reports on the classification performance of the automatically extracted formant-like features in (simulated) noisy acoustic conditions.

A. How formant-like are RFs and HMM2 features?

There are no generally accepted procedures to assess the degree to which formant-like features resemble true formants. In this study we approached the problem in two complementary ways: by means of a formal distance measure that captures the goodness-of-fit in a single measure, and by means of a graphical illustration of the physical nature of the differences that underlie the summary measures.

1. Statistical distance

The Mahalanobis distance was introduced in Sec. II B as a means to select the RF feature sets that were closest to their hand-labeled counterparts, in terms of statistical distance. The minimum values from Table I are repeated in Table II, together with the mean Mahalanobis distances between the HLFs and the HMM2 features. The values in Table II clearly indicate that, in terms of statistical distance, the RFs are more similar to the HLFs than the HMM2 features.

2. Graphical illustration

Some of the issues involved in comparing HLFs, RFs, and HMM2 features can be illustrated by means of a typical example, in the form of a representative token of the vowel /ɜ˞/. Figure 2 shows the HLF tracks corresponding to a female pronunciation of the vowel overlaid on a spectrogram representing the frequency range between 0 and 4000 Hz.

The same example was used to create the graphs in Fig. 3. In each of the subplots in Fig. 3, the *y* axis corresponds to frequency index, the *x* axis to time, and darker shades of gray to higher intensity levels. Figure 3(a) shows the same HLFs as in Fig. 2, but overlaid on the mel-weighted log-energy within each frame. The mel-scaled filterbank that was used to obtain the energy values consisted of 14 filters that were linearly spaced in the mel frequency domain between 0 and

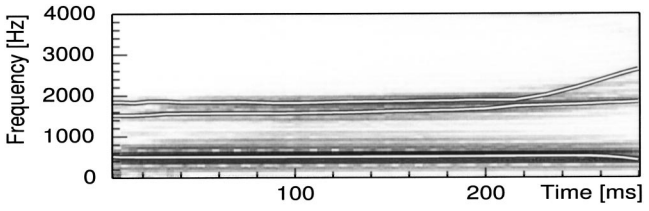


FIG. 2. HLF tracks corresponding to a female pronunciation of the vowel /ɜ˞/.

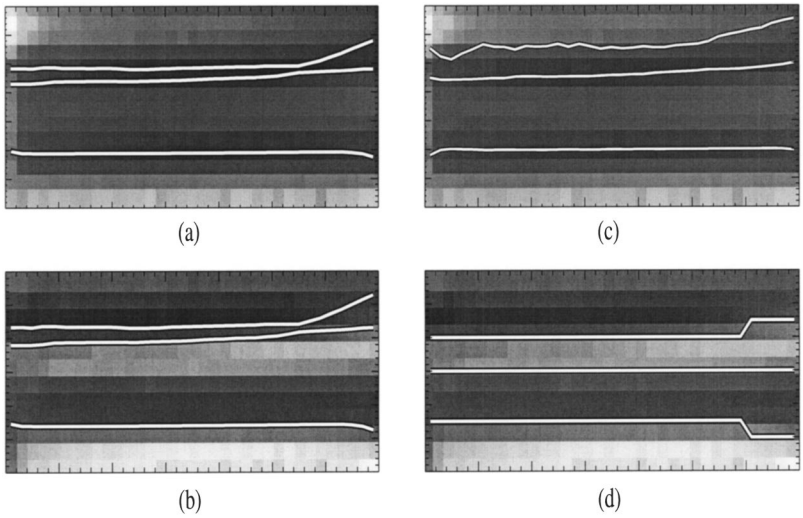


FIG. 3. Feature tracks corresponding to a female pronunciation of the vowel /ɝ/: (a) HLFs overlaid on the mel-scaled log-energy of each frame, (b) the same HLFs on the corresponding FF features; (c) RFs on the mel-scaled log-energy of each frame, and (d) HMM2 feature tracks on FF features.

2146 mel (corresponding to the frequency range between 0 and 4000 Hz, as in Fig. 2). Figure 3(b) shows the HLFs overlaid on 12 FF features, which were derived form the 14 filterbank values, and which were used to train the HMM2 feature extractor. It can be seen that, while the HLFs follow spectral maxima in the filterbank domain, they are positioned at the transitions from low to high intensity regions in the FF domain. Figure 3(c) shows the tracks of the RF features overlaid on the mel-scaled filterbank features, while Fig. 3(d) shows the HMM2 feature tracks overlaid on FF features.

The data in Fig. 3 show that the RF feature tracks are fairly similar to the HLFs. Most importantly, there are no obvious examples of missing formants or wrong labels. The RF features exhibit more frame-to-frame variation than their hand-labeled counterparts. In this example, the LPC spectrum of the vowel contained multiple peaks in the F2–F3 region, while the human labelers consistently preferred a peak at a lower frequency than the RF procedure. We have not been able to verify whether this type of frame-to-frame variation is related to those parts of the vowels in which the human labelers found it most difficult to find the ''correct'' spectral peaks. It is also not clear whether this variation has affected the classification performance of the RF features, relative to the more smooth HLF features. During normal human speech the articulators move relatively slowly. The smooth HLF feature tracks therefore seem to be more plausible than the slightly more ''noisy'' RF features. The short-term variations in the RF features are the result of the attempt of the low-order LPC analysis to account for the spectral envelope in the original acoustic signal, which is not only determined by the vocal tract resonances, but also by the excitation. For the RF extractor to yield feature tracks as smooth as the HLF an additional smoother would have to be applied to the raw RF values.

The HMM2 features are very crude and do not resemble either the HLF or the RF tracks. The crudeness is due to the fact that the HMM2 features are derived from 12 FF features, instead of spectral envelopes sampled at multiple equidistant frequencies. However, the feature tracks in Fig. 3(d) indicate that, for the example utterance illustrated in the figure, the HMM2 method succeeded in separating high from low intensity regions in the FF domain. While the first and third HMM2 feature tracks are roughly situated near formant positions (corresponding to the transition between low and high intensity in the FF domain, and to spectral peaks in the spectrogram), the HMM2 track in the middle can be supposed to correspond to a spectral valley. General trends present in the signal (such as the upward tendency for the highest formant at the end of the vowel) are also reflected by the HMM2 tracks.

For other data examples, unexpected transitions and oscillations of the HMM2 feature tracks were also observed. These effects are explained in more detail in Weber (2003). However, for most examples, one or two HMM2 feature tracks correspond to a certain degree (given their low accuracy, which is limited by the low frequency resolution of the FF features) to the HLFs, while another one frequently corresponds to a spectral valley.

B. Experimental setup

Given the fact that the AEV database is quite small, a three-fold cross-validation was used for the classification experiments. The classifiers (LDA and HMM) were trained on two subsets of the data, and tested on the third one. Thus, each experiment consisted of a number of independent tests. Moreover, all tests were performed in two conditions, gender-independent and gender-dependent. The gender-independent data sets were defined as three nonoverlapping train/test sets, each containing the vowel data of 60(train)/30(test) speakers, with an equal number of males and females in each set. For the gender-dependent data, three independent train/test sets were defined for males and females separately. Each train/test set consisted of 30(train)/15(test) speakers. For the gender-independent data sets, the classification results reported in the following correspond to the mean value of the three independent tests. The gender-dependent results were obtained by averaging the classification results of six independent experiments (three male and three female).

Five different feature sets were used to conduct the vowel classification experiments, i.e., hand-labeled formants

TABLE III. LDA classification results (% correct): gender-independent data.

Feature type	stst	20%80%	20%stst80%
Hillenbrand	81.0	91.6	91.8
HLF	77.0 (±2.5)	91.4 (±1.7)	91.9 (±1.6)
RF	63.4	81.8	83.0
HMM2	31.7	48.7	52.2
MFCC12	73.1	90.5	91.2
MFCC-LDA3	67.3	88.3	90.1

(HLFs), robust formants (RFs), HMM2 features, and two sets of mel-frequency cepstral coefficients (MFCCs). The MFCCs were included as an example of acoustic features that are commonly used in ASR applications. MFCCs describe the spectral envelope in a small number of orthogonal coefficients (Davis and Mermelstein, 1980; Rabiner and Juang, 1993). Usually, 10 to 15 MFCCs are needed to obtain a sufficiently accurate description of the spectrum. The first set of MFCCs that was used in this study, MFCC12, consisted of 12 MFCCs ($c_1 \ldots c_{12}$).[6] However, the MFCC12 feature set contains four times as many coefficients as the HLF, RF, and HMM2 representations. We therefore decided to create a three-dimensional MFCC set, MFCC-LDA3, by projecting the twelve-dimensional MFCCs into a three-dimensional feature space. In order to accomplish the transformation, an appropriate transformation matrix was derived from the relevant training data by means of LDA.

C. LDA classification results

This section reports on an experiment that compares the performance of RFs, HMM2, and MFCC features to the performance of HLF features on a task that is very similar to the one described in Hillenbrand *et al.* (1995). In contrast with the original study, we used an LDA [instead of quadratic discriminant analysis (QDA)], we included all vowels,[7] and we used only the adult speakers' data. To maintain the equivalence between the LDA experiments described here and the corresponding experiments with HMMs that are described in Sec. III D, we used the three-fold cross-validation scheme described in Sec. III B for training and testing (instead of a leave-1-out jackknifing procedure). As in Hillenbrand's study, we investigated classification performance for a single set of formant values determined in the vowel steady state (stst), pairs of formant values measured at 20% and 80% of the vowel duration (20%80%), and triplets in which the steady state value was added to the values at 20% and 80% of the vowel duration (20%stst80%).

The classification rates obtained for the gender-independent data are given in Table III and those for the gender-dependent data in Table IV. Table III also contains the results from the QDA experiments reported in Hillenbrand *et al.* (1995). The results show that our results for the HLF features, obtained with a simpler discriminant analysis technique, are very close to Hillenbrand's results. Human classification for the same data (based on the complete /h-V-d/ utterances) was 95.4% correct (Hillenbrand *et al.*, 1995). The results in Tables III and IV indicate that the vowel classes can be separated reasonably well (in comparison with human performance) by the steady state values of their first three formants. Information about patterns of spectral change clearly enhances the distinction between classes.

TABLE IV. LDA classification results (% correct): gender-dependent data.

Feature type	stst	20%80%	20%stst80%
HLF	79.4 (±2.4)	93.6 (±1.5)	93.8 (±1.4)
RF	76.1	91.2	92.0
HMM2	48.5	60.1	63.8
MFCC12	81.7	94.5	94.2
MFCC-LDA3	73.9	92.3	93.5

As our goal was to compare the performance of the HLF features with that of the other features, the 95% confidence intervals corresponding to the HLF results are indicated in parentheses. The values in Tables III and IV show that, with the exception of the MFCC12 features, the HLF features outperform all the other features in terms of vowel classification rate. The difference between HLF and the other results is much larger for the gender-independent experiments than for the gender-dependent experiments. This difference is especially evident for the RF features: for the gender-independent experiments the HLF features outperform the RF features by more than 10% (absolute), whereas the corresponding difference for the gender-dependent experiments is less than 3% (absolute).

The data in Tables III and IV also show that the classification performance of the HMM2 features is substantially lower than the results obtained for the other feature sets. This observation indicates that the vowel classes are not linearly separable given these features at just one, two, or three different instances in time. While the HMM2 features at any given moment may not be sufficient to discriminate between the vowel classes, the additional information required to do so may be provided by a complete temporal sequence of HMM2 features. This presupposition will be investigated in the following section within the framework of HMM recognition.

The MFCC12 features achieve classification rates that compare very well with those of the HLF features. Although they perform slightly better than the HLF features in the gender-dependent experiments, this difference is not significant. This result indicates that, for the current vowel classification task, three HLF features and 12 MFCCs are equally able to discriminate between the vowel classes. The three-dimensional MFCCs outperform both the RFs and the HMM2 features and their classification performance is only slightly inferior to the classification rate achieved by the MFCC12 features.

D. HMM classification rates on clean data

The classification rates in Tables III and IV were obtained by means of a LDA. In discriminative training algorithms such as LDA, the aim of the optimization function is to achieve maximum class separability by finding optimal decision surfaces between the data of the different classes. However, the recognition engines of most state-of-the-art ASR systems are trained using a ML optimization criterion. The training algorithms therefore learn the distribution of the data without paying particular attention to the boundaries

TABLE V. HMM classification results (% correct) for gender-independent and gender-dependent data.

Feature type	Gender-independent	Gender-dependent	Feature dimension
HLF	87.7 (±2)	89.6 (±1.8)	6
RF	84.1	90.5	6
HMM2	77.0	87.2	3
MFCC13	92.3	92.1	26
MFCC-LDA3	79.9	81.6	6

between the different data classes. Although discriminative training procedures have been developed for ASR, they are not as commonly used as their more straightforward ML counterparts (e.g., Juang *et al.*, 1996). The LDA classification described in Sec. III C also required a time-domain segmentation of the data. In real-world applications this kind of information will not be available. The aim of the next experiment is therefore to evaluate the classification performance of the different feature sets using HMMs that were derived by means of ML training.

Toward this aim, we compared the vowel classification rates achieved by the five feature sets used in the LDA experiments. With the exception of the HMM2 features, the first-order time derivatives of all the features were also included in the acoustic feature vectors. Since in mainstream ASR it is usual to add overall energy to MFCC features, we extended the MFCC12 vectors to MFCC13 by adding c_0. The resulting feature vector dimensions for the HLF, RF, HMM2, MFCC13, and MFCC-LDA3 features were therefore 6, 6, 3, 26, and 6.

Classification experiments were conducted using both the gender-independent and the gender-dependent data sets defined in Sec. III B. For each of the vowels in the AEV database and for each acoustic feature/data set combination, a three state HMM was trained. The EM algorithm implemented in HTK was used for the ML training (Young *et al.*, 1997). Each HMM state consisted of a mixture of ten continuous density Gaussian distributions. The results of the classification experiments are shown in Table V. Once again, the 95% confidence intervals corresponding to the HLF results are indicated in parentheses. The values in the last column of Table V correspond to the dimensions of the different feature sets.

According to the results in Table V, the HLF features consistently achieved classification rates of almost 90% correct. Even though these values are significantly lower than those measured in the LDA experiments, they do indicate that, in principle, the HLF features are suitable to be used as features in combination with state-of-the-art ASR methods, i.e., using HMMs, ML training, and Viterbi classification.

A remarkable difference between the LDA and HMM experiments is the excellent classification rate achieved by the HMM2 features: these features perform much better in combination with HMMs than with LDA. Table V shows that, for the gender-dependent data, the HMM2 features not only outperform the MFCC-LDA3s but also approximate the performance of the HLF and RF features, in spite of their lower feature dimensionality.

The data in Table V also show that, for the current vowel classification task, HLF features compare very well with MFCCs. Although the MFCC13 features outperform their HLF counterparts on both gender-independent and gender-dependent data, this is at the price of a much higher feature dimensionality. MFCCs with the same dimension (MFCC-LDA3) perform significantly worse than both MFCC13 and HLF. In contrast to what was observed for the LDA experiments, the RFs and HMM2 features also perform much better in comparison with the MFCC-LDA3 features.

A comparison between the gender-independent and gender-dependent results shows that, in general, the gender-dependent systems work better, even in the case of HLF features. This observation is in good agreement with the results of the LDA experiments. Another similarity between the HMM and LDA results is the fact that the classification performance of the automatically extracted formant-like features are especially gender-dependent. Although not to the same extent as the formant-like features, the performance of the MFCC-LDA3 features is also enhanced by using gender-specific modeling. Only the performance of the MFCC13 features seems to be insensitive to gender differences. The MFCC13 features are probably less sensitive to the gender-dependent properties of the data because, in addition to information on the formants, they also contain information about spectral level and general spectral shape.

E. HMM classification rates on noisy data

In this experiment, the models trained on the MFCC13, RF, and HMM2 features that were used for the experiments described in Sec. III D, were tested in noise. The HLF features could not be included in this experiment, because it was not possible to obtain hand-labeled formants for the noisy data. The models were trained on clean data only and noisy acoustic conditions were simulated by artificially adding babble and factory noise to the test data at SNRs of 18, 12, 6, and 0 dB. The babble and factory noise were both taken from the Noisex CD (Noisex, 1990). The Noisex babble noise contains speech from many different people speaking simultaneously and individual speakers and utterances cannot be discerned from the hubbub. As a result, the signal power is fairly constant and the long-term spectrum is quite flat. The long-term spectrum of the Noisex factory noise also does not exhibit any significant peaks. However, the factory noise is not stationary; it contains a number of hammer blows and other noise bursts.

Figure 4 gives an overview of the classification performance of gender-dependent models tested in noise. Classification rate is shown as a function of SNR for both babble and factory noise. Similar, but slightly inferior, results were obtained for the gender-independent models. (These results are not shown here.)

In Sec. I it was argued that, in the presence of additive noise, the lower energy regions in speech spectra will tend to be masked by the noise energy, but that the formant regions (spectral maxima) may stay above the noise level, even if the average signal-to-noise ratio becomes zero or negative. This line of reasoning gave rise to the hypothesis that a representation in terms of formants or formant-like features should be comparatively robust against additive noise. However, the

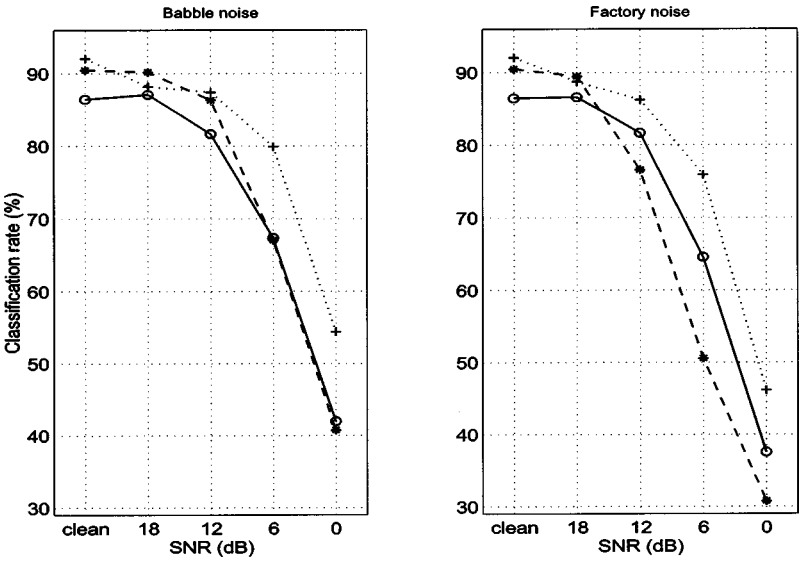


FIG. 4. Average classification rates (% correct) for gender-dependent models trained on clean MFCC13 (+), RF (∗), and HMM2 (○) features and tested in babble (left panel) and factory (right panel) noise. The corresponding feature vector dimensions are 26 (MFCC13), 6 (RF), and 3 (HMM2).

results in Fig. 4 do not support this hypothesis. In fact, the figure shows that the recognition performance of all three systems deteriorates in noise. While the performance of the different features is comparable at SNRs of 18 dB and higher, at lower SNRs the performance degradation of the MFCC13 features seems less severe than that of the formant-related features. To a certain extent, this result may be explained by the fact that the MFCC13 system has a total of 26 feature components at its disposal, while the dimensionality of the RF and HMM2 systems is restricted to 6 and 3, respectively. The higher-order MFCCs—which may contain redundant information in clean conditions—seem to be better at maintaining system performance in adverse acoustic conditions. However, no analysis was done to determine to what extent the errors made when using the different features are complementary. It was also not investigated whether classification performance could be improved by using a combination of different feature streams.

For all three systems the drop in recognition rate is more severe in factory noise than in babble noise. Factory noise also seems to affect the RF features more than HMM2. The type of performance degradation shown in Fig. 4 is equivalent to results obtained for other databases in comparable simulations of noisy conditions (e.g., de Wet *et al.*, 2000).

In principle, the argument that spectral maxima may stay above the noise level seems to be plausible. However, the RF features (which are supposed to model spectral maxima of an all-pole signal) clearly fail in noisy acoustic conditions. This observation suggests that the RF algorithm is ''misled'' by the added noise, such that it is no longer capable to find the spectral maxima that correspond to the formants. The noisier the signal, the more the all-pole character of the speech signal disappears. Consequently, the fixed order all-pole model of the RF-algorithm is no longer able to estimate the parameters of the underlying speech production system, and the RF-extractor is turned into a parametric estimator of the peaks in a spectral envelope, the details of which are increasingly determined by the noise.

The failure of the HMM2 system at low SNRs may be explained as follows: for heavily degraded speech, the number of recognition errors made by the HMM recognizer embedded in the feature extractor is bound to increase. As a result, the corresponding HMM2 features will be calculated by the ''wrong'' HMM2 feature extractor, i.e., the HMM2 model corresponding to the wrong phoneme will give the best likelihood score and will therefore be chosen for feature extraction. Recognition errors made by the HMM2 feature extractor and the conventional HMM recognizer (which uses the erroneous HMM2 features) accumulate, which will forcibly lead to severe degradations at low SNRs.

IV. DISCUSSION

One of the aims of this study was to investigate the degree to which RFs and HMM2 features resemble true formants. The statistical distances and graphical illustrations provided in Sec. III A showed that, of the two automatically extracted formant-like feature sets, the RFs are more similar to the HLFs than the HMM2 features. In fact, the automatically extracted RF features resembled the HLF features quite closely, provided that the RF algorithm was given prior information about the gender of the speaker. This information helps the RF algorithm to avoid one of the most important errors in automatic formant assignment, i.e., labeling spurious peaks as formants, with the results that all higher-order formants in the frame are labeled incorrectly.

Although HMM2 can, in principle, be used as an estimator of true formants, the implementation of HMM2 that was used in this study is not a formant extractor in the classical sense. Because the HMM2 features were derived from a 12-parameter frequency filtered filterbank, they are inherently very coarse. However, the coarse quantization of the HMM2 features is not an intrinsic limitation of this approach to the representation of spectral envelopes. Rather, it is one of the implications of the way in which the current version of HMM2 has been implemented. Other implementations, which use filters with much narrower pass bands than the 14 critical band filters used in this study, should be investigated.

In our comparison of the performance of true formants, RFs, and HMM2 features on a vowel classification task, the

following observations were made. For the gender-dependent data the overall classification performance obtained for the 20%stst80% condition with LDA is better than the results of the HMM classifiers. For the gender-independent data the difference is not equally clear. Apparently, removing the overlap between different vowels from males and females helps the LDA to find an optimal class separation. The ML classifier implemented by the HMMs seems to be less powerful in this regard.

The most salient difference between the LDA and HMM results concerns the classification rates that were obtained for the HMM2 features. While the HMM2 results for the HMM classifier are comparable with the corresponding HLF results, the LDA classifier does not seem to be able to distinguish between the vowel classes if it is trained on HMM2 features. This result indicates that it is not possible to distinguish between the vowel classes in the coarsely quantized HMM2 feature space when only a few points (in time) are taken into consideration. Due to the coarseness of the HMM2 features, HMM2 feature tracks may change rather abruptly at any point in time. For example, an abrupt change may occur before the 20% duration point for some pronunciations of a certain phoneme and after the 20% duration point for other pronunciations of the same phoneme. The LDA classifier does not seem to be able to deal with these differences. The HMM classifier, on the other hand, is able to handle these changes in the data because it classifies vowels in terms of a complete temporal sequence of HMM2 features.

In both the LDA and the HMM classification experiments, the classification rates measured for the gender-dependent data sets were higher than the corresponding results for the gender-independent data sets. Classification performance is determined by two factors, i.e., the degree of estimation noise in the features and the overlap between the vowels in the feature space. The observation that the automatically extracted formant-like features generally yielded much better results for the gender-dependent data sets may be explained by the fact that the vowel classes are better separated in a gender-dependent feature space. However, the RF and HMM2 features clearly benefit more from the gender separation than the HLF and MFCC features. This suggests that, for the RF and HMM2 features, the gender separation also achieved a certain degree of reduction in estimation noise in the features themselves.

The classification experiments also showed that the difference between the gender-dependent and gender-independent results was much smaller for the HLF features than for the other feature representations. This observation can probably be explained by the fact that the human labelers knew the gender of the speakers. The labelers also knew the identity of the tokens while they were assigning the formant labels. This gives the HLF features another advantage over the automatically derived features: these either rely on imperfect classification results (in the case of HMM2) or have no knowledge about the token for which feature extraction is attempted (in the case of the RF features). However, a comparison of the results obtained with HLF and gender-dependent RF features suggests that, for the vowel classification task investigated in this study, the advantage of expert knowledge is rather small when the gender of the speakers is taken into account by the automatic feature extraction procedures. This observation may not generalize to other databases. Especially in fluent, continuous speech the phonetic context of the vowels will be richer and have a bigger impact on the spectral envelopes. After all, the /h-V-d/ context was chosen to minimize coarticulation effects, which will be especially cumbersome for automatic (and manual) formant extraction, e.g., in the case of nasal consonants.

A comparison of the classification performance of HLFs and RFs for the LDA and HMM experiments, and of HLFs and HMM2 features for the HMM experiments, suggests that features that are directly related to vocal tract resonances have very few advantages over formant-like features, as long as the measurement errors in the different feature types are comparable. Especially the results obtained with the HMM2 features, which definitely do not represent formants in the sense of vocal tract resonances, suggest that consistency (including smoothness of the feature tracks over time) is more important than the relation to the underlying, physical speech production process. This result suggests that the formant extraction technique that was recently proposed in Bazzi *et al.* (2003), which guarantees a fixed number of formant values for each frame as well as smooth feature tracks over time, would be a viable candidate to deliver formant-like features that can be used in ASR.

Finally, the results in Sec. III E show that the formant-like features that were investigated in this study are not inherently robust against additive noise. Neither the RFs nor the HMM2 features were able to keep track of the spectral maxima that should remain intact in noisy speech data. For the use of formants in ASR the message appears to be that the theoretical advantages of the formant representation are neutralized by the enormous difficulty of building a reliable automatic formant extractor, especially one that is also able to process noisy speech. The theoretical advantages of the formant concept for processing noisy speech can only be harnessed by signal processing techniques that take full profit of continuity and coherence in the signals, both in time and in frequency.

The relative success of adding formant candidates to MFCC parameters in the work of Holmes *et al.* (1997) suggests that a feasible alternative would be to address formant extraction and ASR simultaneously. Hypotheses about formant values should be conditioned by phone observation probabilities, because knowledge of the recognized sound is a powerful knowledge source to guide the classification of spectral peaks as formants. At the same time, an interpretation of the signal in terms of sounds and words that makes sense against the background of formant candidates should result in more accurate ASR than one that does not. This suggests that, for a formant representation to have its maximum impact on ASR, it is not just the signal processing and feature extraction that must be advanced. Major advances in the search and decision processes that eventually link features to words, meanings, and intentions are also required.

V. CONCLUSIONS

In this paper, a number of issues related to the use and usefulness of the formant concept in ASR were investigated. Because there are no databases available that contain enough true formant data to train ASR systems, we focused on the AEV database introduced in Hillenbrand *et al.* (1995).

The first conclusion that can be drawn from our data is that, of the two automatic formant extraction techniques under investigation, robust formants did approximate hand-labeled formants rather closely, provided that the RF algorithm had prior knowledge of the speaker gender. The HMM2 features, on the other hand, did not resemble vocal tract resonances.

Second, for the automatic classification of vowels, we found little advantage in using acoustic features that have a direct relation to vocal tract resonances. If the features are consistent and feature tracks are smooth, their performance can approximate that of true, hand-labeled formants.

Third, the theoretical robustness of formant measures against additive noise could not be verified for either of the two automatically extracted, formant-like feature sets. Background noise seems to introduce additional spectral peaks in the spectral envelopes, which cannot be effectively discarded as formant candidates by the relatively simple signal processing techniques underlying RF extraction and HMM2 feature computation.

In summary, it seems fair to say that, for the clean experimental conditions that were studied in this investigation, the formant representation of speech signals has no compelling advantages (when used as a conventional feature set) over representations that do not involve error-prone labeling decisions such as MFCCs. In noisy conditions, we found that the theoretical advantages of the formant concept were vastly diminished by the failure of our signal processing techniques to reliably distinguish between spectral maxima that must be attributed to vocal tract resonances and maxima that are introduced by the noise.

ACKNOWLEDGMENTS

We would like to thank Professor James Hillenbrand for making the AEV database available to us and for his swift reply to all our enquiries. The development of the database was supported by a research grant from the American National Institutes of Health (NIDCD 1-R01-DC01661). F.d.W.'s visit to IDIAP was made possible by the I.B.M. Frye grant. K.W. was funded through the Swiss National Science Foundation, Project No. FN 2000-059169.99/1.

[1]Some of the experimental results reported in this study were presented in ''Evaluation of formant-like features for ASR,'' Proceedings of the Seventh International Conference on Spoken Language Processing, Denver, CO, September 2002.

[2]In this paper the term *formant* or *true formants* refers to the resonance frequencies of the vocal tract. The term *formant-like* refers to features that are similar, but not necessarily identical, to true formants.

[3]Vowel steady state was defined by Peterson and Barney as, ''... following the influence of the /h/ and preceding the influence of the /d/, during which a practically steady state is reached'' (Peterson and Barney, 1952).

[4]In Hillenbrand and Gayvert (1993) it was found that, for a vowel classification task, nonlinear frequency transforms significantly enhanced the performance of a linear discriminant classifier. For a quadratic classifier, on the other hand, there was no advantage for any of the nonlinear transforms (mel, log, Koenig, Bark) over linear frequency. During the current investigation HMM classification experiments were also conducted using the original, linear frequency values. No significant difference was observed between the tests performed with the linear frequency values and the mel-scaled values.

[5]The possibility to apply pre-emphasis is incorporated in the acoustic preprocessing of the RF algorithm. One may therefore assume that the inherent spectral tilt in the data is equalized and that all the LPC poles are available to model spectral peaks.

[6]These features were derived using HTK's feature extraction software (Young *et al.*, 1997).

[7]Data from /e/ and /o/ were omitted in Hillenbrand *et al.* (1995) to facilitate comparisons with Peterson and Barney's results.

Bazzi, I., Acero, A., and Deng, L. (**2003**). ''An expectation maximization approach for formant tracking using a parameter-free non-linear predictor,'' in Proceedings of ICASSP 2003, Hong Kong, pp. I.464–I.467.

Davis, S. B., and Mermelstein, P. (**1980**). ''Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**, 357–366.

Delsarte, P., and Genin, Y. V. (**1986**). ''The Split Levinson Algorithm,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 470–478.

de Wet, F., Cranen, B., de Veth, J., and Boves, L. (**2000**). ''Comparing acoustic features for robust ASR in fixed and cellular network applications,'' in Proceedings of ICASSP 2000, Istanbul, Turkey, pp. 1415–1418.

Duda, R. O., Hart, P. E., and Stork, D. G. (**2001**). *Pattern Classification*, 2nd ed. (Wiley, New York).

Flanagan, J. L. (**1972**). *Speech Analysis Synthesis and Perception*, 2nd ed. (Springer, Berlin).

Garner, P., and Holmes, W. (**1998**). ''On the robust incorporation of formant features into hidden Markov models for automatic speech recognition,'' in Proceedings of ICASSP 1998, Seattle, WA, pp. 1–4.

Hillenbrand, J. M., and Gayvert, R. T. (**1993**). ''Vowel classification based on fundamental frequency and formant frequencies,'' J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.

Holmes, J., Holmes, W., and Garner, P. (**1997**). ''Using formant frequencies in speech recognition,'' in Proceedings of Eurospeech 1997, Rhodes, Greece, pp. 2083–2086.

Hunt, M. J. (**1999**). ''Spectral signal processing for ASR,'' in Proceedings of ASRU 1999, Keystone, CO.

Juang, B.-H., Chou, W., and Lee, C.-H. (**1996**). ''Statistical and discriminative methods for speech recognition,'' in *Automatic Speech and Speaker Recognition, Advanced Topics*, edited by C.-H. Lee, F. Soong, and K. Paliwal (Kluwer Academic, Boston).

Ladefoged, P. (**1975**). *A Course in Phonetics* (Harcourt Brace Jovanovich, New York).

Markel, J., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer, Berlin).

Minifie, F. D., Hixon, T. J., and Williams, F., editors (**1973**). *Normal Aspects of Speech, Hearing and Language* (Prentice–Hall, Englewood Cliffs, NJ).

Nadeu, C. (**1999**). ''On the filter-bank-based parametrization front-end for robust HMM speech recognition,'' in Proceedings of Nokia Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, Finland, pp. 235–238.

Noisex (**1990**). NOISE-ROM-0. NATO: AC243/(Panel 3)/RSG-10, ESPRIT: Project 2589-SAM.

Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.

Pols, L. C. W., van der Kamp, L. J. T., and Plomp, R. (**1969**). ''Perceptual and physical space of vowel sounds,'' J. Acoust. Soc. Am. **46**, 458–467.

Rabiner, L. R., and Juang, B. H. (**1993**). *Fundamentals of Speech Recognition* (Prentice–Hall, Englewood Cliffs, NJ).

Rabiner, L. R., and Schafer, R. W. (**1978**). *Digital Processing of Speech Signals* (Prentice–Hall, Englewood Cliffs, NJ).

Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).

Weber, K. (**2003**). ''HMM mixtures (HMM2) for robust speech recognition,'' PhD thesis, Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland.
Weber, K., Bengio, S., and Bourlard, H. (**2000**). ''HMM2—A novel approach to HMM emission probability estimation,'' in Proceedings of ICSLP 2000, Beijing, China, pp. (III)147–150.
Weber, K., Bengio, S., and Bourlard, H. (**2001a**). ''HMM2—extraction of formant structures and their use for robust ASR,'' in Proceedings of Eurospeech 2001, Aalborg, Denmark, pp. 607–610.
Weber, K., Bengio, S., and Bourlard, H. (**2001b**). ''A pragmatic view of the application of HMM2 for ASR,'' IDIAP-RR 23, IDIAP, Martigny, Switzerland.
Weber, K., Bengio, S., and Bourlard, H. (**2001c**). ''Speech recognition using advanced HMM2 features,'' in Proceedings of ASRU 2001, Madonna di Campiglio, Trento, Italy.
Weber, K., de Wet, F., Cranen, B., Boves, L., Bengio, S., and Bourlard, H. (**2002**). ''Evaluation of formant-like features for ASR,'' in Proceedings of ICSLP 2002, Denver, CO.
Welling, L., and Ney, H. (**1996**). ''A model for efficient formant estimation,'' in Proceedings of ICASSP 1996, Atlanta, GA, pp. 797–800.
Willems, L. F. (**1986**). ''Robust formant analysis,'' in IPO Annual Report 21, Eindhoven, The Netherlands, pp. 34–40.
Young, S., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**1997**). *The HTK Book (for HTK Version 2.1)* (Cambridge University, Cambridge).

A nonmusical paradigm for identifying absolute pitch possessors

David A. Ross[a)]
Department of Diagnostic Radiology, Yale School of Medicine, Box 208043, New Haven, Connecticut 06520

Ingrid R. Olson
Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104-6196

Lawrence E. Marks
John B. Pierce Laboratory and Departments of Epidemiology/Public Health and Psychology, Yale School of Medicine, 290 Congress Avenue, New Haven, Connecticut 06519

John C. Gore
Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, 1161 21st Avenue South, R-1032 MCN, Nashville, Tennessee 37232

(Received 17 November 2003; revised 12 April 2004; accepted 12 April 2004)

The ability to identify and reproduce sounds of specific frequencies is remarkable and uncommon. The etiology and defining characteristics of this skill, absolute pitch (AP), have been very controversial. One theory suggests that AP requires a specific type of early musical training and that the ability to encode and remember tones depends on these learned musical associations. An alternate theory argues that AP may be strongly dependent on hereditary factors and relatively independent of musical experience. To date, it has been difficult to test these hypotheses because all previous paradigms for identifying AP have required subjects to employ knowledge of musical nomenclature. As such, these tests are insensitive to the possibility of discovering AP in either nonmusicians or musicians of non-Western training. Based on previous literature in pitch memory, a paradigm is presented that is intended to distinguish between AP possessors and nonpossessors independent of the subjects' musical experience. The efficacy of this method is then tested with 20 classically defined AP possessors and 22 nonpossessors. Data from these groups strongly support the validity of the paradigm. The use of a nonmusical paradigm to identify AP may facilitate research into many aspects of this phenomenon. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1758973]

PACS numbers: 43.75.Cd, 43.66.Hg [SEM] Pages: 1793–1799

I. INTRODUCTION

Most individuals perceive melodic sequences without being able to identify absolutely the notes involved. For example, most people could recognize Beethoven's Fifth Symphony but would not know that the first note is a ''G.'' This example illustrates that the perception of tonal sequences is generally dependent on the ability to recognize the relationships between notes, rather than on recognizing the exact frequency of each note. When further refined, this skill—the ability to identify or produce musical intervals accurately—is referred to as ''relative pitch'' (RP).

A small subset of the population is also capable of quickly and accurately labeling tonal stimuli on the basis of their fundamental frequencies (without the use of a reference tone). This skill is generally referred to as ''absolute pitch'' (AP), or, colloquially, as ''perfect pitch.'' This phenomenon has been of interest to musicians and scientists alike for many years, in large part because of the musical advantages it may endow. For example, it may be easier for AP possessors to transcribe or remember a piece of music because they can directly identify each of the notes that are played. Similarly, it may be easier for them to sight-sing difficult passages by producing each note individually rather than trying to tune challenging intervals. AP possessors may also be limited in several important ways. Strict adherence to the absolute frequencies of tones as found in the equal-tempered scale could hinder a possessor's ability to remain in tune with a choir that drifts flat or to maintain proper intonation on intervals that differ from equal temperament. There are also reports of AP musicians who find that their sense of pitch shifts as they grow older, possibly due to changes in the elasticity of the basilar membrane.[1,2]

A central point of controversy in the study of AP has been the relative roles of heredity and early musical experience. While a parsimonious approach might allow that both genetic and epigenetic factors are involved, few authors in the field have adopted such a moderate stance. Early models of AP suggested that ''true'' AP is a genetically determined trait[3,4] and several recent papers have provided convincing evidence that AP has a genetic component.[2,5–7] Other models of absolute pitch have argued that AP results from early learning experiences.[8–11] These models generally propose that early in life some individuals memorize the exact frequency of each note in the musical scale, either consciously or unconsciously, resulting in absolute pitch abilities. Indirect evidence in support of this theory is drawn from the

a)Electronic mail: david.a.ross@yale.edu

finding that all reported AP individuals possess extensive early musical training,[4,11] and that the degree of AP tends to correlate with years of musical training.[9] However, the data supporting this theory are problematic since AP testing methods have been biased towards identifying individuals who possess musical training. Specifically, most paradigms have employed a labeling technique in which subjects are played a series of musical tones and are asked to name the notes according to the Western musical scale. Performance on such a test is predicated on a basic level of musical expertise and is therefore insensitive to the possibility of discovering AP in either a nonmusician or a musician of non-Western training. As such, there is no convincing evidence that only musicians possess AP.[12]

The goal of the present study was to develop a paradigm that could distinguish between AP possessors and nonpossessors without requiring subjects to have knowledge of (or at least without requiring subjects to use) conventional musical nomenclature. The paradigm that we developed for this purpose resembles one used by Siegel,[13] in which she asked AP possessors and nonpossessors to make same/different judgments about a target stimulus following a series of interfering tones. We refined Siegel's design so as to present subjects a target stimulus followed either by a silent interstimulus interval (ISI) or one filled with interfering tones, after which the subjects were asked to reproduce the original tone by adjusting the knob of a sine function generator. We expected that both groups would perform this task with relative accuracy after a silent ISI. In contrast, a significant body of literature indicates that passive exposure to as few as four to seven interfering tones should be sufficient to destroy completely a non-AP subject's memory for the target[14–17] whereas AP subjects should remain accurate.[13,18,19] Thus, we expected that this paradigm would be able to distinguish accurately between AP and NAP individuals.

Importantly, this paradigm does not require subjects to make use of any musical knowledge or training: subjects are never asked to name musical notes nor are they asked to produce musical notes by name. As discussed by Siegel, we do not deny the possibility that a subject *could* apply categorical labels to the target stimuli in order to maintain representations of them. However, it is important to emphasize that this task focuses strictly on the ability of subjects to encode durable long-term representations of target stimuli. The extent to which this skill overlaps with the ability to name notes (and thus corresponds to classically defined ''perfect pitch'') will be a point of great interest.

II. EXPERIMENT 1

A. Method

1. Participants

At this time, because we were strictly interested in testing the validity of our paradigm, only experienced musicians were included (i.e., individuals who could easily be classified as AP possessors or nonpossessors using standard methods). Nine musicians participated in the original version of this experiment (3 AP, 6 NAP) and 33 subjects (17 AP, 16 NAP) took part in an abbreviated version. The 20 AP possessors (10 male) had an average age of 23.1 ± 7.6 and an average of 17.2 ± 8.0 years of performance on their primary instrument. The 22 NAP subjects had an average age of 21.5 ± 5.5 and an average of 11.0 ± 6.5 years experience playing their primary instrument.

2. Screening for note naming ability

The following test was used to verify the claims of subjects who reported to possess AP. The method is similar to those used by Baharloo *et al.*[5] and Gregersen *et al.* (personal communication). We used a predefined criterion of 80% accuracy to qualify a subject as an AP possessor.

The test required subjects to orally name 60 musical notes. The tones were presented in six blocks of ten notes each. Between blocks, subjects were allowed to rest for as long as desired. Notes were pseudo-randomly selected with replacement from within the range of C2–C7 (65.4–2093.0 Hz) such that no consecutive notes were within nine whole steps of each other and no note values repeated within any five-note span. All stimuli had a duration of 500 ms with an interstimulus interval of 2100 ms. Subjects were asked to label orally the musical note name (without regard to octave) of each stimulus (e.g., ''F'' or ''B^b'') during the interstimulus interval. Responses were manually recorded by the experimenter and equivalent enharmonic labels were accepted as correct (e.g., ''G^b'' vs. ''$F^\#$''). The test was administered using PsyScope 1.2.1 with a script that was identical for all subjects. To minimize the success of skilled RP musicians, subjects were not given a reference tone and at no time during the experiment was any sort of feedback provided.

The first 30 stimuli consisted of sinusoidal tones. The notes were generated using the ''tone generator'' function of SoundEditII (SEII) for the Macintosh. Frequency values were based on the equal-tempered scale with $A_4 = 440$ Hz. The next 30 stimuli consisted of piano tones. The tones were generated using a Roland JV-50 synthesizer. In order to control for note duration and envelope the tones were recorded directly into SEII from a Finale '98 file. Fundamental frequency values were again based on an equal-tempered scale with $A_4 = 440$ Hz.

Data were scored in terms of percent accuracy. Accuracy was measured for each subject as the proportion of the total stimuli that received correct responses (i.e., omitted notes were considered incorrect).

Twenty subjects qualified as AP possessors, performing with an average of $91\% \pm 7\%$ accuracy (range: 80%–100%; chance = 8%). Eighteen subjects claiming not to possess AP performed at $9\% \pm 6\%$ (range: 2%–20%).[20]

3. Stimuli and procedure

Subjects were played a tone followed by a silent interval of varying length (either 2, 8, or 32 s). At the end of the ISI a nontonal auditory click (similar to the striking of wooden block; duration ~100 ms) was sounded, indicating that participants should reproduce the original target frequency by adjusting the knob of a digital sine-wave function generator. Subjects were explicitly instructed that they should try to match both the chroma and octave but that chroma was more

important (i.e., that it was better to reproduce the chroma accurately in the wrong octave then to get the octave correct but lose the exact chroma). Before each stimulus, the function generator was reset to a frequency of 1000 Hz. Subjects indicated successful completion of the task either orally or by tapping their finger on the top of the function generator. The experimenter recorded the participants' responses manually and was not aware of the correct frequency values.

Ten target tones were played for each condition in a fixed order (2s, 8s, 32s, 2s,...). The target stimuli consisted of pure tones generated using the frequency generator function of SoundEditII (SEII) for the Macintosh. Targets were selected from a set of 120 possible tones. Sixty tones were notes from the Western, equal-tempered scale ($A_4 = 440$), ranging from C2 to B6 (65.4 to 1975.5 Hz). A second set of 60 tones was then generated by deliberately mistuning the original notes upwards by a variable amount (ranging from 20 to 80 cents). The unfamiliar tones had a frequency range from 66.4 to 2060.6 Hz. Tones were presented binaurally for a duration of 300 ms. Within each experiment, targets were pseudo-randomly selected, with replacement, so as to include an equal number of familiar and unfamiliar tones from each octave. Consecutive target tones were never within a half step (or octave $\pm$ 1 half step) of each other.

An identical computer script was used for each participant. To prevent visual monitoring of the function generator, participants were blindfolded during the performance of all tasks. Individuals were not given feedback on their performance until the completion of all tasks and at no time was a reference tone provided.

In the abbreviated version of the experiment, subjects were presented eight tones each for 2-, 8-, and 16-s ISI's and targets were drawn from between C3 and B6. All other parameters were the same.

4. Data processing

Because we were primarily interested in the encoding of chroma and sought to avoid the confounding influence of octave errors, subjects were instructed that reproducing the chroma exactly should be their priority. Accordingly, each response was corrected to the nearest octave. This correction is consistent both with the majority of AP studies[2,13,19,21–23] and with physiological evidence supporting the concept of octave equivalence.[24–27] This manipulation is unlikely to confound group differences given the substantial body of literature that fails to find differences between AP possessors and nonpossessors in the frequency of octave errors (cf. Ref. 9). After octave correction, the response distance from the target stimulus was calculated in half steps (hs) away (where a half step is defined as $\frac{1}{12}$ of the distance of an octave).[28] Half steps were chosen as a convenient and familiar unit of measure that respects the fact that our perceptual response to musical sound frequency is logarithmic. It is important to note, however, that this measure is independent of any specific musical values—it is used solely as a means of quantifying distance.

Individual performance was judged using constant error (CE) as a measure of bias and standard deviation (σ) as a measure of variability. Accurate performance would be expected to yield a normal distribution of responses that is centered at zero (low CE and low σ). In contrast, random performance (i.e., if responses bore no relationship to the original target) would be expected to produce a uniform distribution from -6 to $+6$ hs. Thus, it would have a low CE but high variability.

Given the expected distributions for accurate and inaccurate performance (and the potential for deviation from normality), the median of the positive distribution (MPD) was used to determine whether observed distributions differed from the prediction of the null hypothesis. Accurate performance should have a MPD close to 0 whereas chance performance would have a MPD near 3. Group performance was compared to the null hypothesis by calculating a simple z score from the average and standard error of MPD values (where $z = |3 - \mu|/\mathrm{SE}$). Group differences for CE, σ, and MPD were calculated using standard ANOVAs.

B. Results and discussion

Histograms of combined AP and NAP performance for each time interval are shown in Fig. 1. AP subjects performed at $(\mathrm{CE} \pm \sigma) - 0.14 \pm 0.47$, 0.05 ± 0.44, and -0.02 ± 0.47 hs and NAP subjects performed at -0.05 ± 2.14, -0.23 ± 2.53, and -0.13 ± 2.10 hs for time intervals of 2, 8, and 16/32 seconds.[29] ANOVA with CE as the dependent variable[30] (G-G $\varepsilon = 0.89$) showed no significant effect of group [$F(1,40) = 0.35$; $p > 0.50$] or time interval [$F(2,80) = 0.08$; $p > 0.90$]. There was no interaction between group and time interval [$F(2,80) = 2.39$; $p > 0.10$]. ANOVA with σ as the dependent variable (G-G $\varepsilon = 0.91$) showed a significant effect of group [$F(1,40) = 68.26$; $p = 0.0001$] but not of time interval [$F(2,80) = 2.41$; $p = 0.10$]. The interaction between group and time interval was not significant [$F(2,80) = 3.13$; $p = 0.05$].

The AP subjects performed at $(\mathrm{CE} \pm \sigma) - 0.04 \pm 0.45$ and -0.03 ± 0.48 for in-tune and mistuned stimuli, respectively. NAP subjects performed at -0.29 ± 2.10 and 0.08 ± 2.31. None of these differences was significant.

The average MPDs of the AP subjects were (avg. $\pm$ se): 0.33 ± 0.03, 0.29 ± 0.03, and 0.33 ± 0.04 for the three time intervals, and for the NAP subjects 1.48 ± 0.25, 1.55 ± 0.23, and 1.26 ± 0.21. All of these values differ significantly from the prediction of the null hypothesis. An ANOVA with MPD as the dependent variable (G-G $\varepsilon = 1.00$) showed a significant effect of group [$F(1,40) = 27.84$; $p = 0.0001$] but not of time [$F(2,80) = 0.82$; $p > 0.40$]. The interaction between group and time was not significant [$F(2,80) = 1.19$; $p > 0.30$].

These data show that the AP group was more consistent than the NAP group in their ability to reproduce the target tones. This difference may be accounted for by the experimental setup—specifically, because the frequency generator was always turned on at 1000 Hz and the output function was continuous, there was a small amount of tonal interference intrinsic to the task. NAP subjects frequently reported this to be distracting. In both groups, it is likely that this interference led to some degree of a gap between subjects' ability to *remember* tones and their ability to *reproduce* the tones accurately using the experimental setup. Nevertheless,

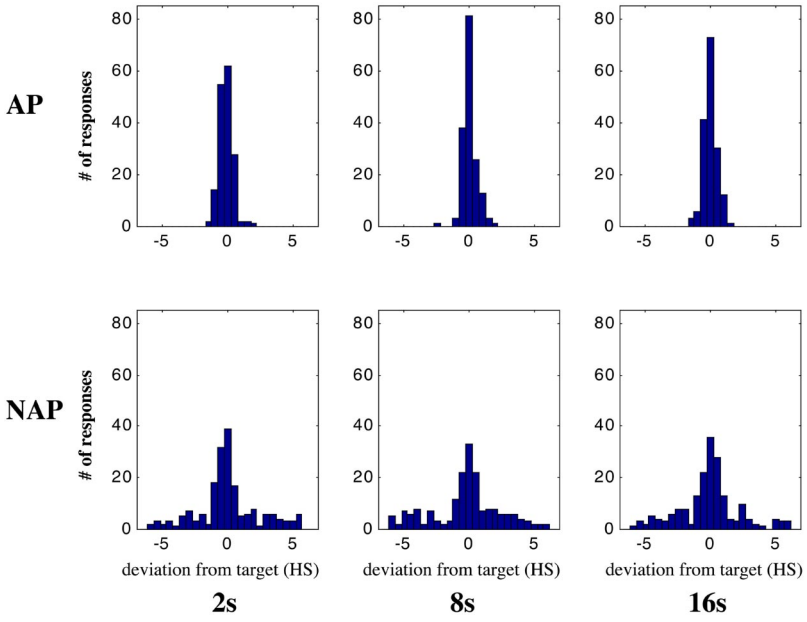


FIG. 1. Experiment 1: Histogram of responses by the AP and NAP groups for silent delays of 2, 8, and 16/32 s. Accurate performance would appear as a normal distribution that is centered at zero with little variability. Random performance would be a uniform distribution from −6 to +6.

both groups demonstrated that they *were* able to reproduce the target stimulus accurately with our experimental apparatus. Further, data show that neither group's performance diminished significantly *below their own baseline* with time intervals of up to 32 s. These data are consistent with previous literature in pitch memory.[17,31]

III. EXPERIMENT 2

In this experiment we tested whether subjects could reproduce a stimulus frequency accurately after passive exposure to interfering tones. Four conditions were used, each with an increasing number of distracting tones. Given that neither AP nor NAP performance was significantly affected by an increased time delay, any decrease in performance on the present task must be attributed to the effect of the distracting tones on the maintained memory trace.

A. Method

Experiment 2 differed from experiment 1 by filling the ISI with a variable number of distracting tones. Four conditions were created: a 2-s ISI with 1 distracting tone; a 4-s ISI with 11 distracting tones; an 8-s ISI with 31 distracting tones; and a 16-s ISI with 71 distracting tones.[32] The distracting tones were randomly selected from the 120 stimulus tones and presented for 200 ms each. There was a 200-ms interval between the target tone and the first distractor and a 1600-ms interval between the last distractor and the cue for subjects to reproduce the target stimulus. Again, ten target tones were played for each condition in a fixed order (2 s, 4 s, 8 s, 16 s, 2 s,...). In the abbreviated version, eight tones each were presented in 2-, 8-, and 16-s conditions. All other parameters were the same as in experiment 1.

B. Results and discussion

Histograms of AP and NAP performance for each time interval are shown in Fig. 2. AP subjects performed at -0.04 ± 0.60, -0.35 ± 0.97 (three subjects only), 0.00 ± 0.58, 0.05 ± 0.76 hs and NAP subjects performed at -0.05 ± 2.43, 0.27 ± 2.29 (six subjects only), 0.25 ± 3.19, 0.22 ± 3.18 hs for the four time intervals. ANOVA with CE as the dependent variable (G-G $\varepsilon=0.96$) showed no effect of group [$F(1,40)=0.88$; $p>0.30$] or time [$F(2,80)=0.47$; $p>0.60$] and there was no interaction between group and time [$F(2,80)=0.34$; $p>0.70$]. ANOVA with σ as the dependent variable (G-G $\varepsilon=0.89$) showed significant effects of group [$F(1,40)=385.66$; $p<0.0001$] and time [$F(2,80)=6.16$; $p<0.01$] as well as a significant interaction between group and time [$F(2,80)=4.53$; $p<0.025$].

The AP subjects performed at (CE$\pm\sigma$) -0.03 ± 0.73 and -0.04 ± 0.68 for in-tune and mis-tuned stimuli, respectively. NAP subjects performed at -0.10 ± 2.98 and 0.34 ± 2.83. None of these differences was significant. These data illustrate that stimulus intonation did not have any consistent effect on the performance of either AP or NAP subjects.

The average MPDs of the AP subjects were 0.31 ± 0.03, 0.48 ± 0.14, 0.39 ± 0.04, and 0.44 ± 0.06 for the four time intervals; for the NAP subjects they were $1.63\pm.29$, 1.30 ± 0.17, 2.73 ± 0.17, and 2.86 ± 0.19. The AP values differed from the null hypothesis at all four time intervals. The NAP group differed from the null hypothesis for the first two time intervals but not the third ($p>0.05$) or fourth ($p>0.20$). An ANOVA with MPD as the dependent variable (G-G $\varepsilon=0.77$) showed significant effects of group [$F(1,40)=207.44$; $p=0.0001$] and time [$F(2,80)=9.63$; $p<0.001$]. The interaction between group and time was also significant [$F(2,80)=6.54$; $p<0.005$].

Because there was a trend towards increased musical experience in the AP group, additional analyses were conducted to evaluate the extent to which total years of musical experience and the age at which a subject first began musical training affected performance on our paradigm (see Fig. 3). Neither of these variables was well correlated with subjects' average performance for the tasks with interfering tones (for years experience, $r^2=0.16$; for age of onset, $r^2=0.36$). Similarly, there was no significant correlation of either of

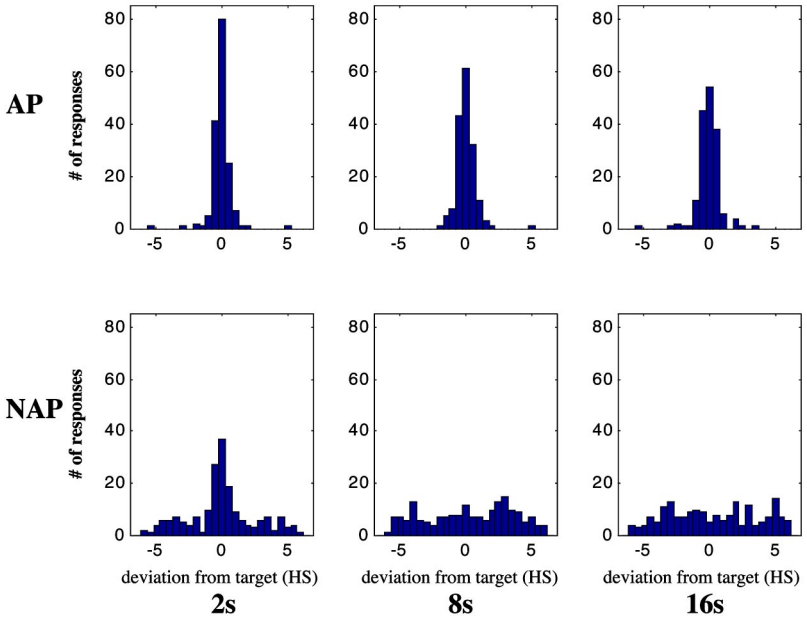


FIG. 2. Experiment 2: Histogram illustrating the distance away from target of responses by the AP and NAP groups with 1, 31, and 71 interfering tones.

these variables for within-group analyses (for AP subjects, $r^2 = 0.01$ and 0.16 for years experience and age of onset, respectively; for NAP, $r^2 = 0.02$ and 0.03). It should also be noted, in stark contrast to the idea of a "critical period" during which training must take place to develop AP (e.g., Refs. 11, 33, and 34), that two of our best AP possessors (both of whom scored perfect on the note-naming test and near perfect on the reproduction experiments) began training at ages 10 and 12, respectively.

While these data illustrate clear effects of group, they do not address whether this paradigm would be suitable to distinguishing between AP and NAP possessors at an individual level. To this end, data were separately analyzed for each subject as the average of the two conditions with the highest amount of interference (i.e., the 8- and 16-s intervals in experiment 2 with 31 and 71 interfering tones). The AP subjects had an average σ of 0.66 with a range of 0.27–1.43 and a median of 0.59. The NAP subjects had an average of 3.18 with a range of 2.31–3.82 and a median of 3.19. Thus, there was a gap of almost one full standard deviation between the worst performance by an AP subject and the best performance by an NAP subject; the group medians suggests an even wider gap between prototypical AP and NAP possessors.

These data confirm at the individual level the major effect of the group analysis: AP and NAP subjects were differentially affected by the presence of interfering tones. With a large number of distracting tones, NAP performance was indistinguishable from chance. Only the AP subjects were able to remember and reproduce stimuli accurately after the initial sensory trace was destroyed. These data are consistent with previous research[13–16,18] and support the notion that this

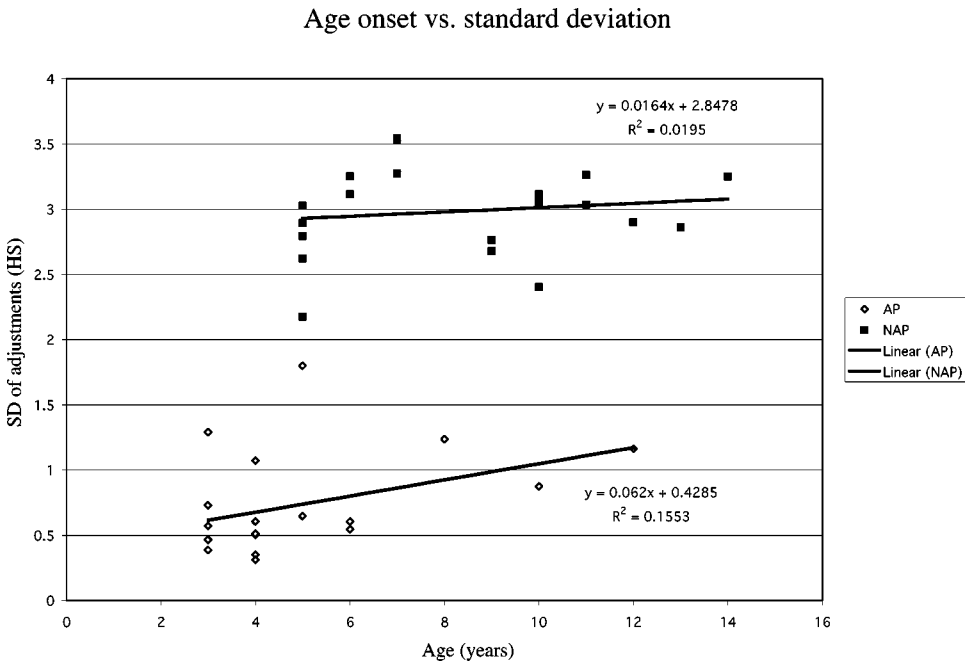


FIG. 3. Correlation between age of onset of musical training and standard deviation of performance for experiment 2. AP subjects are represented by squares and NAP subjects by triangles. Solid lines represent the linear regression indicated by the equations.

paradigm may be an effective way to identify individuals as either possessors or nonpossessors of AP.

IV. DISCUSSION

In this paper, we present a paradigm that is able to distinguish accurately between AP possessors and nonpossessors without requiring subjects to make use of any knowledge of musical nomenclature. Many previous data support the theoretical basis of this paradigm. Principally, Bachem[18,35] argued that the fundamental difference between AP possessors and nonpossessors is that NAP individuals are able to encode only the height of tonal stimuli whereas AP individuals encode representations of both height and chroma. Further, Bachem was careful to point out that this skill could not be acquired—rather, he characterized pseudo-AP (that which could be learned later in life) as the ability to become particularly adept at encoding height. However, this skill allows individuals to encode frequency at best to a precision of 5–9 semitones. Given that our paradigm corrected all responses to the proper octave, this level of precision would be ineffective. Instead, the accuracy of responses should reflect subjects' precision at encoding and then reproducing the chroma of the targets. Over the years, a significant body of data has accrued that corroborate Bachem's original assertion: AP and NAP individuals are comparable at performing tasks that involve echoic or short-term memory, but only AP possessors are able to encode and maintain an accurate long-term representation of stimulus chroma.[13–16,18,19,31]

Additional theoretical support for our paradigm comes from Takeuchi and Hulse (Ref. 33, p. 353) who discuss the many factors (e.g., timbre, chroma, and height) that may affect individuals' absolute pitch abilities, and suggest that susceptibility to these factors may be a hallmark of ''lower'' forms of AP (this is comparable to Bachem's description of ''limited universal'' absolute pitch[35]). In contrast, Takeuchi and Hulse suggest that the ability to reproduce accurately the absolute frequency of stimuli may be a hallmark of ''higher'' forms of AP. Obviously, as originally intended by the authors, this would not be a suitable means to test for AP in nonmusicians because it requires subjects to have memorized specific frequency values. However, in conjunction with the findings of Bachem and others, it is reasonable to believe that if subjects were given a target to remember, followed by an imposing interval that destroyed their short-term trace of the stimulus, this may be an excellent way of testing for AP. Reasonable options for such an interval could be either a very long delay (on the order of minutes, hours, or days), although this would be experimentally prohibitive for any large-scale study, or, alternatively, some sort of tonal interference (which is known to destroy rapidly the ability of normal individuals to maintain a tonal representation).

It should be emphasized that this skill is *not* equivalent to the ''latent AP'' described by Levitin and others.[36–38] Paradigms used by those groups test the ability to evoke the memory of a specific, spectrally complex stimulus that has accumulated across many repeated presentations. In contrast, the present paradigm explicitly isolates the ability to encode an immediate representation of stimulus fundamental frequency without the presence of any extrinsic cues. Elsewhere, we expand on the significance of this difference and present a new model of AP that identifies subpopulations within the traditional group of note namers and describes the differential effects of timbre, chroma, and intonation on subjects' ability to encode accurate representations of stimulus frequency. We believe this model may help reconcile conflicting views on the etiology of AP.[39]

There are several important applications of the new experimental paradigm presented here. Foremost, use of this paradigm may be able to provide key information in resolving the dispute on the etiology of AP. Advocates of the early-learning theory have argued that musical training is required in order to develop AP. Thus, the only way to rigorously test this theory would be to test for AP in individuals both with *and without* musical training. However, as long as the ability to name notes is considered the hallmark of AP, such a test will be impossible—only individuals with musical experience will be able to pass the test (cf. Ref. 12). Perhaps a better definition of AP—based on Bachem's original description—would be as the ability to encode (and reproduce) durable representations of the chroma of periodic stimuli. To prevent confusion with traditional definitions of AP, we might refer to this skill as the ability to perceptually encode, or APE.

Obviously, this revised definition would include all individuals capable of passing conventional AP tests. It would also include those individuals who are commonly recognized as AP possessors but who cannot name notes accurately according to the Western equal-tempered system based on $A_4 = 440$ Hz—such as musicians of non-Western training or AP possessors whose sense of pitch has shifted with age[1,40] or medication.[41,42] Elsewhere, we have presented data that demonstrate a dissociation between the ability to name notes and the ability to encode meaningful representations of stimulus frequency.[39,43] Importantly, being more inclusive, under our definition it should also be possible to test whether musical training is required to develop APE. Indeed, the paradigm presented here may be ideally suited for this purpose.

V. CONCLUSION

All previous tests of AP have required subjects to possess some degree of musical expertise—as such, it has never been possible to test the hypothesis that early musical training is required to develop AP. We present a new paradigm that does not require subjects to make use of any musical knowledge or experience. Data from over 40 classically defined AP and NAP musicians strongly support the validity of this method.

Because it does not require labeling of notes, this new paradigm may be usefully applied in several ways—in quantifying more accurately than heretofore the prevalence of AP, in genetically assessing the possible hereditary contribution to AP, and, perhaps most centrally, in determining the extent to which musical experience is, or is not, needed to develop AP. To the best of our knowledge, this is the first test explicitly designed as a possible way of detecting AP in nonmusicians. Given the performance of experienced musicians with

and without AP, it would be of great interest to discover a nonmusician who is able to perform accurately on this test.

ACKNOWLEDGMENTS

The authors would like to thank Linda Bartoshuk, Robert Schultz, Eric Odgaard, David Katz, Junhyong Kim, and Pawel Skudlarski for their expert consultation on this project and Thomas C. Duffy and Thomas Cantey for invaluable musical consultation. This research was supported in part by MSTP Training Grant No. GM07205 from the NIH to the Yale School of Medicine and NIH Grant No. R01 DC00271-19 from the National Institute of Deafness and Other Communication Disorders to the John B. Pierce Laboratory. Equipment was donated by Bill Rabinowitz and the Bose Co.

[1] P. E. Vernon, "Absolute pitch: a case study," Br. J. Psychol. **68**, 485–489 (1977).
[2] S. Baharloo, P. A. Johnston, S. K. Service, J. Gitschier, and N. B. Freimer, "Absolute pitch: an approach for identification of genetic and nongenetic components [see comments]," Am. J. Hum. Genet. **62**, 224–231 (1998).
[3] C. D. Seashore, *Psychology of Music* (McGraw-Hill, New York, 1938).
[4] A. Bachem, "Absolute pitch," J. Acoust. Soc. Am. **27**, 1180–1185 (1955).
[5] S. Baharloo, S. K. Service, N. Risch, J. Gitschier, and N. B. Freimer, "Familial aggregation of absolute pitch," Am. J. Hum. Genet. **67**, 755–758 (2000).
[6] P. K. Gregersen, "Instant recognition: the genetics of pitch perception [editorial; comment] [see comments]," Am. J. Hum. Genet. **62**, 221–223 (1998).
[7] P. K. Gregersen, E. Kowalsky, N. Kohn, and E. W. Marvin, "Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment," Am. J. Med. Genet. **98**, 280–282 (2001).
[8] H. J. Watt, *The Psychology of Sound* (Cambridge U.P., Cambridge, Press 1917).
[9] A. H. Takeuchi and S. H. Hulse, "Absolute pitch," Psychol. Bull. **113**, 345–361 (1993).
[10] L. L. Cuddy, "Training the absolute identification of pitch," Percept. Psychophys. **8**, 265–269 (1970).
[11] K. Miyazaki, "Musical pitch identification by absolute pitch possessors," Percept. Psychophys. **44**, 501–512 (1988).
[12] P. K. Gregersen, E. Kowalsky, N. Kohn, and E. W. Marvin, "Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component [letter; comment]," Am. J. Hum. Genet. **65**, 911–913 (1999).
[13] J. A. Siegel, "Sensory and verbal coding strategies in subjects with absolute pitch," J. Exp. Psychol. **103**, 37–44 (1974).
[14] T. Pechmann and G. Mohr, "Interference in memory for tonal pitch: implications for a working-memory model," Mem. Cognit. **20**, 314–320 (1992).
[15] D. Deutsch, "Tones and numbers: specificity of interference in immediate memory," Science **168**, 1604–1605 (1970).
[16] D. Deutsch, "Mapping of interactions in the pitch memory store," Science **175**, 1020–1022 (1972).
[17] C. Semal and L. Demany, "Dissociation of pitch from timbre in auditory short-term memory," J. Acoust. Soc. Am. **89**, 2404–2407 (1991).
[18] A. Bachem, "Time factors and absolute pitch determination," J. Acoust. Soc. Am. **26**, 751–753 (1954).
[19] R. J. Zatorre and C. Beckett, "Multiple coding strategies in the retention of musical tones by possessors of absolute pitch," Mem. Cognit. **17**, 582–589 (1989).
[20] N. B. Four NAP subjects did not take this test; however, all four were accomplished musicians of comparable age and experience to the rest of the cohort and were separately tested to confirm that they did not have AP.
[21] M. Klein, M. Coles, and E. Donchin, "People with absolute pitch process tones without producing a P300," Science **223**, 1306–1309 (1984).
[22] E. C. Hantz, K. G. Kreilick, E. W. Marvin, and R. M. Chapman, "Absolute pitch and sex effect event-related potential activity for a melodic interval discrimination task," J. Acoust. Soc. Am. **102**, 451–460 (1997).
[23] G. C. Crummer, J. P. Walton, J. W. Wayman, E. C. Hantz, and R. D. Frisina, "Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch," J. Acoust. Soc. Am. **95**, 2720–2727 (1994).
[24] K. Ohgushi, "The origin of tonality and a possible explanation of the octave enlargement phenomenon," J. Acoust. Soc. Am. **73**, 1694–1700 (1983).
[25] M. F. McKinney and B. Delgutte, "A possible neurophysiological basis of the octave enlargement effect," J. Acoust. Soc. Am. **106**, 2679–2692 (1999).
[26] P. Cariani, "Temporal codes, timing nets, and music perception," New Music Percept. **30**, 107–135 (2001).
[27] J. D. Warren, S. Uppenkamp, R. D. Patterson, and T. D. Griffiths, "Analyzing pitch chroma and pitch height in the human brain," Ann. N.Y. Acad. Sci. **999**, 212–214 (2003).
[28] $d = (12/\log 2)\log(f_{response}/f_{stimulus})$.
[29] Data were tested for both AP and NAP groups to determine if there was a difference in performance for the 16- and 32-s conditions. ANOVA of CE showed no effect of time interval for either the AP [$F(1,18) = 1.22$; $p > 0.20$] or NAP group [$F(1,20) = 0.82$; $p > 0.30$]. Similarly, ANOVA with σ as a factor showed no effect of time interval for either the AP [$F(1,18) = 0.60$; $p = 0.45$] or NAP group [$F(1,20) = 0.04$; $p > 0.80$].
[30] For this and all other tests (unless otherwise specified) we used a repeated-measures ANOVA, with AP/NAP status as a between-group factor and time interval as a within-group factor. To account for possible violations of sphericity, Greenhouse–Geisser (G-G) corrections were calculated and the corrected p-values are reported in the main text.
[31] T. A. Keller, N. Cowan, and J. S. Saults, "Can auditory memory for tone pitch be rehearsed?" J. Exp. Psychol. Learn. Mem. Cogn. **21**, 635–645 (1995).
[32] These parameters were chosen to allow for a broad assessment of the effect of interference. It was not our intent to systematically study the interaction between ISI duration and the number of interfering tones at this time.
[33] A. H. Takeuchi and S. H. Hulse, "Absolute-pitch judgments of black-and white-key pitches," Music Percept. **9**, 27–46 (1991).
[34] D. Sergeant, "Experimental investigation of absolute pitch," J Res Mus Educ **17**, 135–143 (1969).
[35] A. Bachem, "Various types of absolute pitch," J. Acoust. Soc. Am. **9**, 146–157 (1937).
[36] D. J. Levitin, "Absolute memory for musical pitch: evidence from the production of learned melodies," Percept. Psychophys. **56**, 414–423 (1994).
[37] A. R. Halpern, "Memory for the absolute pitch of familiar songs," Mem. Cognit. **17**, 572–581 (1989).
[38] E. G. Schellenberg and S. E. Trehub, "Good pitch memory is widespread," Psychol. Sci. **14**, 262–266 (2003).
[39] D. A. Ross, "Auditory Processes of Musical Pitch Perception," doctoral dissertation, Yale University, 2004.
[40] V. T. Wynn, "Absolute pitch in humans, its variations and possible connections with other known rhythmic phenomena," Prog. Neurobiol. **1**, 111–149 (1973).
[41] T. Kobayashi, K. Nisijima, Y. Ehara, K. Otsuka, and S. Kato, "Pitch perception shift: A rare side-effect of carbamazepine," Psychiatry Clin. Neurosci. **55**, 415–417 (2001).
[42] V. Chaloupka, S. Mitchell, and R. Muirhead, "Observation of a reversible, medication-induced change in pitch perception," J. Acoust. Soc. Am. **96**, 145–149 (1994).
[43] D. A. Ross, I. R. Olson, and J. C. Gore, "Absolute Pitch Does Not Depend on Early Musical Training," Ann. N.Y. Acad. Sci. **999**, 522–526 (2003).

Detection of random alterations to time-varying musical instrument spectra

Andrew Horner[a)]
Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

James Beauchamp[a)]
School and Music and Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, 2136 Music Building, 1114 West Nevada Street, Urbana, Illinois 61801

Richard So[b)]
Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

(Received 10 January 2003; accepted for publication 14 June 2004)

The time-varying spectra of eight musical instrument sounds were randomly altered by a time-invariant process to determine how detection of spectral alteration varies with degree of alteration, instrument, musical experience, and spectral variation. Sounds were resynthesized with centroids equalized to the original sounds, with frequencies harmonically flattened, and with average spectral error levels of 8%, 16%, 24%, 32%, and 48%. Listeners were asked to discriminate the randomly altered sounds from reference sounds resynthesized from the original data. For all eight instruments, discrimination was very good for the 32% and 48% error levels, moderate for the 16% and 24% error levels, and poor for the 8% error levels. When the error levels were 16%, 24%, and 32%, the scores of musically experienced listeners were found to be significantly better than the scores of listeners with no musical experience. Also, in this same error level range, discrimination was significantly affected by the instrument tested. For error levels of 16% and 24%, discrimination scores were significantly, but negatively correlated with measures of spectral incoherence and normalized centroid deviation on unaltered instrument spectra, suggesting that the presence of dynamic spectral variations tends to increase the difficulty of detecting spectral alterations. Correlation between discrimination and a measure of spectral irregularity was comparatively low.
© *2004 Acoustical Society of America.* [DOI: 10.1121/1.1778741]

PACS numbers: 43.75.Cd, 43.66.Jh [SEM] Pages: 1800–1810

I. INTRODUCTION

It is common knowledge that musical instruments can be identified even when their spectra have been substantially altered. A trumpet is recognizable when performed in a vast cathedral or small bathroom, played through an advanced 3D-surround system or cheap PC speakers, or modified in various ways through a spectrum equalizer, even though all of these systems substantially modify the trumpet's source spectrum. What is less obvious is the amount of spectral change required for an audible change of timbre to occur. If the band-levels of a spectrum equalizer are set at random within a range of $\pm x$ dB, at what level of x would a listener begin to distinguish the modified signal from the input signal? Does discrimination vary from instrument to instrument? How much does time-varying spectral variation affect this process? For example, are spectral alterations of a relatively static organlike sound more easily detected than those that have more pronounced spectral variations? Answering these questions is the primary objectives of this paper.

Some related previous studies have considered the effect of simple spectral alterations of static spectra, speech, and audio signals. However, to our knowledge, no previous work has addressed the effect of random spectral variations of individual musical instrument sounds with time-variant spectra.

Plomp (1970) considered the correspondence between an error metric and discrimination data based on static musical instrument and vowel spectra. He concluded that differences in timbre can be predicted well from spectral differences. However, spectral differences between the sounds in that study were not systematically varied from small to large.

Toole and Olive (1988) reviewed previous research on the audibility of resonances and explored the effect of modifying the spectra of noise, music recordings, and speech by adding a single resonance to a unity gain signal path. They stated that ''it is surprising just how much the ... signal ... can be modified without significantly altering perceived timbre.'' They concluded that resonances are more easily heard in noise input signals than in speech and music input signals.

The effect of transmission channel spectral-envelope distortion of vowel sounds in speech was investigated by Watkins (1991) and Watkins and Makin (1996). In the 1991 study, short-test vowel stimuli ranging on a continuum from

a)Address correspondence to either A. Horner at HKUST (electronic mail: horner@cs.ust.hk) or to J. Beauchamp at UIUC (electronic mail: jwbeauch@uiuc.edu).
b)Electronic mail: rhyso@ust.hk

''itch'' to ''etch'' were preceded by a ''carrier'' consisting of a four-word phrase. The carrier phrase was processed by a filter that could turn ''itch'' into ''etch'' in varying degrees. It was found that filtering the carrier phrase would cause subjects to compensate for the filtering and shift their boundary between itch and etch perception. In the 1996 study, spectral distortion was accomplished by increasing the contrast between high-amplitude and low-amplitude portions of the spectrum in such a way that the spectral slope was preserved. By manipulating spectral-envelope distortions of the carrier independently of the test sounds, the authors concluded that listeners compensate for distortion before perceptual features of the vowel are extracted. The principal consequence of this work for the current study is that the perception of a sound can be influenced by the sound that precedes it, suggesting that spectral variations might also have a significant influence on discrimination.

A related area of research is *spectral profile analysis* (Green, 1988). Like random spectrum alteration, spectral profile analysis studies the ability of listeners to discriminate an original stimulus from a spectrally altered version of the stimulus. However, there are significant differences between profile analysis and the approach of the current study. For one, spectral profile analysis usually has only considered static spectra with log-spaced rather than harmonic-spaced components. Also, the reference spectra are usually flat, which is quite different from the spectra of acoustic instruments, which include frequency roll-off. One exception to this is the study by Versfeld and Houtsma (1991) that examined the effect of slope in downward- and upward-ramp-shaped spectra. They found that the detection of spectral changes in a sound is strongly dependent on the frequency spacing of the components, concluding that harmonic spectra would give quite different results than log-spaced spectra. Some researchers have also used ''perturbed'' or ''jagged'' spectra (Kidd *et al.*, 1991). Lentz and Richards (1998) noted that spectral jaggedness increases the threshold of alteration detection.

Another basic difference from our approach is that profile analysis studies have usually only attempted to determine the threshold of discriminating an increase in the amplitude of a single spectral component, or at most a few components in a band. Bernstein *et al.* (1987) observed that thresholds for amplitude changes to a single component cannot be used for accurate prediction of thresholds for changes to multiple components. Hall *et al.* (1984) and Buus (1985) investigated single tone masking by spectra consisting of slowly varying noise or sinusoids. They found that the masking threshold for a single tone was lowered (or ''released'') when masking spectral components were comodulated, meaning that they were amplitude modulated in-phase by a common low-frequency signal. However, Mendoza *et al.* (1996) found that as time-envelopes of narrow-band noise maskers become more incoherent (less correlated), the amount of masking release decreases (i.e., the masking increases).

It is difficult to generalize profile analysis results to the discrimination of broadband alterations to time-varying (dynamic) harmonic spectra, although the above results suggest that changes to multiple components of acoustic spectra are more difficult to detect than are single components. Also, comodulation-masking-release results suggest that alterations to coherent spectra are easier to detect than are alterations to incoherent spectra.

In the present study, we sought to determine discrimination scores for sustained musical instrument sounds with different levels of time-invariant random spectrum alteration. Noting that discrimination varied significantly from instrument to instrument, we then correlated the discrimination scores with spectral incoherence, spectral centroid deviation, and spectral irregularity of the original sound spectra, to determine whether these different measures of spectral variation are significantly related to discrimination. We present the techniques used for analysis and synthesis of the stimuli and then discuss the discrimination experiment. We then present the results in terms of the effect on the discrimination of the instrument, musical experience, and measured spectral variations of the original stimuli.

II. STIMULUS PREPARATION

Eight sustained musical instrument sounds (six of them also used by McAdams *et al.*, 1999) were selected as prototype signals for stimulus preparation. Each entire sound was used as a stimulus, including the attack, sustain, and release. The sounds were first subjected to time-variant spectrum analysis using a computer-based phase vocoder method (Beauchamp, 1993). This phase vocoder is different from most in that it allows a fixed analysis frequency (f_a) to be tuned to an estimated fundamental frequency of the input signal. The analysis method yields frequency deviations between harmonics of the analysis frequency and the corresponding frequencies of the input signal, which are assumed to be approximately harmonic relative to the fundamental. The harmonic frequency deviations are assumed to be within $\pm 2\%$ of the corresponding harmonic of the analysis frequency.

A. Signal representation

Each sound stimulus was represented as a sum of sinusoids with time-varying amplitudes and frequencies:

$$s(t)=\sum_{k=1}^{K} A_k(t)\cos\left(2\pi\int_0^t (kf_a+\Delta f_k(\tau))d\tau+\theta_k(0)\right), \tag{1}$$

where $s(t)$=sound signal, t=time in s, k=harmonic number, K=number of harmonics, $A_k(t)$ is the amplitude of the kth harmonic at time t, f_a=analysis frequency (approximately 311 Hz), $\Delta f_k(\tau)$ is the kth harmonic's frequency deviation, so that $f_k(\tau)=kf_a+\Delta f_k(\tau)$ is the estimated frequency of the kth harmonic, and $\theta_k(0)$ is the initial phase of the kth harmonic.

The parameters used for resynthesis in this study are f_a and $A_k(t)$. The frequency deviations, $\Delta f_k(t)$, were set to zero in order to restrict listener attention to the amplitude data. Although the $A_k(t)$ were only stored as samples occurring every $T_o=1/(2f_a)$ s, the synthetic signals were gener-

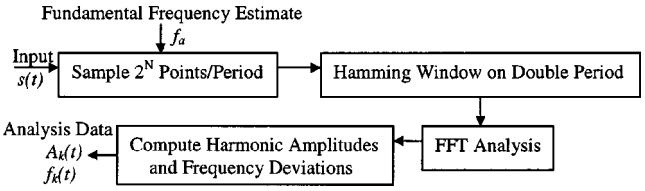


FIG. 1. Method for time-varying spectral analysis that yields the amplitude and frequency deviations for each harmonic, k. The estimated frequency for harmonic k is given by $f_k(t)=kf_a+\Delta f_k(t)$, where f_a is the analysis fundamental frequency.

ated at a much higher data rate (sample frequency 22 050 or 44 100 Hz) by using linear interpolation between these values.

B. Prototype instrument sounds

Sounds of a bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin were used to represent the several wind and the bowed string instrument families. Five of the sounds were taken from the McGill University Master Samples recordings, two from Prosonus (bassoon and oboe), and one (trumpet) had been recorded at the University of Illinois Urbana–Champaign School of Music. Except for the bassoon and horn, these sounds were also used by McAdams *et al.* (1999), who give more details about their characteristics.

C. Analysis method

The phase vocoder method used for analysis consists of the following steps (see Fig. 1):

(1) Band-limited interpolation of the input signal to produce the lowest possible power-of-two number of samples per analysis period ($T_a=1/f_a$), which exceeds the number of original samples in this time interval.
(2) Segmentation of the input signal into contiguous frames whose lengths are equal to twice the analysis period ($2T_a$) and which overlap by half an analysis period ($T_o=T_a/2$, where T_o is the time between frames). This amount of overlap minimally satisfies the sample-rate/bandwidth criterion (Allen, 1977).
(3) Multiplication of each signal frame by a Hamming window function whose length is two analysis periods ($2T_a$). When f_a is tuned to the input signal's fundamental frequency, the Hamming window provides separation of neighboring harmonics by at least 42 dB.
(4) Fast Fourier transform (FFT) of the resulting product to produce real and imaginary components (a_k and b_k) at frequencies 0, $f_a/2$, f_a, $3f_a/2$,...,$f_s/2-f_a$, where f_s is the sampling frequency. Components that are not positive integer multiples of f_a are discarded.
(5) Conversion of each retained real and imaginary component pair to give the amplitude and phase of each harmonic [$A_k=\sqrt{a_k^2+b_k^2}$; $\theta_k=\text{atan}\,2(b_k,a_k)$].
(6) Computation of the frequency deviation for each harmonic by a trigonometric identity, which essentially gives the difference in phase between adjacent frames for each harmonic. The phase difference divided by $2\pi T_o$ gives the frequency deviation.
(7) Storage of the harmonic amplitude and frequency-deviation data in an "analysis file." The number of harmonics stored is less than $f_s/(2f_a)$. The analysis file for each sound is the basis for further sound processing.

For $f_s=44\,100$ Hz and $f_a=311.1$ Hz (Eb4), the maximum number of harmonics that can be analyzed is 70. For $f_s=22\,050$ Hz, this reduces to 35. Because harmonic amplitudes were judged (by visual inspection of spectra) to be insignificant beyond $K=35$ for the bassoon, oboe, and trumpet sounds, $f_s=22\,050$ Hz was used for these. The other sounds were sampled at 44 100 Hz.

Further details on the analysis procedure are discussed by Beauchamp (1993).

The analysis system may be viewed as a set of contiguous bandpass filters that have identical bandwidths (f_a) and are centered at the harmonics of the analysis frequency (f_a). The basic assumption is that the signal consists of harmonic sine waves whose frequencies line up with the filter band centers, so that the output of each filter is one of the sine waves. The analysis gives the amplitude and frequency of each sine wave. When the filter outputs are summed, the signal is almost perfectly reconstructed. In fact, the output signal can be viewed as that created by processing the input signal by the sum of the bandpass-filter characteristics. For the Hamming window, it can be shown that this sum varies only by 1.4 dB over the range [$f_a/2$,$f_s/2$], with maxima occurring at the harmonic frequencies and minima at the halfway points. Figure 1 shows a block diagram of the basic analysis/synthesis system.

D. Duration and loudness equalization

Sound duration is a potential factor in discrimination. For example, a sound lasting 5 s might well be easier to discriminate than a sound lasting 0.3 s. In order that duration would not be a factor in the study, the sounds were standardized to a 2-s duration by interpolating the analysis data. Next, the eight duration-equalized prototype sounds were compared, and amplitude multipliers were determined such that the sounds were judged to have equal loudness. McAdams *et al.* (1999) give more details on these equalizations. It should be mentioned, however, that these equalizations were not essential for the present study, since each discrimination pair was always derived from a single prototype sound.

E. Frequency flattening

In order that they would not be factors in this study, frequency variations and inharmonicity were eliminated from the sounds by setting each harmonic's frequency equal to the product of its harmonic number (k) and the fixed analysis frequency (f_a), resulting in flat, equally spaced frequency envelopes; i.e.,

$$f_k=kf_a. \tag{2}$$

Frequency flattening was previously shown to have an effect on discrimination by Gray and Moorer (1977), Charbonneau (1981), and McAdams *et al.* (1999). For the five sustain instruments they tested, McAdams *et al.* found that frequency

flattening resulted in discrimination scores in the range of 48% to 82% and that discrimination is weakly correlated (34% to 57%) with the amount of frequency deviation in the original sound. In the current study, frequency flattening was employed to focus listeners' attention on harmonic amplitude variations and also to avoid frequency fluctuations which may influence subjects' judgments, especially considering that audible artifacts can result from these fluctuations when they are amplified by altered harmonic amplitudes. The frequency-flattened sounds then served as the reference sounds for this study, and their corresponding time-varying harmonic amplitude spectra are henceforth referred to as the *analysis data*.

F. Random spectrum alteration

Time-invariant random alteration was performed on the analysis data, after which the sounds were regenerated by additive synthesis. Randomly altered harmonic amplitudes are indicated with the prime symbol, e.g., $A_k'(t)$. Alteration was accomplished by multiplying each harmonic amplitude by a randomly selected scalar, r_k:

$$A_k'(t) = r_k A_k(t). \tag{3}$$

The $\{r_k\}$ were selected uniformly in the range $[1-2\varepsilon, 1+2\varepsilon]$, where ε is referred to as the *error level*. Equation (3) describes a linear stationary process, as the value of each r_k remains the same throughout the sound. The goal of this method is to perturb each harmonic amplitude so that a desired overall spectral error is achieved, without changing the spectral centroid or rise time.

The *relative-amplitude spectral error* is defined as

$$\varepsilon' = \frac{1}{N}\sum_{n=0}^{N-1} \sqrt{\frac{\Sigma_{k=1}^{K}(A_k(t_n) - A_k'(t_n))^2}{\Sigma_{k=1}^{K} A_k^2(t_n)}}, \tag{4}$$

where n is the analysis frame number, $t_n = nT_o$ is the frame time, and N is the number of frames used in the calculation. For this study, $N=20$, where 10 values are taken from the perceptually important attack and 10 values are taken from the remainder of the sound, both equally spaced in time. The attack time is defined as the time from the beginning of the sound to the sound's maximum rms amplitude. Equation (4) is similar to the spectral distance measure used by McAdams *et al.* (1999), except that the normalization is slightly different and fewer amplitude values are used.

Under the definitions of Eqs. (3) and (4), the relative-amplitude spectral error usually varies between 0 and 1. Note that Eq. (4) computes the average of the spectral error at each instant relative to the rms amplitude at that instant. Due to the rms amplitude in the formula's denominator, low-amplitude portions of the sound are given the same weight as high-amplitude portions. (It is assumed that proportional-amplitude errors are more relevant than absolute-amplitude errors.) The normalized-squared errors are then accumulated over the harmonics and averaged over time. One could argue that the error metric might be perceptually improved by first accumulating amplitudes by critical bands before averaging, but we have investigated this possibility and, for a similar set of sound signals, have found no improvement in terms of correlation with discrimination results.

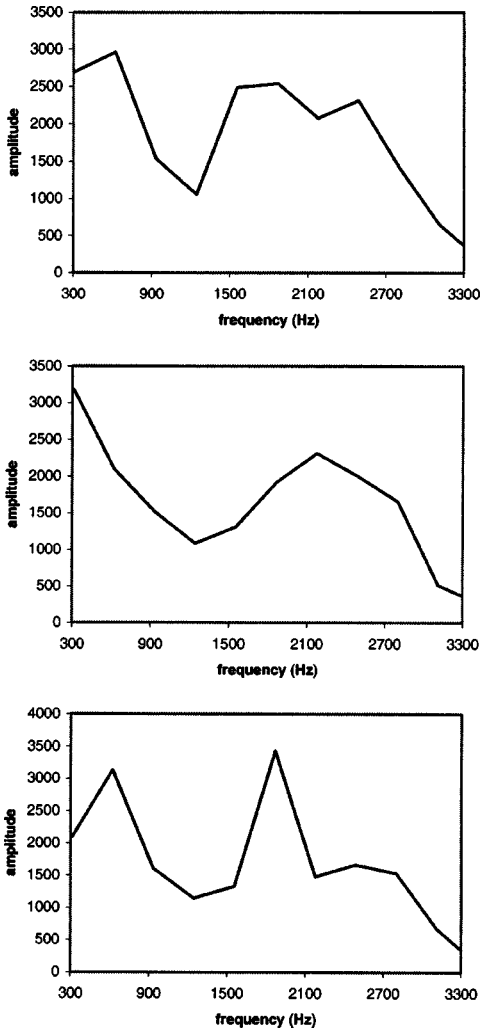


FIG. 2. Original trumpet sound spectral envelope (left) and two spectral envelopes (center and right) generated by random spectrum alteration of the original sound resulting in approximately 24% relative-amplitude spectral error.

Since the $\{r_k\}$ were selected uniformly in the range $[1-2\varepsilon, 1+2\varepsilon]$, the relative-amplitude spectral error was expected to be approximately ε. However, the relative-amplitude error, ε', as defined by Eq. (4), is slightly different from ε. For our purposes, we randomly generated the $\{r_k\}$ and then calculated ε'. If ε' was in the range $[\varepsilon-0.01, \varepsilon+0.01]$, we accepted the $\{r_k\}$; otherwise, we continued to generate and test the $\{r_k\}$ until ε' fell in the desired range.

For example, Fig. 2 shows an original spectral envelope and two randomly altered spectral envelopes, each with 24% error. Note that the first altered spectral envelope is smoother than the original, while the second one is more irregular than the original.

Spectral centroid has been shown to be strongly correlated with one of the most prominent dimensions of timbre as derived by multidimensional scaling (MDS) experiments (Gray and Gordon, 1978; Wessel, 1979; Krumhansl, 1989; Iverson and Krumhansl, 1993; Krimphoff *et al.*, 1994; Kendall and Carterette, 1996; Lakatos, 2000). Normalized cen-

troid versus time functions for original and altered sounds are defined as

$$NSC(t_n)=\frac{\Sigma_{k=1}^{K} kA_k(t_n)}{\Sigma_{k=1}^{K} A_k(t_n)} \quad \text{and}$$

$$NSC'(t_n)=\frac{\Sigma_{k=1}^{K} kA'_k(t_n)}{\Sigma_{k=1}^{K} A'_k(t_n)}. \tag{5}$$

Preserving the spectral centroid after random spectral alteration has been applied eliminates this parameter from being a correlate of discrimination. Therefore, the $\{r_k\}$ were only accepted if the peak (over time) centroid of the original and altered spectra were matched within $\pm 2.5\%$.

The final spectrum alteration algorithm then consists of the following steps:

(1) Pick a vector $\{r_k\}$ such that, for each harmonic k, $1-2\varepsilon<r_k<1+2\varepsilon$.
(2) Generate modified harmonic amplitude envelopes, $A'_k(t)=r_kA_k(t)$.
(3) Calculate the relative-amplitude spectral error, ε' [see Eq. (4)].
(4) If ε' is outside the range $[\varepsilon-0.01,\varepsilon+0.01]$, go to step (1) to repick $\{r_k\}$.
(5) Calculate the peak spectral centroids of the original and randomly altered sounds [see Eq. (5)].
(6) If the peak centroid NSC' of the altered spectrum is outside the range $[0.975\cdot NSC_{\text{peak}},1.025\cdot NSC_{\text{peak}}]$, go to step (1) to repick $\{r_k\}$.
(7) End.

G. Resynthesis method

Resynthesis was accomplished by additive (or Fourier) synthesis of the altered harmonic sine waves:

$$\hat{s}(t)=\sum_{k=1}^{K} A''_k(t)\cos(2\pi k f_a t+\theta_k(0)), \tag{6}$$

where $A''_k(t)$ is the linear interpolation of the $A'_k(t_n)$ amplitude envelope data between frames, so that

$$A''_k(t)=\frac{t_{n+1}-t}{T_o}A'(t_n)+\frac{t-t_n}{T_o}A'_k(t_{n+1}),$$
$$t_n\leqslant t\leqslant t_{n+1}. \tag{7}$$

While linear interpolation introduces a smoothing operation to the harmonic amplitude functions, it has a very minor effect on the resulting sound. Also, both the randomly altered sounds and the original reference sounds are resynthesized by additive synthesis in the same fashion, with strictly fixed harmonic frequencies, before being compared in the listening experiment.

III. EXPERIMENTAL METHOD

A. Subjects

Twenty listeners participated in the experiment. They were undergraduate students at the Hong Kong University of Science and Technology, ranging in age from 20 to 23 years, who reported no hearing problems. They included 17 males and 3 females, and their experience playing a musical instrument was recorded. For the 13 listeners with musical experience, the range in musical experience varied from 1 to 5 years with a mean of 2.5 years. The listeners were paid to compensate for their time spent in the experiment.

B. Stimuli

The eight musical instruments used belong to the air column (air reed, single reed, lip reed, double reed) and bowed string families: bassoon, clarinet, flute, horn, oboe, saxophone, trumpet, and violin. Each sound was analyzed and resynthesized using the reference analysis data with no frequency variations and no inharmonicity. The fixed harmonic frequencies encouraged listeners to focus their attention exclusively on the amplitude data. They were prevented from detecting cues stemming from the amplitudes of originally weak sinusoids with noisy harmonic frequency variations possibly made more audible by random spectral alteration. Also, the original sustained sounds had very nearly strictly harmonic frequencies with relatively small frequency deviations.

The sounds were stored in 16-bit integer format on a hard disk. All ''reference sounds'' (resynthesized using the analysis data with strictly fixed harmonic frequencies) were equalized for duration and loudness. The randomly altered sounds for each instrument were resynthesized with the method described in Sec. II G on an Intel PC.

Ten different randomly altered sounds were synthesized for each of five spectral error levels (8%, 16%, 24%, 32%, and 48%), yielding a total of 50 modified sounds for each instrument. The initial random multipliers, r_k, used were the same for each of the instruments. However, the random-alteration algorithm guaranteed that the relative-amplitude spectral errors, as measured by Eq. (4), were within $\pm 1\%$ of the prescribed error levels while matching the reference sounds in duration, fundamental frequency, and peak centroid. Like the reference sounds, the randomly altered sounds were also synthesized using strictly fixed harmonic frequencies. Also, using the Moore–Glasberg loudness program (Moore *et al.*, 1997), it was determined that loudnesses of the altered sounds matched those of the reference sounds within 2 phons.

C. Test procedure

A two-alternative forced-choice (2AFC) discrimination paradigm was used. The listener heard two pairs of sounds and chose which pair was ''different.'' Each trial structure was one of AA-AB, AB-AA, AA-BA, or BA-AA, where A represents the reference sound and B one of the 50 randomly altered sounds. This paradigm had the advantage of not being as susceptible to variations in listeners' criteria across experimental trials compared to the simpler A-B method. All four combinations were presented for each randomly altered sound. The two 2-s sounds of each pair were separated by a 500-ms silence, and the two pairs were separated by a 1-s silence. For each trial, the user was prompted with ''which

pair is different, 1 or 2?'' and the response was given by the keyboard. The computer would not accept a response until all four sounds in a trial had been played.

For each instrument, a block of 200 trials was presented to each of the listeners (four trial structures×5 error levels×10 different random alterations). The order of presentation of these 200 trials was randomized. For each random alteration, discrimination performance was averaged using the results of the four trials for each listener. Because these four trials were presented in random order within the 200 trials, the effects of possible learning were averaged out. The same trials were presented to each listener, although in a different random order. The duration of each block was about 40 min, and listeners took 5–10-min breaks after each block. Eight blocks were presented to each listener, corresponding to the eight instruments. The order of presentation of the instruments was also randomized for each listener. Sessions of four blocks were scheduled on two different days. The average time to complete a block was 45 min (including breaks). The total duration of the experiment was about 6 h. A custom program written at HKUST ran on an Intel PC to control the experiment.

Listeners were seated in a ''quiet room'' with less than 40 dB SPL background noise level (mostly due to computers and air conditioning). The headphones also provided some reduction of the noise level. Sound signals were converted to analog by a SoundBlaster Audigy soundcard and then presented through Sony MDR-7506 headphones at a level of approximately 75 dB SPL as measured with a sound-level meter. The Audigy DAC utilized 24 bits with a maximum sampling rate of 96 kHz and a 100 dB S/N ratio. The sounds were actually played at 22.05 or 44.1 KHz.

At the beginning of the experiment, the listener read the instructions and asked any necessary questions of the experimenter. Five test trials (chosen at random from the instruments) were presented before the data trials for each instrument.

IV. SPECTRAL CORRELATE MEASURES

Discrimination of spectral alterations may be affected by some aspect of the spectral variations of the original sounds. Several different measures of these variations are possible. Gray (1977) referred to ''spectral fluctuation'' as an interpretation of a dimension revealed by multidimensional scaling (MDS) of musical instrument dissimilarity judgments. ''Spectral flux'' was qualitatively described by Krumhansl (1989) as ''the temporal evolution of the spectral components'' and by McAdams *et al.* (1999) as ''the change in shape of a spectral envelope over time.'' Krimphoff (1993) defined *spectral variation* as the average correlation between adjacent time points in the spectrum:

$$VS = \frac{1}{N}\sum_{n=1}^{N} \frac{\Sigma_{k=1}^{K} A_k(t_{n-1})A(t_n)}{A_{rms}(t_{n-1})A_{rms}(t_n)}, \tag{8}$$

where

$$A_{rms}(t_n) = \sqrt{\sum_{k=1}^{K} A_k^2(t_n)} \tag{9}$$

is the instantaneous rms amplitude and N is the total number of frames. With the definition of Eq. (8), VS is theoretically between 0 and 1, but because spectral amplitudes vary slowly, VS is expected to be close to 1, with closeness to 1 being an indicator of how static the signal is. Instead of this measure, we opted to use a spectral incoherence measure as defined by Eqs. (11) and (12) below.

Krimphoff also defined ''flux'' as the root-mean-squared deviation of the normalized spectral centroid over time given by

$$FL = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} (NSC(t_n) - NSC_{xx})^2}, \tag{10}$$

where NSC_{xx} could be the average, rms, or maximum value of NSC. A time-average value was used in the current study.

Beauchamp and Lakatos (2002) measured spectral fluctuation in terms of *spectral incoherence*, a measure of how much a spectrum differs from a coherent version of itself. Larger incoherence values indicate a more dynamic spectrum, and smaller values indicate a more static spectrum. A perfectly static spectrum has an incoherence of zero.

The coherent spectrum is defined to have the same average spectrum as the original and the same instantaneous rms amplitude, but, unlike the original, all harmonic amplitudes vary in time proportional to the rms amplitude and, therefore, in fixed ratios to each other. Put another way, harmonic amplitudes normalized by the rms amplitude are fixed. The coherent version of the kth harmonic amplitude is defined by

$$\hat{A}_k(t_n) = \frac{\overline{A}_k A_{rms}(t_n)}{\sqrt{\Sigma_{k=1}^{K} \overline{A}_k^2}}, \tag{11}$$

where $\overline{A}_k$ is the time-averaged amplitude of the kth harmonic. Then, spectral incoherence of the original spectrum is defined as

$$SI = \left(\frac{\Sigma_{n=0}^{N-1}\Sigma_{k=1}^{K}(A_k(t_n) - \hat{A}_k(t_n))^2}{\Sigma_{n=0}^{N-1}(A_{rms}(t_n))^2} \right)^{1/2}. \tag{12}$$

With the definitions of Eqs. (11) and (12), spectral incoherence (SI) varies between 0 and 1 with higher values indicating more incoherence (more dynamic). Both VS and SI are measures of spectral fluctuation, but FL is a measure of the normalized centroid change over time. Since it is possible for FL to be significant while SI is relatively small (or VS is close to 1) and for SI to be large while FL is small, SI (or VS) and FL are approximately independent measures, although they may be correlated for a particular set of musical sounds.

Krimphoff (1993) also introduced the concept of *spectral irregularity*, which was redefined by Beauchamp and Lakatos (2002) as

$$SIR = \frac{1}{N}\sum_{n=0}^{N-1} \frac{\Sigma_{k=2}^{K-1} A_k(t_n)\left|A_k(t_n) - (A_{k-1}(t_n) + A_k(t_n) + A_{k-1}(t_n))/3\right|}{A_{rms}(t_n)\Sigma_{k=2}^{K-1} A_k(t_n)}. \quad (13)$$

This formula defines the difference between a spectrum and a spectrally smoothed version of it, averaged over both harmonics and time and normalized by rms amplitude. It may be hypothesized that spectra that are irregular or jagged to begin with are more likely to be confused by listeners with randomly altered versions of the spectra.

In general, the expectation is that the more complex an original signal is, either in terms of spectral fluctuation over time or in terms of spectral jaggedness, the less perceptible the effect of random alteration on a spectrum would be.

V. RESULTS

A. Data analysis methods

A test of normality indicated that the discrimination data were not normally distributed even after various transformations (e.g., sqrt[arcsin(data)]) (Stevens, 2002). Therefore, the data were analyzed using both parametric and nonparametric statistical tests (parametric: ANOVAs, Student Newman–Keuls tests; nonparametric: Friedman ANOVAs by ranks, Wiloxcon signed-ranks tests, Kruskal–Wallis tests, and Mann–Whitney U tests). The results of parametric and nonparametric tests were found to be similar for the overall main effects and two-way interaction effects, although they differed in some detailed analyses (e.g., statistical groupings of data using parametric Student Newman–Keuls tests were sometimes different from the results of Wilcoxon signed-ranks tests). In the rest of the paper, the main effects and the two-way interactions were tested with both ANOVAs and nonparametric tests. For detailed analyses of interactions, results of nonparametric tests are presented because the data were not normally distributed.

B. Effects and interactions of error level, instrument, and musical experience

Discrimination scores were computed for each random alteration across the four trial structures for each listener. Because the presentation order of each of the four trials was randomized, any potential effects of learning were averaged out. Figure 3 shows the scores averaged over the 20 listeners for the 50 random alterations (ten versions of each of five error levels) on the eight instruments. It can be observed that as the error level increases from 8% to 48%, the mean discrimination increases asymptotically towards 100%. For 8% error levels, most scores are in the range 50% to 60%. The ranges are wider and more variable for intermediate errors such as 16% error, where the scores are between 60% and 85% (except for the trumpet, which is about 10% lower). For 24%, 32%, and 48% error, most scores are between 80% and 95%, 88% and 100% and 92% and 100%, respectively.

Figure 4 shows a graph of average discrimination versus error level for each instrument. The scores are in close agreement, especially for the 8%, 32%, and 48% error cases. Even for 16% and 24% error levels, the average discriminations for the various instruments are within 12% of one another.

ANOVA analyses of the results used the ten different random alterations for each instrument as repeated measures to test the main effects of instrument (eight instruments), error level (five levels: 8%, 16%, 24%, 32%, 48%), musical experience (two groups: no musical experience, at least 1 year of musical experience) and their two-way interactions (see Table I).

Inspection of Table I indicates that both experience and instrument, and instrument and error level, have strong significant two-way interactions and experience and error level have weak significant interactions. Further analyses of the weak interactions between the effects of musical experience and the effects of error level indicate that for all 16 combinations of instrument and experience, the effects of error level are significant and the trends are the same ($p < 0.0001$, Friedman ANOVAs by ranks). This suggests that the effects of error level are strong and consistent for all conditions.

Mean discrimination scores for listeners with no experience in playing a musical instrument and listeners with at least one year of experience as functions of the five error levels are compared in Fig. 5. Mann-Whitney U tests were conducted to compare the discrimination scores between listeners with and without musical experience at each error level. Consistent with Fig. 5, no significant effects of experience were found when error levels were either 8% or 48% ($p > 0.05$). However, when error levels were 16%, 24%, and 32%, the scores of musically experienced listeners were found to be significantly higher than those of listeners with no musical experience ($p < 0.001$, Mann-Whitney U tests).

Since there were significant two-way interacting effects with experience and instrument, and instrument and error level, Friedman ANOVAs by rank and Wilcoxon signed-ranks tests were conducted to examine the interaction patterns between instrument and error level for data collected from listeners with and without musical experience. The results of these analyses are presented in Table II. Table II shows that when the error levels were 8% or 48%, instrument did not have a significant effect on discrimination. However, when the error levels were 16%, 24%, or 32%, instrument did have a significant effect. Also, for all listeners regardless of their musical experience, the violin and trumpet were consistently associated with lower discrimination scores, and the bassoon and horn were consistently associated with higher discrimination scores.

C. Correlation of spectral incoherence, spectral centroid change, and spectral irregularity with discrimination

Since discrimination varies significantly with instrument for intermediate error levels, we correlated the discrimination

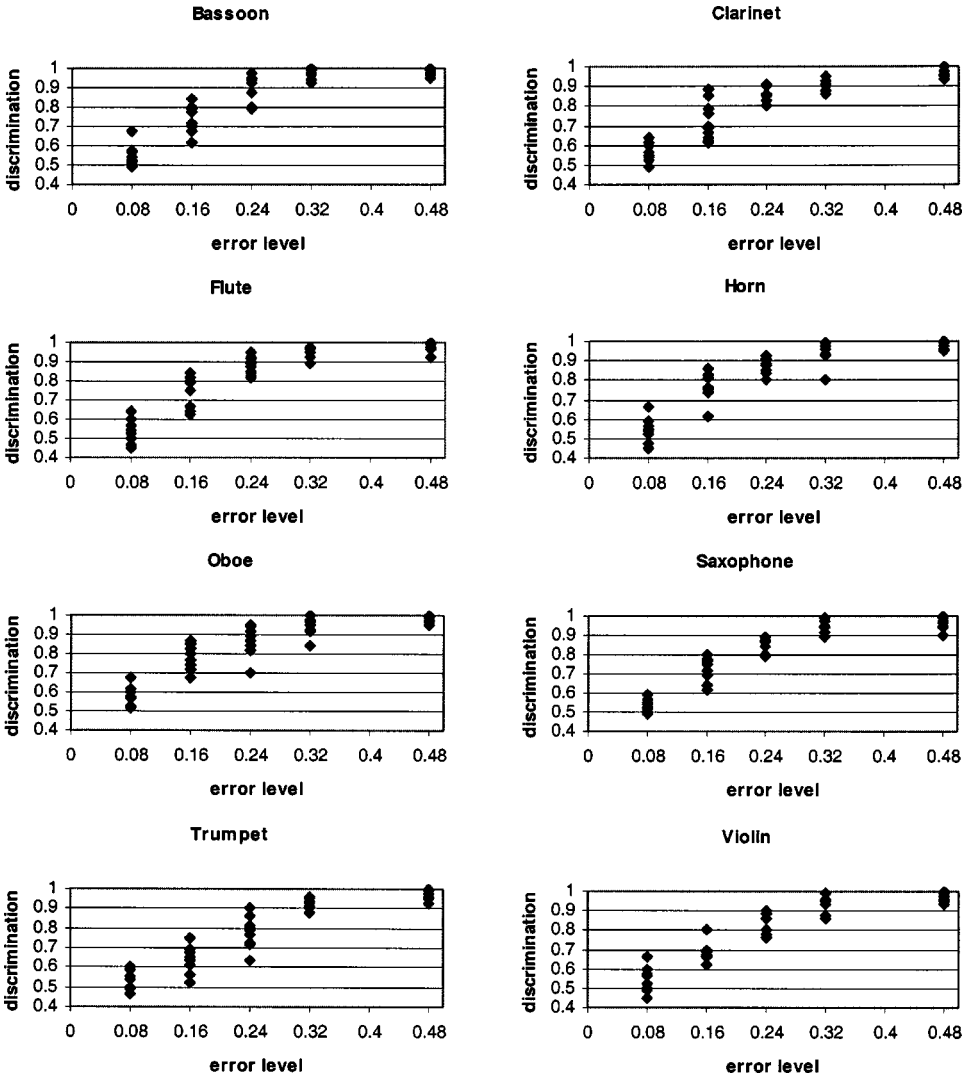


FIG. 3. Mean discrimination scores for ten randomly altered sounds versus error level for the eight instruments.

scores with spectral incoherence, normalized centroid deviation, and spectral irregularity to determine whether these different measures of spectral variation were significantly related to discrimination. Table III gives spectral incoherence, spectral centroid deviation, spectral irregularity, and average discrimination values for error levels of 16% and 24% for each of the eight instruments.

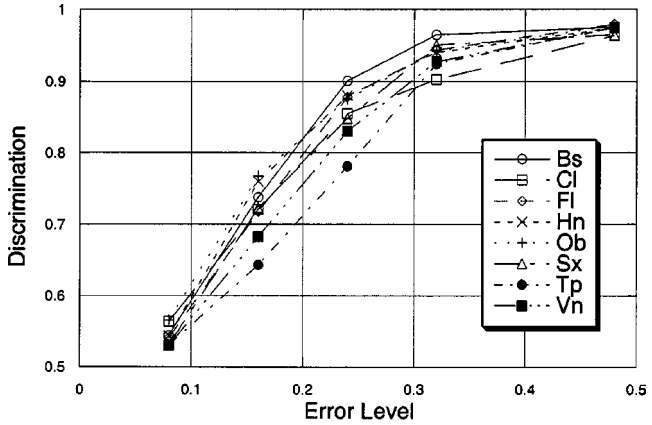


FIG. 4. Average discrimination scores versus error level for the eight instruments.

Table III shows that the trumpet and violin have the largest incoherence values, as well as the lowest discrimination scores. Conversely, the bassoon and horn have among the lowest spectral incoherence values and the highest discrimination scores. This suggests that spectral incoherence is related to difficulty in detecting random spectral alterations. When discrimination scores were averaged over the ten alterations and the 20 subjects (as shown in Fig. 4), strong negative correlations were found for the 16% and 24% error levels [$r(6)=-0.91$, $p<0.005$ at 16%; $r(6)=-0.64$, $p<0.1$ at 24%, Spearman correlation].

Correlations between discrimination scores and normalized centroid deviations were also calculated. Again, strong negative correlations were found [$r(6)=-0.72$, $p<0.05$ at 16%; $r(6)=-0.85$, $p<0.01$ at 24%, Spearman correlation]. On the other hand, the correlations between discrimination scores and spectral irregularity were relatively weak and not statistically significant [$r(6)=0.12$, $p=0.78$ at 16%; $r(6)=-0.17$, $p=0.69$ at 24%, Spearman correlation].

VI. DISCUSSION

The results presented in the previous section show that, for the instruments tested, both spectral incoherence and

TABLE I. ANOVA table illustrating the main effects and two-way interactions of instrument (eight instruments), error level (8%, 16%, 24%, 32%, and 48% error), and musical experience (no experience versus 1 year or more) on the data collected from 20 listeners participating in the discrimination experiment. Data are the percentage of correct discrimination scores (100%, 75%, 50%, 25%, and 0%) over each group of four trials containing the same reference and altered sounds. For each error level, there were 40 trials with 10 different random spectral alterations in the test sounds.

Source	DF	Sum of squares	Mean square	F value	Pr>F
Instrument	7	40.56	5.59	10.69	0.0001[a]
Error level	4	3186.14	796.53	1469.79	0.0001[b]
Musical experience	1	15.61	15.61	28.80	0.0001[c]
Instrument and error level	28	42.03	1.50	2.77	0.0001[d]
Experience and instrument	7	16.14	2.31	4.26	0.0001[d]
Experience and error level	4	5.36	1.34	2.47	0.0425[e]
Measurement error	7948	4307.31	0.54		
Corrected total	7999	7613.14			

[a]The significant main effects were confirmed with nonparametric Friedman ANOVA by ranks: $p<0.0001$, chi-square=3337.4.
[b]The significant main effects were confirmed with nonparametric Friedman ANOVA by ranks: $p<0.0001$, chi-square=77.87.
[c]The significant main effects were confirmed with nonparametric Kruskal–Wallis test: $p<0.001$, chi-square =50.3.
[d]Nonparametric statistical tests also indicated a significant two-way interacting effect.
[e]Nonparametric statistical tests did not indicate a significant two-way interacting effect.

spectral centroid deviation are strongly correlated with difficulty of detecting random spectral alterations. While it seemed reasonable to expect that increased incoherence would make detection of random spectral alterations more difficult, it was not obvious that increased centroid deviation would have the same effect. Even though spectral incoherence and centroid deviation are theoretically quite independent, in our case they appear to be tightly correlated measures of time-variant spectral variation. On the other hand, spectral irregularity was expected to affect discrimination, but correlations between this parameter and discrimination were found to be small. For example, the trumpet and violin have similar spectral incoherences and centroid deviations and similar 16% and 24% error level discrimination scores, but the trumpet has a much lower spectral irregularity than the violin has. Time-varying spectral variations seem to be more important than ''jaggedness'' of a spectrum with respect to difficulty of detecting spectral alterations.

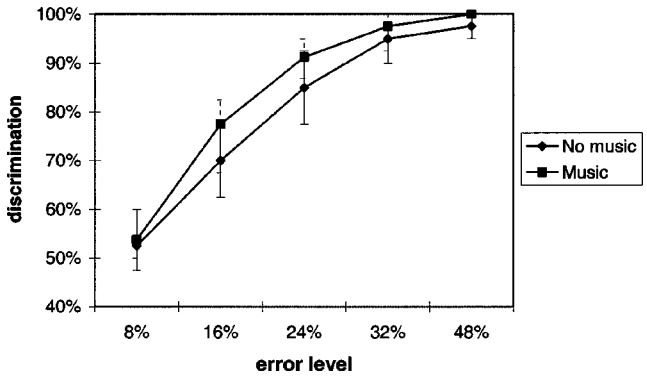


FIG. 5. Median of discrimination scores collected from listeners with and without musical experience as functions of error level (8%,16%,24%,32%,48%). Scores from the ten repeated runs per each condition have been averaged first before the median was calculated. The interquartile ranges are also shown.

From Fig. 4 it can be seen that the 16% error level roughly corresponds to a 75% discrimination threshold, where subtle differences between original and randomly altered sounds are sometimes detected. However, at the 24% error level, we have found (in informal listening tests) that random spectral alteration produces interesting collections of ''similar, yet different'' sounds compared to the originals. Random spectrum alteration could then provide an efficient time-invariant method for generating diverse sets of musical sounds with the same temporal behavior and spectral centroid. This method may possibly provide an improvement for existing synthesizers, which have often been criticized as sounding too tepid, presumably because repeated notes, played at the same amplitude, typically sound exactly the same. Knowing how to change sounds rapidly by random spectrum alteration so that the sounds sound subtly different from one note to the next may result in a process for production of more dynamic and interesting synthetic musical sound sequences.

The distribution of error among the harmonics may make a large difference in the perceptual effect of spectral modification. A previous spectral matching study using a method that often lumped most of the error into one or two of the most prominent harmonics (Horner, 2001) found 75% discrimination thresholds at about 8% error. This is approximately the profile analysis result found by Kidd *et al.* (1991), who observed that jagged spectra with a single altered partial have 75% discrimination thresholds at error levels roughly corresponding to 8%. The current study found that evenly distributing the error among all the partials appears to double the 75% discrimination threshold approximately to 16% error. This indicates that errors distributed over all the harmonics are more difficult to detect than those lumped into one or two of the most prominent harmonics.

TABLE II. This table presents the effects of instrument on the ability of listeners with and without musical experience to discriminate randomly altered tones with five levels of error. Within each column, the instruments (BS: bassoon; CL: clarinet; FL: flute; HN: horn; OB: oboe; SX: saxophone; TP: trumpet; VN: violin) are listed in descending order of discrimination scores collected from 20 listeners for the five error levels (8%, 16%, 24%, 32%, and 48%). The table shows the results of Wilcoxon signed-ranks tests and Friedman ANOVA tests, where instruments on the ranked lists associated with similar (i.e., not significantly different at $p=0.05$ level) data were grouped into the same group as represented by the same capital letter (''A'' to ''D'').

Listeners with no musical experience

8%		16%				24%					32%				48%	
SX	A	OB	A			BS	A				SX	A			FL	A
BS	A	HN	A	B		HN	A	B			BS	A	B		OB	A
VN	A	SX	A	B		SX	A	B			FL	A	B		HN	A
OB	A	BS	A	B		OB	A	B	C		HN	A	B		SX	A
CL	A	FL		B		FL	A	B	C		OB	A	B	C	TP	A
FL	A	VN		B		CL		B	C		TP		B	C	VN	A
TP	A	CL		B	C	VN			C	D	VN			C	BS	A
HN	A	TP			C	TP				D	CL			C	CL	A

Listeners with musical experience

8%		16%					24%					32%				48%	
OB	A	HN	A				BS	A				BS	A			BS	A
CL	A	OB	A	B			HN	A				VN	A	B		VN	A
HN	A	CL	A	B			FL	A	B			OB	A	B	C	HN	A
FL	A	BS	A	B	C		CL	A	B			HN		B	C	FL	A
TP	A	FL	A	B	C		OB	A	B	C		FL		B	C	TP	A
BS	A	SX		B	C	D	VN		B	C	D	SX		B	C	CL	A
VN	A	VN			C	D	SX			C	D	TP		B	C	OB	A
SX	A	TP				D	TP				D	CL			C	SX	A

VII. CONCLUSIONS

The results of the current study show a significant monotonic relationship between discrimination and the amount of random spectrum alteration. The trend of increasing discrimination scores with increasing error level is consistent for all combinations of musical experience and instrument, although the discrimination scores are significantly higher for listeners with at least one year of musical experience when the errors are 16%, 24%, and 32%. Significant interacting effects between error level and instrument were found and investigated. When error levels were 16%, 24%, and 32%, musically experienced listeners discriminated significantly better than those with no experience and instrument had a significant effect.

In general, discrimination was poor for sounds with less than 16% error, the 75% discrimination threshold. Sounds with 16% and 24% random alteration yielded the widest range of discrimination, while sounds with 32% and 48% error levels were clearly and consistently distinguishable by most listeners. Random spectrum alteration leads to lower discrimination scores than other spectral modification methods, such as profile analysis, which do not evenly distribute the error among all harmonics.

Listeners had more difficulty discriminating alterations to instrument sounds containing more pronounced spectral variations. Average spectral incoherences and spectral centroid deviations were found to have a strong negative correlation with average discrimination scores. This suggests that dynamic spectral variations result in increased difficulty of detecting spectral alterations. Such a high correlation was not found for spectral irregularity, a measure of the ''jaggedness'' of a spectrum.

TABLE III. Spectral incoherence, normalized centroid deviation, spectral irregularity, and average 16% and 24% error level discrimination scores for the eight instruments.

Instrument	Spectral incoherence	Normalized centroid deviation	Spectral irregularity	16% error average discrimination	24% error average discrimination
HN	0.057	0.2	0.073	0.765	0.889
OB	0.069	0.8	0.137	0.761	0.871
BS	0.075	0.4	0.093	0.738	0.901
CL	0.085	0.7	0.174	0.721	0.855
SX	0.101	0.6	0.195	0.723	0.848
FL	0.118	0.6	0.129	0.718	0.878
TP	0.184	1.6	0.039	0.643	0.781
VN	0.193	1.4	0.131	0.683	0.830

ACKNOWLEDGMENTS

This work was supported in part by the Hong Kong Research Grants Council's CERG Project Nos. HKUST6194/02E and HKUST6167/03E. We would like to thank Francis Hing-Cheung Wong for running subjects in the control experiment, and Simon Cheuk-Wai Wun for writing the program for Intel personal computers used for the listening tests. Thanks to the anonymous reviewers for their careful and helpful comments.

Allen, J. B. (**1977**). ''Short term spectral analysis, synthesis, and modification by discrete Fourier transform,'' IEEE Trans. Acoust., Speech, Signal Process. **25**, 235–238.

Beauchamp, J. W. (**1993**). ''Unix workstation software for analysis, graphics, modifications, and synthesis of musical sounds,'' 94th Conv. Audio Eng. Soc. (Berlin), Audio Eng. Soc. Preprint 3479 (L-I-7).

Beauchamp, J. W., and Lakatos, S. (**2002**). ''New Spectro-Temporal Measures of Musical Instrument Sounds Used for a Study of Timbral Similarity of Rise-Time- and Centroid-Normalized Musical Sounds,'' in Proc. 7th Int. Conf. on Music Perception and Cognition, Univ. of New South Wales, Sydney, Australia, pp. 592–595.

Bernstein, L. R., Richards, V. M., and Green, D. M. (**1987**). ''The detection of spectral shape changes,'' in *Auditory Processing of Complex Sounds*, edited by W. Yost and C. Watson (Erlbaum, Hillsdale, NJ), pp. 6–15.

Buus, S. (**1985**). ''Release from masking caused by envelope fluctuations,'' J. Acoust. Soc. Am. **78**, 1958–1965.

Charbonneau, G. R. (**1981**). ''Timbre and the perceptual effects of three types of data reduction,'' Comput. Music J. **5**(2), 10–19.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Grey, J. M. (**1977**). ''Multidimensional perceptual scaling of musical timbres,'' J. Acoust. Soc. Am. **61**, 1270–1277.

Grey, J. M., and Gordon, J. W. (**1978**). ''Perceptual effects of spectral modification on musical timbres,'' J. Acoust. Soc. Am. **63**, 1493–1500.

Grey, J. M., and Moorer, J. A. (**1977**). ''Perceptual evaluations of synthesized musical instrument tones,'' J. Acoust. Soc. Am. **62**, 454–462.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). ''Detection in noise by spectrotemporal pattern analysis,'' J. Acoust. Soc. Am. **76**, 50–56.

Horner, A. (**2001**). ''A Simplified Wavetable Matching Method Using Combinatorial Basis Spectra Selection,'' J. Audio Eng. Soc. **49**, 1060–1066.

Iverson, P., and Krumhansl, C. L. (**1993**). ''Isolating the dynamic attributes of musical timbre,'' J. Acoust. Soc. Am. **94**, 2595–2603.

Kendall, R. A., and Carterette, E. C. (**1996**). ''Difference thresholds for timbre related to spectral centroid,'' in Proc. 4th Int. Conf. Music, Perception and Cognition (Montreal), Faculty of Music, McGill University, pp. 91–95.

Kidd, Jr., G., Mason, C. R., Uchanski, R. M., Brantley, M. A., and Shah, P. (**1991**). ''Evaluation of simple models of auditory profile analysis using random reference spectra,'' J. Acoust. Soc. Am. **90**, 1340–1354.

Krimphoff, J. (**1993**). ''Analyse acoustique et perception du timbre,'' unpublished DEA thesis, Université du Maine, Le Mans, France.

Krimphoff, J., McAdams, S., and Winsberg, S. (**1994**). ''Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique,'' J. Phys. **4**(C5), 625–628.

Krumhansl, C. L. (**1989**). ''Why is musical timbre so hard to understand?'' in *Structure and Perception of Electroacoustic Sounds and Music*, edited by S. Nielzen and O. Olsson (Excerpta Medica, Amsterdam), pp. 43–53.

Lakatos, S. (**2000**). ''A common perceptual space for harmonic and percussive timbres,'' Percept. Psychophys. **62**, 1426–1439.

Lentz, J. J., and Richards, V. M. (**1998**). ''The effects of amplitude perturbation and increasing numbers of components in profile analysis,'' J. Acoust. Soc. Am. **103**, 535–541.

McAdams, S., Beauchamp, J. W., and Meneguzzi, S. (**1999**). ''Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters,'' J. Acoust. Soc. Am. **105**, 882–897.

Mendoza, L., Schultz, M. L., and Schulz, R. A. (**1996**). ''Comodulation masking release as a function of masking noise-band temporal envelope similarity in normal hearing and cochlear impaired listeners,'' J. Acoust. Soc. Am. **99**, 2565.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). ''A model for the prediction of thresholds, loudness and partial loudness,'' J. Audio Eng. Soc. **45**, 224–240.

Plomp, R. (**1970**). ''Timbre as a multidimensional attribute of complex tones,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Eliden, The Netherlands), pp. 405–408.

Stevens, J. (**2002**). *Applied Multivariate Statistics for the Social Sciences*, 4th ed. (Erlbaum, Hillsdale, NJ), p. 264.

Toole, F. E., and Olive, S. E. (**1988**). ''The modification of timbre by resonances: perception and measurement,'' J. Audio Eng. Soc. **36**, 122–142.

Versfeld, N. J., and Houtsma, A. J. M. (**1991**). ''Perception of spectral changes in multi-tone complexes,'' Q. J. Exp. Psychol. A **43**, 459–479.

Watkins, A. J. (**1991**). ''Central auditory mechanisms of perceptual compensation for spectral-envelope distortion,'' J. Acoust. Soc. Am. **90**, 2942–2955.

Watkins, A. J., and Makin, S. J. (**1996**). ''Effects of spectral contrast on perceptual compensation for spectral-envelope distortion,'' J. Acoust. Soc. Am. **99**, 3749–3757.

Wessel, D. L. (**1979**). ''Timbre space as a musical control structure,'' Comput. Music J. **3**(2), 45–52.

Neuro-computational processing of moving sonar echoes classifies and localizes foliage

Roman Kuc
Intelligent Sensors Laboratory, Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520-8284

(Received 8 January 2004; revised 4 June 2004; accepted 16 June 2004)

Echoes from *in situ* tree trunks, similar to those observed by flying bats, are processed. A moving sonar converts echoes into spike sequences and applies neural-computational methods to classify objects and estimate passing range. Two classes of tree trunks act as retro-reflectors that generate strong echoes (SEs), identified by a locally dense spike pattern. Linear drive-by sonar trajectories cause SEs to follow hyperbolic curves specified by passing range. A glint is a collection of consecutive range readings matching expected values on a specific hyperbolic curve. Passing-range detectors compare successive SE data with expected values in a table and tally coincidences. Counters increment when coincidences occur and decrement when they do not. A glint terminates after tallying a sufficient number of coincidences and coincidence failure occurs in the maximum-count detector. Reflector roughness, deviations in sonar trajectory, and echo jitter necessitate a coincidence window to define matches. Short windows identify small glints over piecewise linear sonar trajectories, while long windows accommodate deviations in sonar speed and trajectory, and associate multiple glints observed with shorter windows. The minimum coincidence window size yielding glints classify smooth and rough retro-reflectors. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1779273]

PACS numbers: 43.80.Ev, 43.80.Ka, 43.80.Lb [WWU] Pages: 1811–1818

I. INTRODUCTION

Individual sonar echoes contain limited information about the environment, requiring multiple echoes to be integrated during sonar movement. Some sensing tasks inherently require motion, such as moving a finger along a surface to determine its texture. This paper examines ecological echoes similar to those observed by flying bats to classify foliage and estimate passing range. Rather than asking the standard question of what echoes a particular object produces, this paper examines the informational content of strong echoes, or those that are most easily detected in the presence of noise.

This paper extends previous work in sonar recognition of ecological objects in several ways. In Ref. 1 a custom biomimetic sonar consisting of a separate transmitter and a pair of receivers was situated on a robot arm. The sonar translated along a linear 1 m trajectory to acquire echoes from potted foliage located at meter range. A wide-band swept-frequency probing pulse was transmitted and echoes were applied to a bank of bandpass filters followed by statistical signal processing. While the results were encouraging, the separation in the echo feature space was marginal. The experiments described in this paper moved from the laboratory into the ''real world.'' Precise robot movements were replaced by a moving sonar platform that experienced anticipated deviations from a desired linear trajectory.

Previous experiments moved from the laboratory into natural settings to examine larger and more varied foliage at greater ranges.[2] A conventional Polaroid ranging module was employed to exploit its portability and robust detection at ranges up to 3.5 m, at the expense of narrow-band pulses,[3,4] which we interpret as the output of a single auditory channel in our previous study.[1] The conventional sonar normally detects only the first echo that exceeds a threshold and uses the corresponding time-of-flight (TOF) to measure target range, assumed along the sonar axis. Our biomimetic sonar transforms echoes into a neural-like spike sequence by repeatedly clearing and enabling threshold detection. The spike firing rate provides a measure of echo amplitude. Our initial experiments with this sonar employed a sector scan, with a stationary sonar rotated about the vertical axis, to acquire echoes from plants exhibiting different scattering properties.[2] Those results demonstrated that stable features could be extracted from spike sequences if the sonar was oriented to detect specularly reflected echoes. This paper considers scanning the environment with a sonar that is oriented at a constant direction while moving along a piecewise linear trajectory.

Our investigations of navigating robots in indoor settings employed this sonar.[5] Orienting the sonar obliquely (45°) with respect to the travel direction identified and localized right-angle corner retro-reflectors. This paper extends the approach to naturally occurring targets. Our previous work applied standard classification approaches using template matching to the entire echo sequence. The sonar described here employs neuro-morphic processing of spikes with delays, coincidence detection, and thresholds.

The paper has the following organization. The next section describes the spike sequences produced by echoes from simple and complex ecological objects. Section III describes the processing of the spike sequences to identify strong echo coincidences while the sonar is travelling. Section IV presents the processing results. Section V interprets the results and Sec. VI speculates on equivalent processes in bat sonar.

II. SONAR SPIKES

The echoes were acquired with a conventional Polaroid ranging sonar[6] mounted on an electric scooter moving along an intended linear trajectory at 0.22 m/s. Probing pulses containing 16 cycles of 50 kHz were emitted every $T_E = 0.15$ s. While the Polaroid transducer has a wide bandwidth,[1,7] this long pulse of single sinusoidal energy produces a narrow-band sonar that has robust detection properties. The conventional sonar processes echoes with rectification and low pass filtering, producing an echo TOF indication when the filter output first exceeds a threshold. Our sonar clears this TOF indication within 1 μs, thus forming a short output pulse, or spike, and resets the detector.[2,5] Another sonar spike occurs whenever the processed waveform again exceeds the threshold, at which time our sonar again resets the detector. The time between spikes relates to the echo amplitude because strong echoes cause the integrator output to ramp up to the threshold more quickly than weak echoes. Consequently, strong echoes produce three spikes exhibiting an interspike spacing as small as 140 μs, while weaker echoes produce less dense spike sequences. Echoes from range-distributed reflectors cause long spike sequences of varying density. This spike generation process illustrates a novel control of a conventional sonar and an inexpensive analog-to-digital conversion that provides sufficient information for target classification.

The sonar processes these spikes to extract only the first SE produced by each emission. While all echoes contain information, SEs are useful for characterizing objects because they are detected more reliably in the presence of reverberation and noise, and produce unbiased range readings.[4] The sonar identifies an SE when three sonar spikes occur within 300 μs. We convert echo times to equivalent range readings by multiplying by the sound speed 343 m/s and dividing by 2.[4]

The sonar was oriented approximately 45° in azimuth with respect to the travel direction to scan objects off to the left-hand side, as shown in Fig. 1(a). This allows the sonar to examine interesting targets while passing. Figure 1(b) shows the sonar located on a joystick-controlled electric scooter (Electric Mobility Rascal). The scooter was driven at 0.22 m/s along a desired straight path by maintaining the joystick in the forward position, although road conditions often caused the scooter to deviate slightly from a linear path. The sonar acquired echo data while driven past targets of interest, passing the target at the minimum range, or passing range (PR). Echoes occur when favorably oriented reflectors on the target surface scatter echoes back to the transducer. Retro-reflectors, such as tree trunks, are particularly interesting candidates for classification, producing echoes over a long section of sonar path. Two tree trunks having very different scattering properties illustrate the processing of observed echo spike sequences. A young red maple *Acer rubrum* shown in Fig. 1(c) had a 0.17 m diam trunk with a relatively smooth surface, while a large Norway maple *Acer platanoides* in Fig. 1(d) had a 0.8 m diam trunk with a rough surface compared to the wavelength of the sonar pulse (7 mm).

Sonar spike sequences and strong echoes observed during a drive-by scan of the red maple are shown in Fig. 2. Each emission produced from 1 to 6 sonar spikes. The set of spike sequences over successive emissions shows decreasing range as the sonar approaches the trunk. Weak echoes produce single and double spike sequences and occur when the trunk is either distant (emissions 5–6) or located on the trailing beam edge (emissions 70–81). SEs extracted from the spikes by the density criterion are shown as circles placed over the first spike in the triplet that generated it. SE range values decrease continuously as the main beam passes over the trunk in emissions 7–64. SEs from the trailing beam side lobe can be seen at emissions 67–70. Although the intended scooter trajectory was linear, uneven road conditions caused the scooter path to veer slightly toward the tree. Such environmentally induced path deviations may not be unusual for biosonars as well and the processing below will identify or accommodate such trajectory deviations.

The sonar spikes observed during a drive-by scan of a Norway maple are shown in Fig. 3. Many more spikes were generated by each emission because scatterers are distributed over range on the rough trunk surface. Even though the spikes are more complex than those for the red maple, the SEs show a clear trend of decreasing range as the sonar passes the Norway maple. The distant range readings come from the near-range section of the trunk. As the sonar passes the tree, SEs come from different parts along the surface of the large trunk.

III. SPIKE PROCESSING

To aid in the analysis of the SE sequence, we consider the simple model of range readings from an ideal vertical line target as the sonar travels along a linear trajectory at a specified passing range and speed. Figure 4 shows the expected range readings at a set of emission points following a hyperbola.[5] Echoes occur only when the target lies within the sonar beam. A small planar patch, such as a large leaf, would produce similar range readings, but over a smaller path length, when the planar patch surface normal lies within the sonar beam. In contrast, a retro-reflector does not have this constraint and produces echoes over a larger sonar path length. It is for this reason we choose to examine retro-reflectors.

A SE sequence containing a significant number of consecutive range readings having consistent hyperbolic delays defines a glint. Our sonar recognizes glints by counting coincidences between SE range readings and a table of expected values. Each row in the table corresponds to a passing-range detector and lists readings from a line target located at a specific passing range from 0 to 3 m, in steps of 0.025 m, for a total of 121 rows. The standard deviation of the error associated with this quantization is $0.025/\sqrt{12} = 0.007$ m.[8] The table contains expected range values computed at a dense set of points along the sonar linear trajectory to allow accurate registration with asynchronously acquired sonar data. Using the nominal sonar speed (0.22 m/s) and a finer sampling period $T_E/4 = 0.0375$ s, each row contains 405 values, corresponding to 8 mm sonar translations along a 3 m path that terminates when the sonar passes the target. An

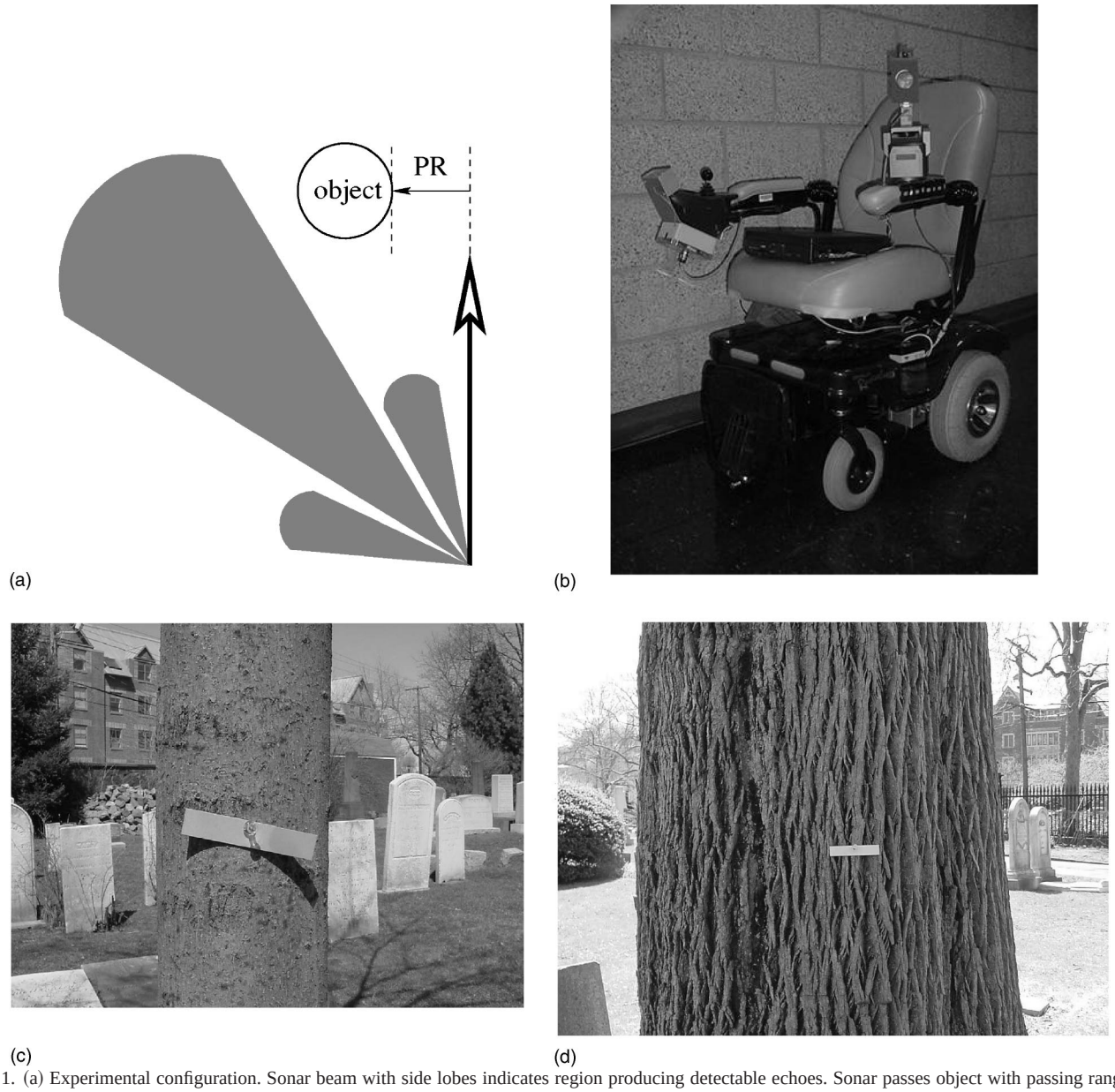


FIG. 1. (a) Experimental configuration. Sonar beam with side lobes indicates region producing detectable echoes. Sonar passes object with passing range PR at a height of 1.2 m. (b) Electric scooter with sonar mounted on left handle and laptop for data acquisition and processing. (c) Young red maple tree trunk (*Acer rubrum*, 0.17 m diam). (d) Norway maple tree trunk (*Acer platanoides*, 0.8 m diam). Bar width is 10 cm.

omnidirectional sonar beam is assumed, producing readings over the full sonar trajectory. This table then accommodates a wide-beam sonar, different sonar orientations or a sonar rotation to keep an object in the beam during the drive-by scan.

The matching procedure will be described by considering the first SE encountered with all PR detector coincidence counters equal to zero. This first SE range reading is compared with the expected range values in each PR detector to find a match within 4 mm, a threshold that guarantees a match will be found in each row corresponding to a valid passing range candidate. Each detector has a pointer that points to the range value expected from the next emission. The pointer is incremented by 4, because of the $T_E/4$ table construction, at each emission to synchronize actual and expected values. SE range values from successive emissions are compared to the predicted expected values and PR detector-specific counters increment when there is a match. Counters decrement when there is no match and minimum counter values are clamped to zero. Matches between data and expected values occur when an SE range reading falls within a coincidence window centered about the expected value. The coincidence window sizes that produced glints in this paper varied from 5 mm (29 μs), or ± 2.5 mm (± 15 μs) from the expected value, to 100 mm (583 μs). Other experiments with smooth poles exhibited glints with coincidence window sizes as small as 1 mm (6 μs). Small coincidence window sizes typically find small glints that accurately match the predicted values within echo jitter,[9] which usually occur over piecewise linear sections of the trajectory. Larger coincidence windows accommodate greater reflector size and surface roughness, as well as deviations from constant sonar velocity in both speed and direction.

Glint termination is determined using a stopping rule

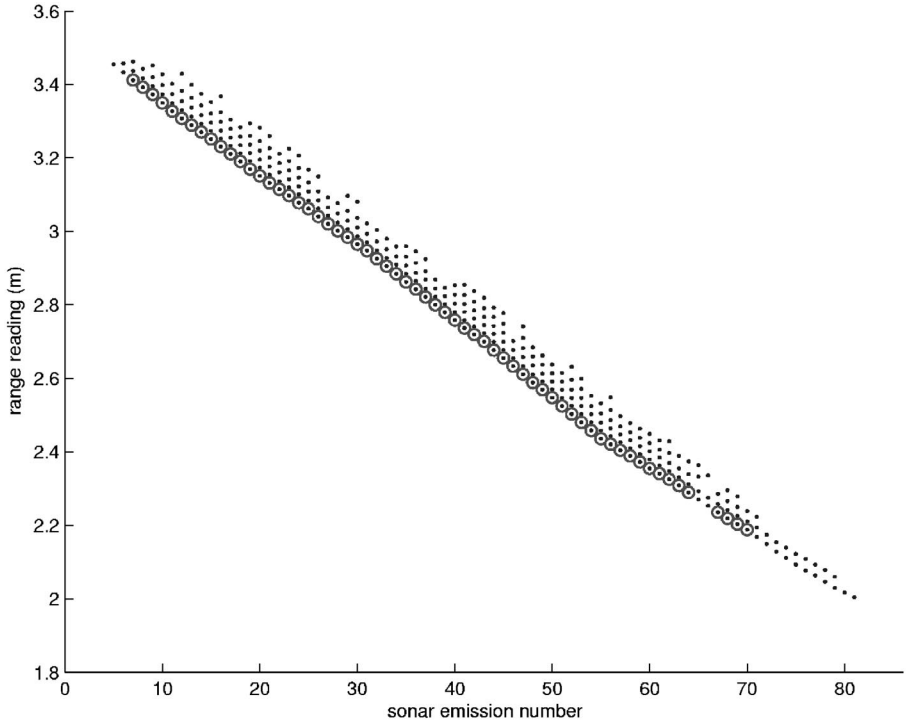


FIG. 2. Sonar range spikes (points) observed during by-drive scan of red maple. SEs are indicated by circles.

applied when at least one PR detector counter exceeds a threshold value that is proportional to the log of the coincidence window size. A glint is then terminated when the PR detector having the maximum count fails to find coincidence from an emission. Upon termination the maximum-count detector provides the count, indicating the glint size, and the passing range estimate (PRE). The counts in the 20 PR-detector neighbors, 10 on each side, of the maximum-count PR detector are processed to derive an equivalent standard deviation (SD) measure of spread by scaling the 21 counts values to form a probability mass function.[8] The maximum SD would equal 0.151 m if all 21 counters exhibited the same count. Glint termination also resets all counters to zero.

IV. RESULTS

The results for the shortest coincidence window size (5 mm) that produced glints for the red maple are shown in Fig. 5. These data are typical of similar small diameter trunks with smooth surfaces and plastic poles.[5] The Norway maple results using a 50 mm coincidence window are shown in Fig.

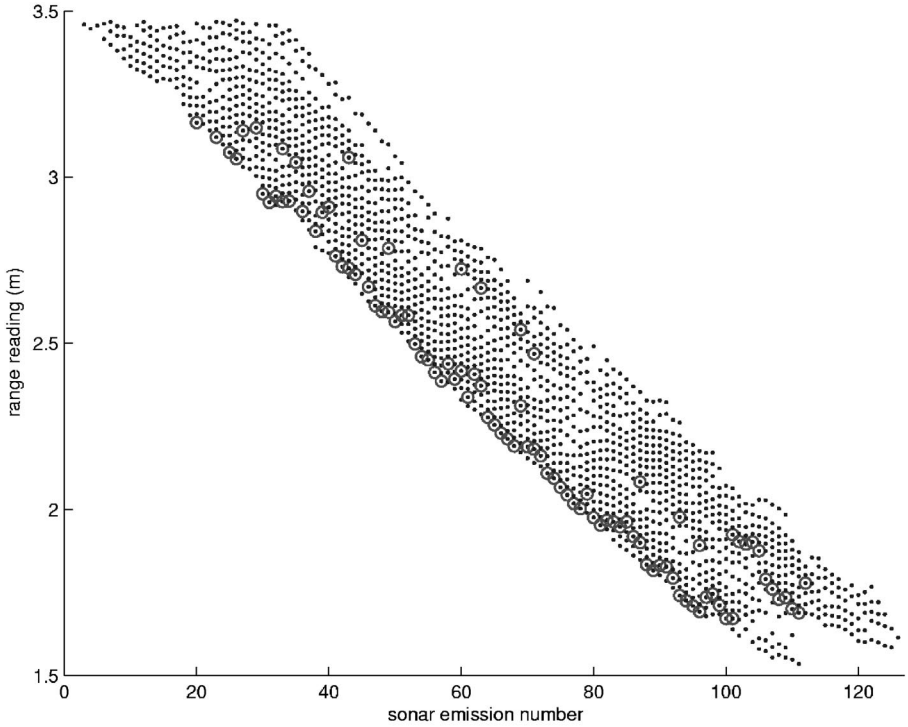


FIG. 3. Sonar range spikes (points) observed during by-drive scan of Norway maple. SEs are indicated by circles.

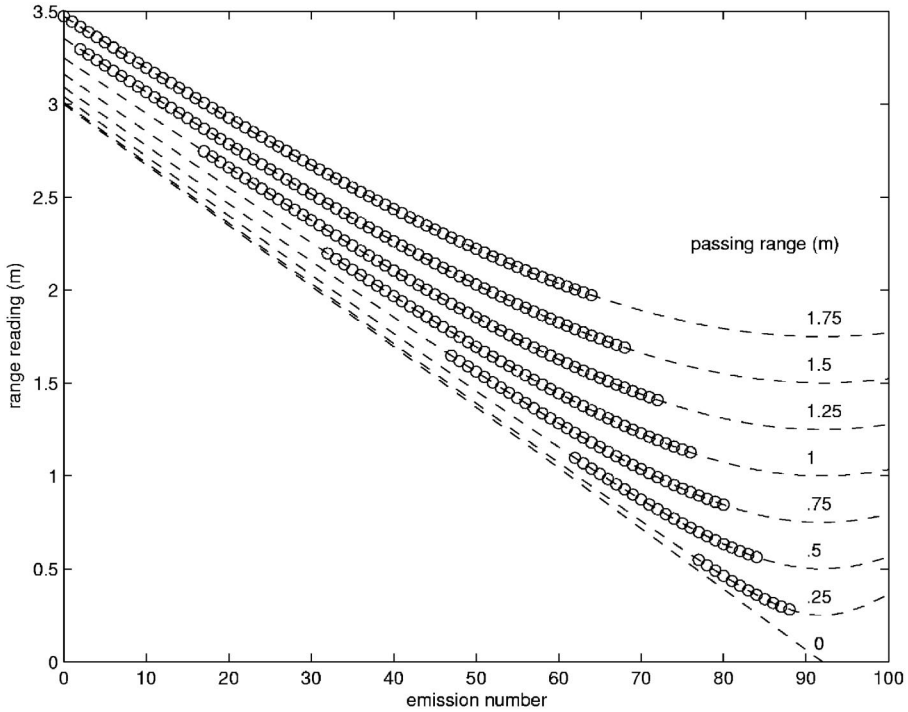


FIG. 4. Expected range readings from an ideal vertical line target observed during a drive-by scan along a linear trajectory at various passing ranges.

6. These data are typical of targets with range-distributed reflectors. All glints observed for different window sizes are listed in Table I.

The figures show the SE range readings that contribute to a glint, with dotted lines showing the beginning of the glint and dashed lines showing the last SE range in the glint. Predicted values coincident with the observed SE range reading are shown by points near their respective SE circles. These are tightly clustered for the 5 mm coincidence window and extend over a noticeable interval for the 50 mm window. Parts (b) and (c) in the figures show counts exhibited by all the PR detectors at glint terminations.

V. DATA INTERPRETATION

The red maple data in Fig. 5 yielded two glints. The PR detector counts at glint termination show only 3 or 4 detectors with nonzero counts and distinct PR detectors having maximum counts. Adjacent counters having significant counts indicates that the table resolution in expected values was adequate. These sharp maxima, quantified by their small SD values, indicate that the data were in very good agreement with the expected values. This is characteristic of a smooth surfaced retro-reflector or an isolated planar patch reflector. Although the trunk is not a line target, echoes originate from a narrow patch on the surface[5] that is adequately modelled by a line target. The two corresponding PREs are consistent with the experimental configuration. The second PRE is smaller, showing the drift of the scooter toward the tree during data acquisition.

The Norway maple data in Fig. 6 also yielded two glints, although at the larger 50 mm window. The second glint had an unexpected small SD value. We speculate that the 50 mm window extracted the expected hyperbolic range trend from a collection of scatterers whose echoes added constructively at the receiver, forming a pseudoplanar patch. The two PREs are approximately equal, with the second one being slightly smaller because the corresponding reflectors on the large-diameter trunk had a smaller passing range.

Table I lists the glints extracted with different window sizes for both targets. The minimum glint size threshold for each window size is given. Both glint size and SD tend to increase with coincidence window size. The larger windows accept more matches over successive emissions and cause more matches to occur in neighboring PR detectors. The smooth red maple bark produces glints at smaller coincidence windows than the rough Norway maple bark. The larger windows serve to associate the individual glints observed with smaller windows. For example, the 7–40 emission glint with the 50 mm coincidence window for the red maple combines the two glints observed with the 5 mm coincidence window. This example also shows that the PREs from combined glints average the individual-glint PREs. Small and large SD values occur for both targets, reducing the classification value of the SD. Large coincidence windows can cause multiple adjacent PR detectors to exhibit the same maximum count. In this case, our PRE comes from the smallest PR detector having the maximum count as a conservative estimate to reduce collision possibility. Such an estimate occurred for the red maple with the 100 mm window, indicated with an asterisk, and produces an SD value close to the maximum that would be observed for a uniform distribution of counts over 21 PR detectors (0.151 m).

It should be noted that PREs relate to glint-producing reflectors rather than the target as a whole. For the small isolated-reflector red maple, there is little difference. In contrast, the Norway maple has glint-producing collections of reflectors that occur at different passing ranges. Also, while applying glint models to the Norway maple data produce a minimum PRE$=1.725$ m, strong echoes not part of a glint occur at smaller ranges, as shown in Fig. 3. The interpreta-

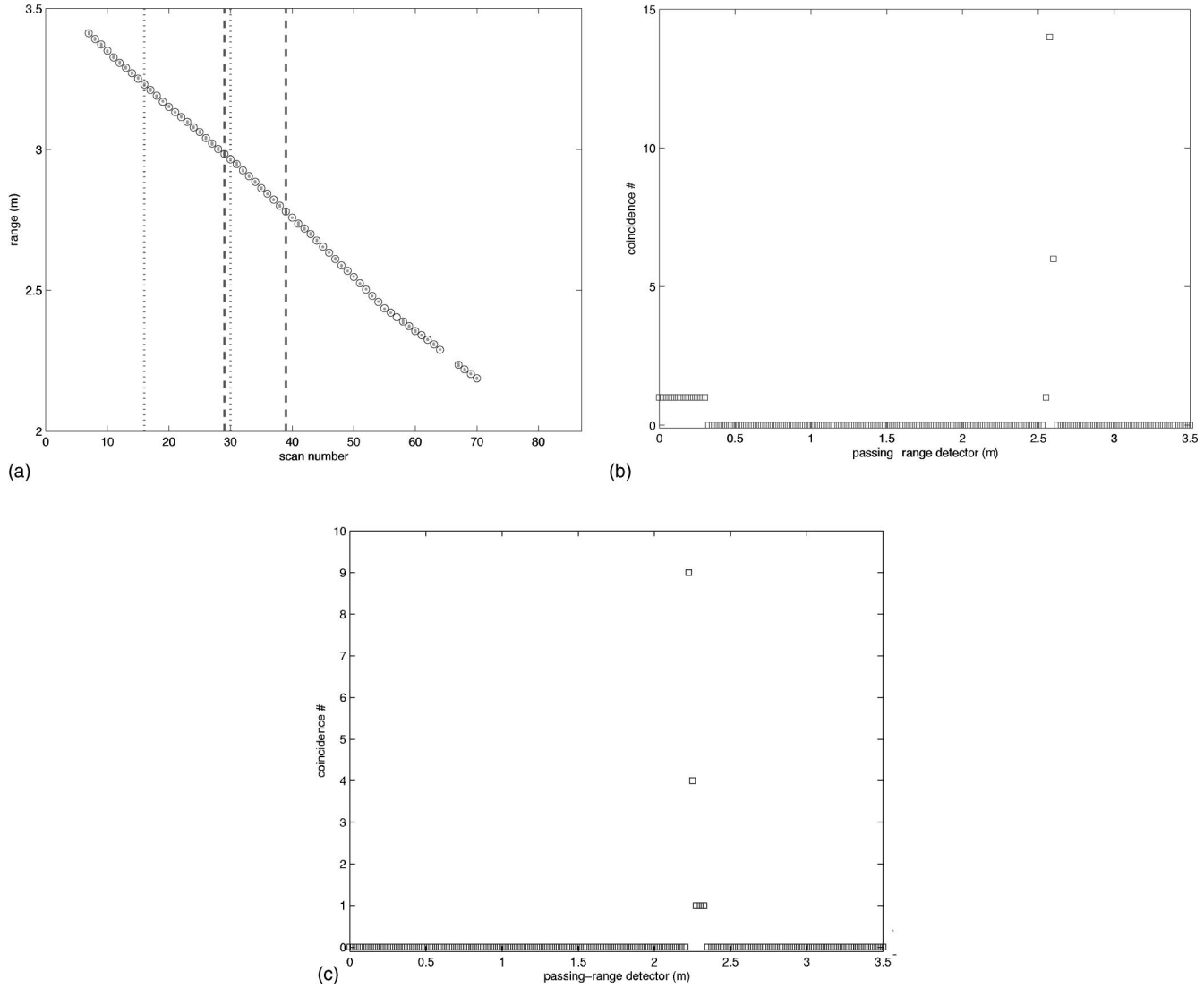


FIG. 5. Red maple results with 5 mm coincidence window. (a) Strong echoes (circles) and matching table values (points). Dotted lines show beginning and dashed lines ends of valid glints. (b) Distribution of detector counts at first glint termination between emission 15–28, size=14, SD=0.014 m, and PRE =2.575 m. (c) Distribution of detector counts at second glint termination between emissions 30–38, size=9, SD=0.036 m, and PRE=2.225 m.

tion of sonar spikes at smaller ranges is difficult because of the occlusion problem observed with wide sonar beams.[10] This indicates the main utilities of our glint model are for characterizing reflector smoothness and for associating glints, rather than estimating safe PRs.

VI. DISCUSSION

Processing of sonar echo data in this paper was performed off-line using Matlab's graphic capabilities to assist in data interpretation and display. A real-time glint extraction system was implemented on an IBM 380D laptop computer that processes echoes with multiple coincidence windows to mimic massively parallel feature extraction typical of neural systems. This real-time capability was useful for determining conditions under which glints are formed.

Closer examination of the Norway maple data in Fig. 3 shows that SEs are not always located on the initial sonar spikes, the ones that produce the conventional sonar TOF, and that our sonar looks beyond these initial weak echoes, something conventional sonars do not. A useful classification feature, to be examined in future experiments, is the placement of SE within the entire spike sequence. In a previous experiment[2] foliage having small leaves or needles typically produce weak leading echoes.

Sonar travel speed was considered to be known and constant, 0.22 m/s, in the spike interpretation. Deviations from this speed could produce anomalous results, similar to deviations from a linear trajectory. The results indicate that spikes produced valid glints under the experimental conditions that did not intentionally control for speed or trajectory. Both deviations would degrade glint quality, most noticeably at the shortest coincidence windows. Environmental conditions that cause deviations in speed and trajectory should be expected to degrade perceptual performance. The minimum coincidence window size producing glints would be one measure of such degradation.

Biological sonars could generate sonar spikes analogous to those shown here in two ways: either sparsely, with each emission yielding a single SE from a single reflector, the case here, or densely, with multiple SEs produced by each emis-

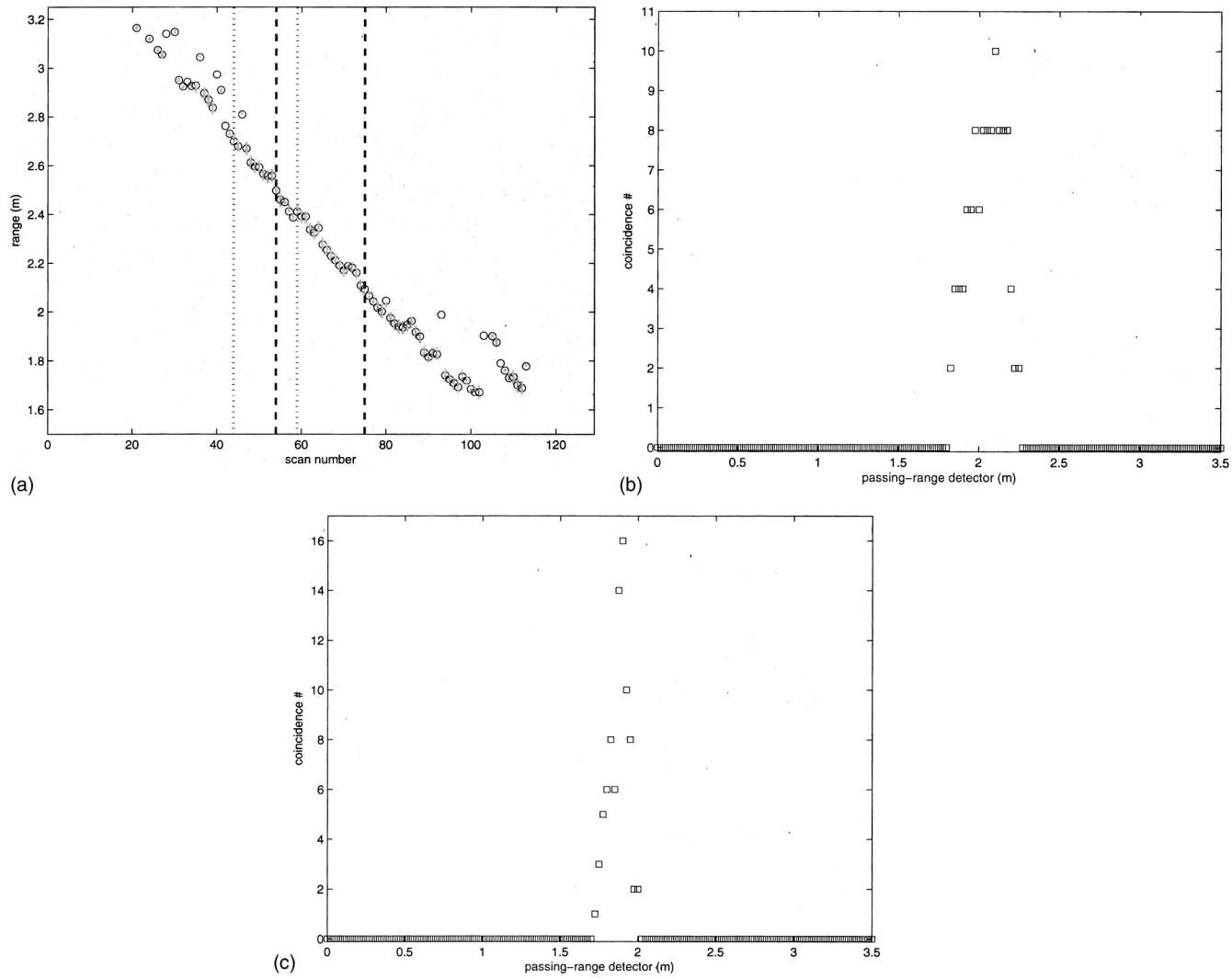


FIG. 6. Norway maple results with 50 mm coincidence window. (a) Strong echoes (circles) and matching table values (points). Dotted lines show beginning and dashed lines ends of valid glints. (b) Distribution of detector counts at first glint between emissions 45–54, size=10, SD=0.115 m, and PRE =2.100 m. (c) Distribution of detector counts at second glint between emissions 60–75, size=16, SD=0.067 m, and PRE=1.900 m.

sion. Dolphin sonars[11] use short pulses, suggesting sparse readings. Some bats emit swept-frequency pulses.[12] With echoes having sufficient signal-to-noise ratio, each cochlear bandpass filter can extract echo components having a specific frequency, equivalent to the single-tone sonar echoes here. Each filter output can then generate its own SE range reading at the corresponding time along the frequency sweep. This would then produce a sequence of range read-

TABLE I. Red maple and Norway maple glints for different window sizes. Coincidence window size, glint size threshold, glint emission interval, glint size, SD of counts over 20 neighboring Pr-detectors, and passing-range estimate (PRE).

Processing parameters		Red maple glints				Norway maple glints			
Window size (mm)	Glint threshold	Emissions	Size	SD (m)	PRE (m)	Emissions	Size	SD (m)	PRE (m)
5	7	15–28	14	0.014	2.575				
		30–38	9	0.036	2.225				
10	8	11–28	18	0.018	2.575				
		30–40	11	0.032	2.225				
		42–57	16	0.031	1.925				
20	8	17–32	26	0.032	2.550				
		34–50	17	0.046	2.100				
		53–64	12	0.042	1.975				
50	10	7–40	34	0.081	2.475	45–54	10	0.115	2.100
		43–64	22	0.083	1.925	60–75	16	0.067	1.900
100	11	7–50	44	0.112	2.375	42–73	32	0.079	2.000
		43–64	12	0.143*	1.725*	75–92	18	0.104	1.725

ings along a short travel path segment from a single swept-frequency echo, accommodating the piecewise linear model employed in this paper.

Our sonar used monaural processing to classify ipsilateral objects lying within the sonar beam. Binaural sonars[13,14] can localize objects in elevation and bearing from single echoes. Given two ears in biosonars, it is an open question as to whether it is a better strategy to process the data monaurally, as two separate sonars each looking to one side to extract glints, or binaurally to get localization from each strong echo. It is probably a combination of both operating simultaneously in a massively parallel system.

VII. SUMMARY

This paper examined echoes from foliage, similar to those observed by flying bats. A moving sonar converted echoes into spike sequences and applied neural-computational methods to extract glints, differentiate smooth from rough tree trunks, and estimate passing range. Retro-reflectors generated strong echoes, identified by their dense spike firing rate. Linear sonar trajectories past vertical line-like targets cause SE range values to follow hyperbolic curves specified by passing range. Successive SE range values were compared with expected values in a table and PR specific detectors tallied matches. A glint terminates after a sufficient number of coincidences is tallied followed by a mismatch by the maximum-count detector. The sonar accommodated rough reflectors, deviations from a linear trajectory, and echo jitter by employing a finite-duration coincidence window. Short windows identified individual glints from smooth surfaces detected along piecewise linear sonar trajectories, rough surfaces exhibited no short-window glints, while long windows generalized extended reflectors and associated smaller glints observed with shorter coincidence windows.

ACKNOWLEDGMENTS

We express our appreciation to Electric Mobility for donating the electric scooter used in this study for research in advanced mobility.

[1] R. Mueller and R. Kuc, ''Foliage echoes: A probe into the ecological acoustics of bat echolocation,'' J. Acoust. Soc. Am. **108**, 836–845 (2000).
[2] R. Kuc, ''Transforming echoes into pseudo-action potentials and classifying plants,'' J. Acoust. Soc. Am. **110**, 2198–2206 (2001).
[3] R. Kuc, ''Forward model for sonar maps produced with the Polaroid ranging module,'' IEEE Trans. Rob. Autom. **19**, 358–362 (2003).
[4] R. Kuc, ''Pseudo-amplitude sonar maps,'' IEEE Trans. Rob. Autom. **17**, 767–770 (2001).
[5] R. Kuc, ''Recognizing retro-reflectors with an obliquely-oriented multi-point sonar and acoustic flow,'' Int. J. Robot. Res. **22**, 129–145 (2003).
[6] J. Borenstein, H. R. Everett, and L. Feng, *Navigating Mobile Robots* (A. K. Peters, Wellesley, MA, 1996).
[7] L. Kleeman and R. Kuc, ''Mobile robot sonar for target localization and classification,'' Int. J. Robot. Res. **14**, 295–318 (1995).
[8] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 2nd ed. (McGraw-Hill, New York, 1984).
[9] A. M. Sabatini, ''A stochastic model of the time-of-flight noise in airborne sonar ranging systems,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 606–614 (1997).
[10] J. Budenske and M. Gini, ''Why is it so difficult for a robot to pass though a doorway using ultrasonic sensors?'' Proceedings of the IEEE International Conference on Robotics Automation, 1994, pp. 3124–3129.
[11] W. W. L. Au, *The Sonar of Dolphins* (Springer-Verlag, New York, 1993).
[12] J. A. Simmons, M. B. Fenton, and M. J. O'Farrell, ''Echolocation and pursuit of prey by bats,'' Science **203**, 16–21 (1979).
[13] R. Kuc, ''Sensorimotor model of bat echolocation and prey capture,'' J. Acoust. Soc. Am. **96**, 1965–1978 (1994).
[14] R. Kuc, ''Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion, and surface texture classification,'' IEEE Trans. Biomed. Eng. **49**, 1173–1180 (2002).

Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues

Xiufen Gong,[a] Dong Zhang, Jiehui Liu, Huanlei Wang, Yongsheng Yan, and Xiaochen Xu

State Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 10 February 2004; revised 16 June 2004; accepted 22 June 2004)

Three novel methods for acoustic nonlinearity parameter B/A imaging in reflection mode are developed in this paper. They are: (1) the acoustic nonlinearity parameter B/A tomography by detecting reflective second harmonic wave, (2) the B/A tomography in reflection mode via the measurement of the difference frequency wave generated by a parametric array, and (3) the C-scan imaging of B/A via the measurement of the echo second-harmonic signal. A theoretical analysis and the experimental imaging of normal and pathological biological tissues by using these methods are also present and discussed. Results show that using the acoustic nonlinearity parameter imaging we can more easily distinguish the diseased tissue from the normal one than using the linear acoustic parameters. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1781709]

PACS numbers: 43.80.Ev, 43.80.Cs [FD] Pages: 1819–1825

I. INTRODUCTION

The ultrasound B-scan imaging has been widely used in a modern hospital for clinical ultrasound diagnosis because it can provide important information on the diseased state of the tissues in a human body noninvasively and nondestructively. The principle of this technique is based on the assumption of linear acoustics. However, many reports indicate that the nonlinear acoustical effects must be considered in the range of biomedical frequencies and intensities.[1–3] Due to these nonlinear effects, a monochromatic sound wave will be distorted progressively and harmonics will be generated. Considering that the second harmonics has a better directional performance and spatial resolution compared with the fundamental wave, the most current B-scanning has the function of the second-harmonic imaging.[4–6] It is well known in nonlinear acoustics that as a measure of nonlinearity of a medium, the acoustic nonlinearity parameter B/A determines the pressure amplitude of the second harmonics generated. Also, since this parameter reflects the dynamic properties of the medium, it can give more important messages about the structural features and the pathological change of the tissues than the linear ultrasonic parameters.[7–11] Therefore, the nonlinearity parameter B/A can become a new and useful parameter in biological tissue characterization.

Owing to the inhomogeneous features of biological tissues, most efforts have worked on the B/A imaging in the fields of ultrasound biomedicine and nonlinear acoustics. Ichida *et al.*[12] presented a phase shift method, and Nakagawa *et al.*[13] reported methods using the second-harmonic wave and parametric array for nonlinearity parameter imaging. Cain[14] theoretically studied the ultrasonic imaging of the nonlinearity parameter B/A in reflection mode. Fatemi and Greenleaf proposed a method for read-time assessment of tissue nonlinearity.[15] In our laboratory, several B/A imaging techniques in transmission mode for biological tissues based on the insert-substitution method have been developed,[16–18] e.g., B/A tomography via the measurement of the second-harmonic detection, and B/A tomography via the measurement of the sum or difference frequency wave produced by the parametric effect. The obtained experimental images showed that the difference between healthy and pathological tissues could be determined quite well by using the acoustic nonlinearity parameter.

Up to now, few works on B/A image using echo signals[19] have been reported. Considering that the ultrasound medical diagnostic instruments operate mostly in reflection mode, it is necessary to investigate the B/A imaging in reflection mode. Besides, the B/A imaging via the measurement of the second harmonics is limited by high sound attenuation. Hence, development of novel techniques for reducing the frequency and obtaining lower sound attenuation in the reconstruction of B/A tomography is needed.

In this paper we present three versions of B/A imaging methods in reflection mode, which are: (1) the acoustic nonlinearity parameter B/A tomography by detecting the reflective second-harmonic wave, (2) B/A tomography in reflection mode via the measurement of the low difference frequency wave generated by a parametric array, and (3) the nonlinearity parameter B/A imaging in reflection mode at a constant depth (named C-scan imaging of B/A) via the measurement of the echo signals of second harmonics. A theoretical analysis and the experimental images of normal and pathological biological tissues obtained by using these methods are presented and discussed.

II. METHODS AND RESULTS

A. Tomography of the nonlinearity parameter B/A via the second harmonics in reflection mode

1. Theory

A compound PZT transducer (T), as shown in Fig. 1, is used to transmit the primary wave at frequency ω and also to receive the reflective second-harmonic wave at $\omega_2 = 2\omega$. An

[a]Electronic mail: dzhang@nju.edu.cn

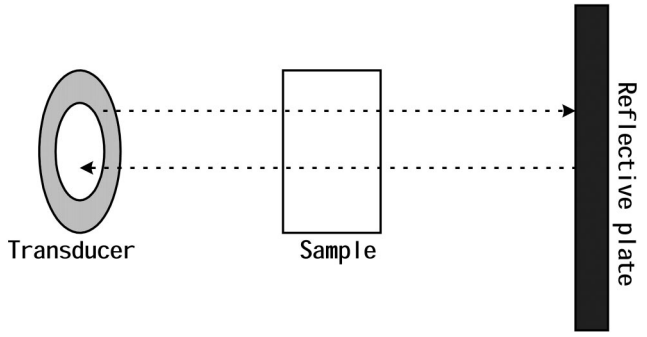


FIG. 1. Propagation path of generated second harmonics in reflection mode.

aluminum plate is fixed at position L from the transducer to enhance the reflective signal. The test sample is inserted into the water (as a reference medium) at distance x from T with $x<L$. When a wave with the finite amplitude $P_1(0)$ is transmitted from the transducer at $x=0$, $P_1=P_1(0)\sin\omega t$, the wave will be distorted along the path of propagation and the second harmonics generated.

The pressure amplitudes of the reflected second harmonics received by the compound transducer consist of two parts, one of which is the reflected second harmonics generated on the forward path and the other is the second harmonics generated on the backward path of the reflected primary wave from the plate and can be expressed as

$$P_{2s}(L)=R\frac{\omega_2}{4}P_1^2(0)\left[\int_0^L\beta_i(x)dx+R\int_0^L\beta_i(L-x)dx\right], \tag{1}$$

where $\beta_i(x)=\beta(x)/\rho_0(x)c_0^3(x)$, $\beta(x)=1+[(B/2A)(x)]$. $\rho_0(x)$, $c_0(x)$ and $(B/A)(x)$ are the density, velocity, and the nonlinearity parameter at x. R is the reflection coefficient of the plate. Considering the sound attenuation of the sample, Eq. (1) is rewritten as

$$\begin{aligned}p_{2s}(L)=R\frac{\omega_2}{4}P_1^2(0)\int_0^L\beta_i(x)\Bigg\{&\exp\left[\int_0^x-2\alpha_1(x)dx-\int_x^L\alpha_2(x)dx-\int_0^L\alpha_2(x)dx\right]\\&+R\exp\left[\int_0^{L-x}-2\alpha_1(x)dx-\int_{L-x}^L\alpha_2(x)dx-\int_0^L2\alpha_1(x)dx\right]\Bigg\}dx,\end{aligned} \tag{2}$$

where $\alpha_1(x)$ and $\alpha_2(x)$ are the attenuation coefficients of the primary and second-harmonic waves.

In order to avoid the measurement of the absolute sound pressure amplitude for the second harmonics, the finite-amplitude insert-substitute method[16,20] is used. When the sample is absent, the pressure amplitude of the second harmonics in the water reflected from the plate is

$$P_{2w}(L)=\frac{\omega L}{2}P_1^2(0)R(1+R)\beta_w, \tag{3}$$

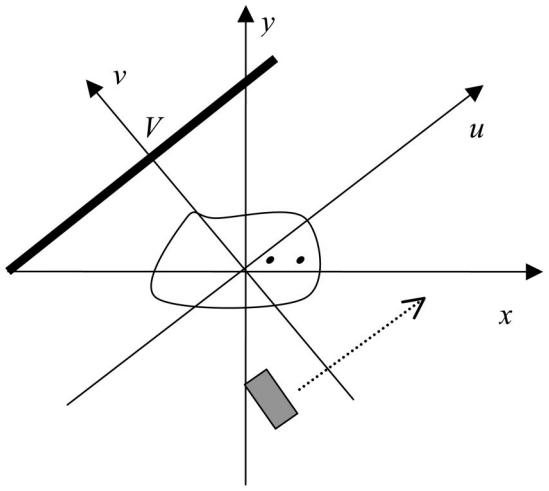


FIG. 2. Geometry of ultrasonic tomography in reflection mode.

where β_w is the β_i of distilled water, whose $B/A=5.2$. The ratio of Eq. (2) to Eq. (3) will be

$$\begin{aligned}p=\frac{p_{2s}(L)}{p_{2w}(L)}=\frac{1}{(1+R)\beta_wL}\int_0^L\beta_i(x)&\\\times\Bigg\{\exp\left[\int_0^x-2\alpha_1(x)dx-\int_x^L\alpha_2(x)dx\right.&\\\left.-\int_0^L\alpha_2(x)dx\right]+R\exp\left[\int_0^{L-x}-2\alpha_1(x)dx\right.&\\\left.-\int_{L-x}^L\alpha_2(x)dx-\int_0^L2\alpha_1(x)dx\right]\Bigg\}dx.\end{aligned} \tag{4}$$

Figure 2 shows a conventional CT scanning system. The sample rotates from 0° to 360° at an equal interval angle $\Delta\theta=2\pi/K$ (K is the number of rotating steps), and at each angle $\theta_j=j\times2\pi/K$ ($j=0,1,...,K-1$) the compound transducer sweeps step by step along one fixed direction (u axis). Assuming that the linear scan is $\Delta u=L/N$ (L and N are the distance and the step number of the linear scan respectively), the projection data are sampled at $u_i=i\times\Delta u$ ($i=-N/2$, $-N/2+1,...,0,1,...,N/2$). Thus, a $K\times(N+1)$ projection data matrix will be obtained for each specimen.

The ratio of $P_{2s}(L)/P_{2w}(L)$ can be regarded as the projection data $p(u,\theta)$ in the CT technique

$$p(u,\theta)=\frac{1}{\beta_0L}\int_{-V}^V\beta_i'(x,y)dv, \tag{5}$$

where $V=L/2$, $\beta_i'(x,y)=\beta_i(x,y)C(x,y)$, $C(x,y)$ is the matrix of the attenuation correction and determined by Eq. (5)[21]

$$\begin{aligned}C(x,y)=\frac{1}{2\pi}\int_0^{2\pi}\Bigg\{&\exp\left[\int_{-V}^v-2\alpha_1(x,y)dv-\int_v^V\alpha_2(x,y)dv\right]+R\exp\left[\int_{-V}^{V/2-v}-2\alpha_1(x,y)dv\right.\\&\left.-\int_{V/2-v}^V\alpha_2(x,y)dv\right]\Bigg\}d\theta.\end{aligned}$$

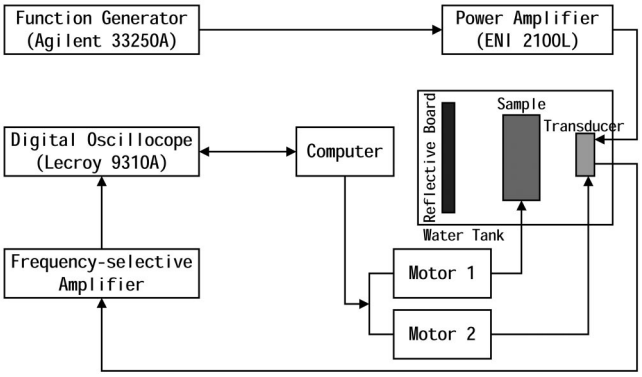


FIG. 3. Block diagram of experimental system for B/A tomography via second-harmonic wave in reflection mode.

Equation (4) shows that $p(u,\theta)$ is a linear integral of $\beta_i'(x,y)$ along the propagation path of the wave, therefore the $\beta_i'(x,y)$ image can be reconstructed by using the filtered convolution method in the CT technique. Then, the nonlinearity parameter B/A tomography can be obtained after the correction of sound attenuation.

2. Experimental system and results

The block diagram of the experimental system is shown in Fig. 3. The outer annular ring of the compound transducer with a resonance frequency of 2.0 MHz is used as the transmitter and the central disk with a resonance frequency of 4.0 MHz as the receiver to detect the second harmonics. A 2.0 MHz tone-burst signal is generated by a programmable function generator (Agilent 33250A, USA). After passing through a broad-band power amplifier (ENI 2100L, USA), this signal is sent to the transmitter. The second-harmonic signals of 4 MHz cumulated in the path of propagation through the sample will be reflected by the aluminum plate and then detected by the receiver. By using a frequency-selective amplifier with a central frequency of 4 MHz, the reflected second-harmonic wave is amplified and recorded by a digital oscilloscope (Lecroy 9310A, USA). The amplitude of the reflected second harmonics in the reference medium (water) P_{2w} can also be measured. Then the value of $P_{2s}(L)/P_{2w}(L)$ is used as the projection data. The two-dimensional mechanical scanning system (motor 1 for sample rotating and motor 2 for linear scanning) is controlled via the GP-IB interface by a computer.

A cylindrical sample container with diameter of 30 mm is surrounded by a polythene membrane. The excised fresh soft tissues taken from a slaughter house are carefully packed into the container and the unoccupied space is filled with a 0.9% saline solution to avoid the influence of gas bubbles. The cross section of a three-layer sample model is shown in Fig. 4(a). The internal layer is the porcine liver tissue, the intermediate layer the porcine fatty tissue and the external layer water. The corresponding experimental B/A tomography is shown in Fig. 4(b). The second sample model consists of two half-cylindrical tissues with normal liver tissue and hepatonecrosis tissue as shown in Fig. 5(a). The corresponding experimental tomography of B/A is shown in Fig. 5(b). The averaged values of B/A for the above-mentioned tissues

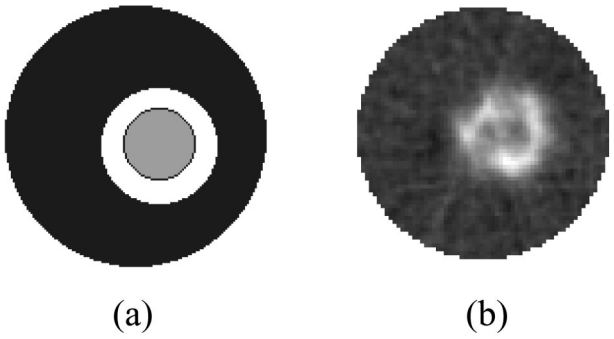


FIG. 4. B/A tomography via second-harmonic wave in reflection mode for a three layers tissue sample: (a) sample model, (b) experimental tomography.

can be determined by the gray scale of their images as follows: for the fatty tissue $B/A=10.9$, for the normal porcine liver $B/A=7.3$, and for the hepatonecrosis liver $B/A=10.5$. These obtained values coincide well with the results in literatures.[7,18] Furthermore, the hepatonecrosis liver can be discerned according as its B/A value is higher than that of the normal liver, therefore, they can be distinguished more easily by means of B/A tomography.

B. Tomography of the acoustic nonlinearity parameter *B/A* via the difference frequency wave from a parametric array in reflection mode

1. Theory and method

When two primary waves with angular frequencies ω_1 and ω_2 and the initial sound pressure amplitudes $P_1(0)$ and $P_2(0)$ are transmitted from a transducer simultaneously, several frequency components will be generated during the sound propagation due to the nonlinear interaction of these two waves in a medium. If only the quadratic effects are considered, the second harmonics $2\omega_1$ and $2\omega_2$ as well as the parametric signals with sum and difference frequencies components $\omega_1 \pm \omega_2$ will be generated.[22]

As shown in Fig. 6, an aluminum plate with rhe reflective coefficient R used as a reflector is located at distance L away from a compound transducer T. Considering only the difference frequency component P_d, the reflected sound pressure amplitude of the difference frequency wave received by the compound transducer is

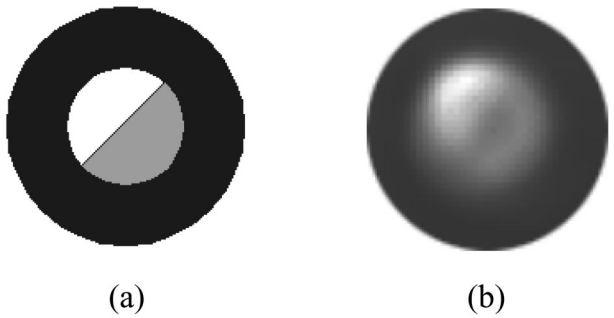


FIG. 5. B/A tomography via second-harmonic wave in reflection mode for normal and hepatonecrosis livers: (a) sample model, (b) experimental tomography.

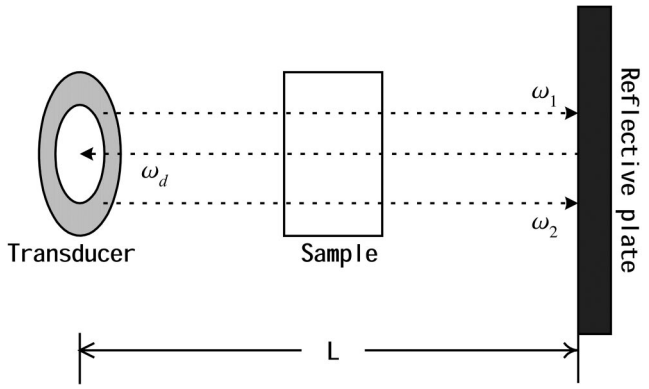


FIG. 6. Propagation path of generated difference-frequency wave via a parametric array in reflection mode.

$$P_{dw}=R(1+R)\frac{\omega_d}{2}P_1(0)P_2(0)\beta_w L, \tag{6}$$

where $\beta_w=[2+(B/A)_w]/2\rho_w c_w^3$, $\omega_d=\omega_2-\omega_1$, ρ_w, c_w are the density and velocity of the reference medium (distilled water), respectively.

For an inhomogeneous medium (biological tissue) the nonlinear coefficient $\beta_i(x)$ is dependent on the position. Then Eq. (6) will be rewritten as

$$P_{ds}=R(1+R)\frac{\omega_d}{2}P_1(0)P_2(0)\int_0^L \beta_i(x)dx. \tag{7}$$

Considering the sound attenuation of the sample, this equation is modified as

$$P_{ds}=R\frac{\omega_d}{2}P_1(0)P_2(0)\int_0^L \beta_i(x,y)C(x,y)dl, \tag{8}$$

where $C(x,y)$ is the attenuation matrix defined as

$$\begin{aligned}C(x,y)=\frac{1}{2\pi}\int_0^{2\pi}\Bigg\{&\exp\Bigg[-\int_0^L[\alpha_1(x,y)+\alpha_2(x,y)]dl\\&-\int_l^L\alpha_d(x,y)dl-\int_0^L\alpha_d(x,y)dl\Bigg]\\&+R\exp\Bigg[\int_0^L-[\alpha_1(x,y)+\alpha_2(x,y)]dl\\&-\int_0^l\alpha_d(x,y)dl\Bigg]\Bigg\}d\theta.\end{aligned} \tag{9}$$

In order to avoid measuring the absolute value of the sound pressure, the finite amplitude insert-substitution method is introduced and the distilled water is used as the reference medium. Using Eqs. (6) and (8), we define $p=P_{ds}/P_{dw}$ as the projection data in the conventional CT technique, the procedure of which is similar to that shown in Fig. 2.

The two-dimensional image of $\beta_i(x,y)C(x,y)$ is reconstructed by using the filtered convolution algorithm to deal with the projection data. Then the $\beta_i(x,y)$ tomography is obtained by correction of $C(x,y)$. Finally the B/A tomography is imaged with algebraic transformation.

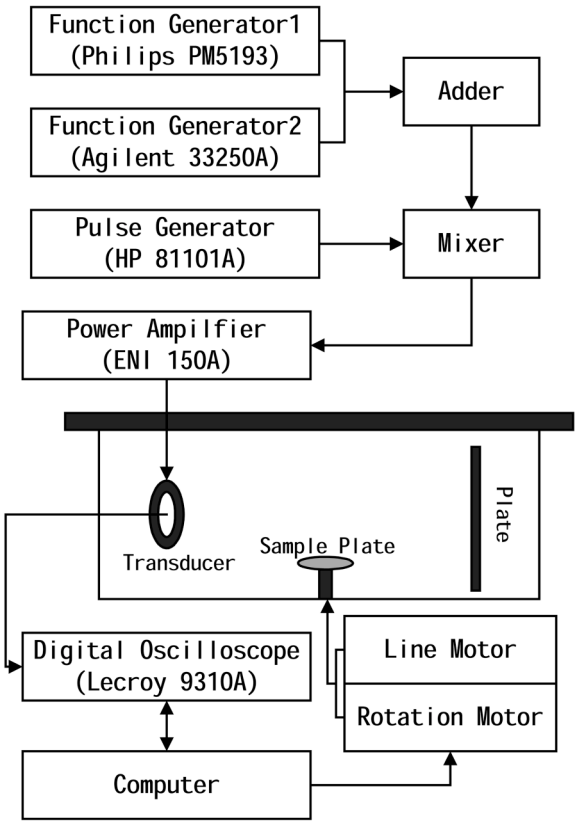


FIG. 7. Block diagram of experimental system for B/A tomography via a parametric array in reflection mode.

2. Experimental arrangement

The block diagram of the experimental system is shown in Fig. 7. A compound transducer is employed as both the transmitter and the receiver. The external circular PZT ring with an outer diameter of 2.2 cm and with a resonance frequency of 4 MHz is used as a transmitter of two primary waves. The internal PZT disk with a diameter of 0.7 cm and with a resonance frequency of 1 MHz is used as a receiver of the difference frequency wave. Two function generators (Philips PM5193 and Agilent 33250A) transmit sine waves at 3.5 and 4.5 MHz, respectively. These two signals are added by an adder, then modulated by a pulse generator (HP8110A, USA). The modulated burst signal is amplified by a power amplifier (ENI A150, USA) and then excites the external PZT of the transducer. The reflected signal of difference frequency is received by the internal PZT disk and recorded by a digital oscilloscope (Lecroy 9310, USA). The amplitudes of the difference frequency wave at 1 MHz are measured with and without inserting the sample (P_{ds} and P_{dw}). The ratio of P_{ds}/P_{dw} is considered as the projection data. In the mechanical scanning, the rotating interval angle is 15°, so 24 angles are needed in one circle. At each angle, the projection data is sampled at the step of 1.25 mm. All projection data are collected in the range of 8 cm and a 24×65 projection matrix is obtained. The B/A tomography is displayed after reconstruction and correction by attenuation $C(x,y)$.

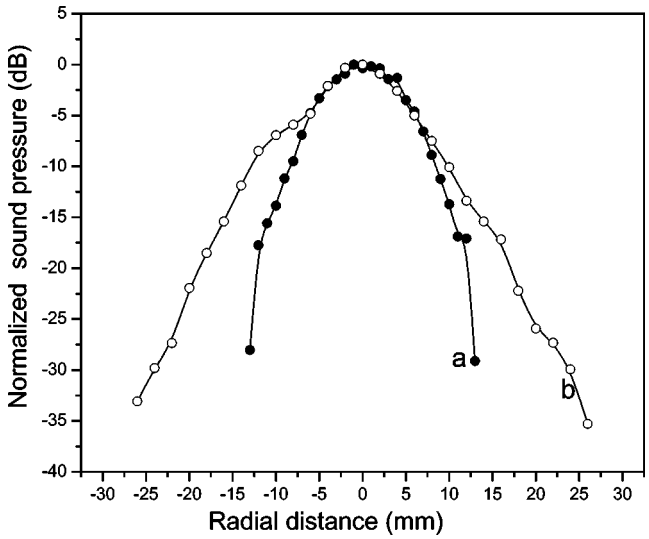


FIG. 8. Comparison of beam patterns at 1 MHz between generation directly from a transducer and that from a parametric array.

3. Experimental results

a. The beam pattern of the difference frequency wave. The beam pattern of the difference frequency wave at 1 MHz for a distance of 8 cm generated by the nonlinear interaction of the two primary waves at 3.5 and 4.5 MHz is shown in Fig. 8 as closed circles. For comparison, the beam pattern at the same distance from the transmitter excited directly at 1 MHz is also shown in this figure as open circles. It can be found that the difference frequency wave has a better directional performance, which will be conducive to improve the resolution of imaging.

b. B/A tomography for biological tissues. B/A tomography was performed for several biological tissues including normal and diseased ones. As shown in Figs. 9(a) and 10(a), the two sample models in different tissue combinations are studied. In Fig. 9(a) the porcine fatty tissue is located in the external ring and the internal part is the normal porcine liver tissue. In Fig. 10(a), the left side is the normal porcine liver tissue and the right side is a kind of pathological porcine liver tissue (hepatocirrhosis). The corresponding experimental images are shown in Figs. 9(b) and 10(b), respectively.

The obtained experimental images coincide well with their sample models and the gray levels represent the B/A values of biological tissues. Based on the gray levels of the

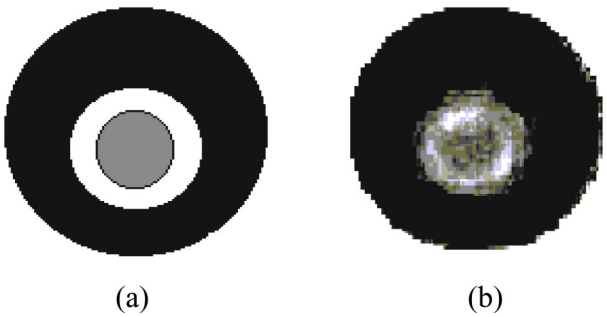


FIG. 9. B/A tomography via parametric array in reflection mode for two kinds of normal tissues, (a) sample model (normal liver and fatty tissues), (b) experimental tomography.

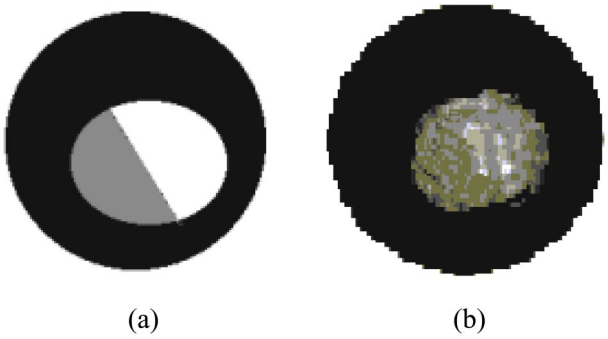


FIG. 10. B/A tomography via parametric array in reflection mode for normal and diseased tissues: (a) sample model (normal and hepatocirrhosis livers), (b) experimental tomography.

images, the averaged values of B/A are obtained as follows: $B/A=11$ for the porcine fat; $B/A=6.5$ for the normal porcine liver and $B/A=8.3$ for the hepatocirrhosis liver. These results are consistent with those presented in Refs. 7 and 18.

The second-harmonic imaging has been successful for years due to its better lateral resolution and improved clarity of the image quality. However, the second harmonics has higher attenuation because of the doubling of the fundamental frequency. The difference frequency scheme is needed because it provides the low frequency wave (corresponding to the low sound attenuation) with a high directionality.

C. Nonlinearity parameter imaging in C-scan mode

1. Theory and method

The nonlinearity parameter C-scan image is constructed by the echo signal of the reflected second harmonics at a constant depth of a biological specimen. Assume that a finite amplitude plane wave $P_1(0)$ with frequency f_0 is emitted at $z=0$ and travels through the medium. The biological specimen can be regarded as a layered medium distributed in the uniform media. At distance z from the source, the reflected echo signal of the second harmonics $2f_0$ is

$$P_{12r}(2f_0,z)=P_{12}(2f_0,z)R(2f_0,z)\exp\left[-\int \alpha_2(2f_0,z)dz\right], \tag{10}$$

where $R(2f_0,z)$ is the reflection coefficient at z, $\alpha_2(2f_0,z)$ is the sound attenuation at $2f_0$, and $P_{12}(2f_0,z)$ is the pressure amplitude of the cumulated second harmonics at distance z and should satisfy

$$\frac{dP_{12}(2f_0,z)}{dz}=K(z)P_1^2(f_0,0)\exp\left[-\int 2\alpha_1(f_0,z)dz\right] - \alpha_2(2f_0,z)P_{12}(2f_0,z), \tag{11}$$

where $K(z)=\omega/4\rho_0c_0^3[B/A(z)+2]$.

If an infinitesimal ultrasound wave $P_i(0)$ with frequency $2f_0$ is transmitted into the medium, the reflected signal of this fundamental wave at z will be

$$P_{ir}(2f_0,z)=P_i(2f_0,z)R(2f_0,z)\exp\left[-\int \alpha_2(2f_0,z)dz\right], \tag{12}$$

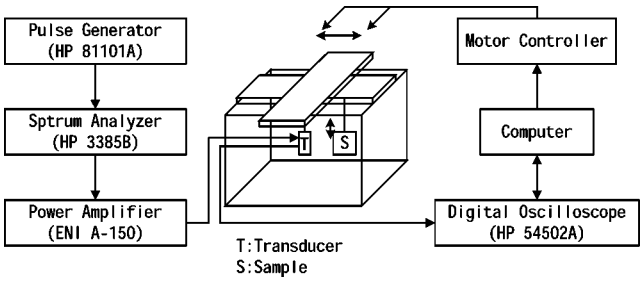


FIG. 11. Block diagram of experimental system for B/A imaging by C-scan method.

where $P_i(2f_0,z)=P_i(2f_0,0)\exp[-\int \alpha_2(2f_0,z)dz]$. Dividing Eq. (10) by Eq. (12) and eliminating $R(2f_0,z)$ and $\exp[-\int \alpha_2(2f_0,z)dz]$, we get

$$\frac{P_i(2f_0,z)}{P_{12}(2f_0,z)}=\frac{P_{ir}(2f_0,z)}{P_{12r}(2f_0,z)}. \tag{13}$$

Differentiating Eq. (13) with respect to variable z and considering the relationship of Eq. (11), we obtain

$$\frac{d}{dz}\left[\frac{P_{12}(2f_0,z)}{P_i(2f_0,z)}\right]=K(z)\frac{P_1^2(f_0,z)}{P_i(2f_0,0)}\exp\left[-\int [2\alpha_1(f_0,z) - \alpha_2(2f_0,z)]dz\right]. \tag{14}$$

For biological tissues, $2\alpha_1(f_0,z)=\alpha_2(2f_0,z)$, then Eq. (14) will be reduced to

$$\frac{d}{dz}\left[\frac{P_{12r}(2f_0,z)}{P_{ir}(2f_0,z)}\right]=K(z)\frac{P_1^2(f_0,0)}{P_i(2f_0,0)}. \tag{15}$$

Parameter z can be converted to t by $z=c_0t/2$. Hence, the distribution of the nonlinearity parameter B/A is obtained as

$$\frac{B}{A}(z)=\frac{8\rho c^2}{\omega}\frac{P_i(0)}{P_1^2(0)}\frac{d}{dt}\left[\frac{P_{12r}(2f_0,z)}{P_{ir}(2f_0,z)}\right]-2. \tag{16}$$

From this equation, we can image the B/A distribution of the biological specimen by measuring the ratio of $P_{12r}(2f_0,z)$ and $P_{ir}(2f_0,z)$.

2. Experiment and results

Figure 11 shows an automatic mechanical system for the B/A C-scan imaging in pulse reflection mode with high precision, which consists of a Newport MM3000 motion controller, a pulse generator, a frequency synthesizer PM5193, a wideband power amplifier ENI A150, a digital oscilloscope HP54502A and a compound PZT transducer used both as a transmitter (for 2 MHz) and a receiver (for 4 MHz). The 3D mechanical scanning system is used to construct the image at a constant depth of the tissue. The reflected signals are detected and measured by the following procedures: (1) passing the signal through the delay switch of the digital oscilloscope and adjusting the imaging depth of the sample, (2) segmenting the wave form, taking FFT and obtaining the pressure amplitude of the reflected second harmonics $P_{12r}(2f_0,z)$ and the reflected fundamental component $P_{ir}(2f_0,z)$ respectively, (3) calculating the ratio P_{12r}/P_{ir}, the slope $(d/dt)[P_{12r}/P_{ir}]$ and the value of $(B/A)(z)$, (4) writing the B/A value into the image file. Then the 2D image of $(B/A)(x,y)$, perpendicular to the z axis, is obtained by grating the C-scan method.

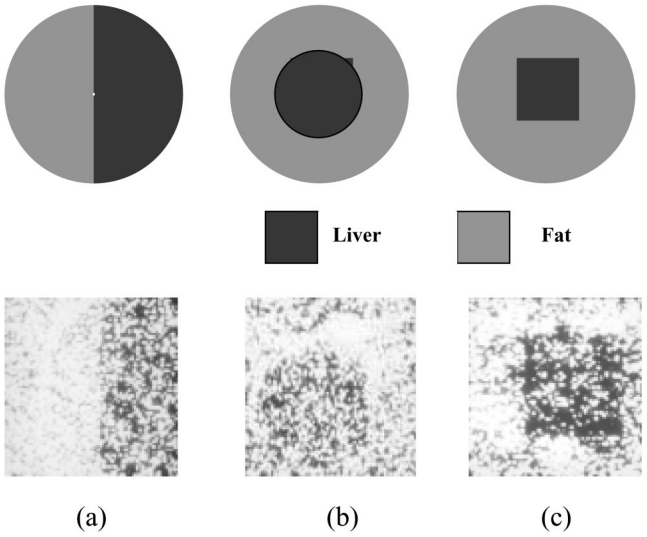


FIG. 12. Sample models and experimental results in B/A imaging by C-scan mode: (a) sample with half liver and half fat tissues, (b) sample with liver in inner circle and fat in outer ring, (c) sample with liver in central square and fat in outside.

Biological samples in the different combinations of normal porcine liver and fat tissues are studied. Figure 12 shows the sample models and corresponding experimental results. In sample (a), liver and fat tissues are located in the right and left half parts, respectively. With a sampling interval of 0.4 mm and a sampling number of 64×64, the imaging area is 2.56×2.56 cm^2. In sample (b), the central square is a liver tissue with a border of 2.0 cm and the outer part is a fat tissue. In sample (c), the ring area is a liver tissue and the other area is a fat tissue. The inner-radius and outer-radius of the ring are 0.6 cm and 1.1 cm, respectively. For samples (b) and (c), the sampling interval is 0.5 mm and the sampling number is 64×64, so the imaging area is 3.2×3.2 cm^2. In experimental images, black and gray areas correspond to the porcine liver and fat tissues, the B/A values of which are 6.9 and 9.6, respectively. The obtained B/A values coincide well with those obtained by other imaging methods. Figure 13(a) demonstrates a sample model with a half normal porcine liver and a half diseased fatty liver. Figure 13(b) shows the

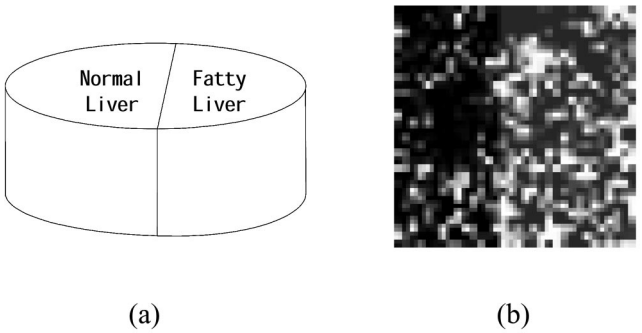


FIG. 13. B/A imaging by C-scan mode for a sample with half normal porcine liver and half diseased fatty liver: (a) sample model (b) experimental result.

TABLE I. Linear and nonlinear parameters of several biological tissues.

Sample	Density (g/cm^3)	Velocity (m/s)	Attenuation (2 MHz, dB/cm)	B/A	B/A (Ref.)
Porcine fat	0.95	1455	1.11	10.9	10.8[a]
Porcine liver	1.06	1588	1.08	6.9	6.8[b]
Fatty liver	1.05	1598	1.20	8.3	8.3[b]
Hepatocirrhosis	1.05	1600	1.26	8.3	8.1[b]
Hepatonecrosis	1.05	1597	1.17	10.5	8.3[b]

[a]See Ref. 16.
[b]Reference 23.

corresponding two dimensional B/A image in C-scan mode. From Fig. 13(b), it is clearly seen that the nonlinearity parameter B/A of the diseased fatty liver ($B/A=8.3$) is significantly higher than that of the normal liver ($B/A=6.8$), thus it is easier to discriminate them from each other in the B/A image according to different gray scales.

III. DISCUSSION AND CONCLUSION

The B/A images obtained in reflection mode demonstrate good clarity and improved contrast between the pathological and healthy states of soft animal tissues. The values of acoustic linear parameters (including densities, sound velocities, attenuations) and nonlinearity parameters are compared and listed in Table I. We found that the acoustic nonlinearity parameter is much more sensitive to the pathological state of the porcine liver tissue than the linear parameters.

For the developed B/A imaging methods in this paper, the energy in the second harmonic frequency is higher than that in the wave at the difference frequency generated by a parametric array. However, the second harmonics has higher sound attenuation because of the doubling of the fundamental frequency. The difference frequency sound wave at 1 MHz provides much lower sound attenuation and has better directional performance compared with that excited directly from a transmitter at 1 MHz, and therefore could improve the imaging resolution. However, it should be noted that the proposed three methods for imaging the B/A parameter are all based on the assumption of the plane wave insonification. Although use of the insert substitution method could reduce the diffraction effect,[16] further work on the diffraction effect is necessary.

In summary, ultrasound identification of the healthy tissues from the diseased ones is sometimes difficult due to the fact that even when tissues are pathologically different, their ultrasonic properties, such as sound velocity, attenuation, and acoustic impedance, may still be quite similar. However, their values of nonlinearity parameter are significantly different and can be used to enhance the discrimination and evaluate the healthy and diseased tissues effectively. The results obtained in the present paper indicate the potential application of nonlinearity parameter B/A imaging in ultrasound diagnosis and bring about a step forward in medical ultrasound.

ACKNOWLEDGMENTS

The Authors wish to thank Professor F. Dunn for his valuable suggestions and comments. This work was supported by Natural Science Foundation of China and JiangSu Province.

[1] T. G. Muir and E. L. Carstensen, ''Prediction of nonlinear acoustic effects at biomedical frequencies and intensities,'' Ultrasound Med. Biol. **4**, 345–357 (1980).
[2] E. L. Carstensen, W. K. Law, N. D. Mckay, and T. G. Muir, ''Demonstration of nonlinear acoustical effects at biological frequencies and intensities,'' Ultrasound Med. Biol. **6**, 359–368 (1982).
[3] F. A. Duck, ''Nonlinear acoustics in diagnostic ultrasound,'' Ultrasound Med. Biol. **28**, 1–18 (2002).
[4] F. Tranquart, N. Grenier, V. Eder, and L. Pourcelot, ''Clinical use of ultrasound tissue harmonic imaging,'' Ultrasound Med. Biol. **6**, 889–894 (1999).
[5] I. Puls, D. Berg, M. Maurer, M. Schliesser, G. Hetzel, and G. Becker, ''Transcranial sonography of the brain parenchyma: comparison of B-mode imaging and tissue harmonic imaging,'' Ultrasound Med. Biol. **26**, 189–194 (2000).
[6] S. Tanaka, O. Oshikawa, T. Sasaki, T. Ioka, and H. Tsukuma, ''Evaluation of tissue harmonic imaging for the diagnosis of focal liver lesions,'' Ultrasound Med. Biol. **26**, 183–187 (2000).
[7] C. M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf, ''Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers,'' Ultrasound Med. Biol. **12**, 865–874 (1986).
[8] J. Zhang, M. S. Kuhlenschmidt, and F. R. Dunn, ''Influences of structural factors of biological media on the acoustic nonlinearity parameter B/A,'' J. Acoust. Soc. Am. **99**, 2397–2402 (1991).
[9] J. Zhang and F. Dunn, ''In vivo B/A determination in a mammalian organ,'' J. Acoust. Soc. Am. **81**, 1635–1637 (1987).
[10] I. Solodov, K. Pfleiderer, H. Gerhard, and G. Busse, ''Nonlinear acoustic approach to material characterization of polymers and composites in tensile tests,'' Ultrasonics **42**, 1011–1015 (2004).
[11] J. Banchet and J. D. N. Cheeke, ''Measurement of the acoustic nonlinearity parameter B/A in solvents: Dependence on chain length and sound velocity,'' J. Acoust. Soc. Am. **108**, 2754–2758 (2000).
[12] N. Ichida, T. Sato, and M. Linzer, ''Imaging the nonlinear parameter of a medium,'' Ultrason. Imaging **5**, 295–299 (1983).
[13] Y. Nakagawa, W. Hou, A. Cai, N. Arnold, and G. Wade, ''Nonlinear parameter imaging with finite amplitude sound waves,'' *IEEE 1986 Ultrasonic Symposium* (IEEE, New York, 1986), pp. 901–904.
[14] C. A. Cain, ''Ultrasonic reflection mode imaging of nonlinear parameter B/A: Theoretical basis,'' J. Acoust. Soc. Am. **80**, 28–32 (1986).
[15] M. Fatemi and J. F. Greenleaf, ''Real-time assessment of the parameter of nonlinearity in tissue using nonlinear shadowing,'' Ultrasound Med. Biol. **22**, 1215–1228 (1996).
[16] X. F. Gong, Z. M. Zhu, T. Shi, and J. H. Huang, ''Determination of nonlinearity parameter for biological tissues using FAIS and ITD methods,'' J. Acoust. Soc. Am. **86**, 1–5 (1989).
[17] D. Zhang, X. F. Gong, and S. G. Ye, ''Acoustic nonlinearity tomography for biological specimens via measurements of the second harmonic wave,'' J. Acoust. Soc. Am. **99**, 2397–2402 (1996).
[18] D. Zhang, X. F. Gong, and X. Chen, ''Experimental imaging of acoustic nonlinearity parameter B/A for biological tissues via a parametric array,'' Ultrasound Med. Biol. **27**, 1359–1365 (2001).
[19] I. Akiyama, ''Nonlinearity parameter B/A imaging by using ultrasound echo signals,'' Proceedings of 12th ISNA, 1990, pp. 379–384.
[20] X. F. Gong, R. Feng, C. Y. Zhu, and T. Shi, ''Ultrasonic investigation of the nonlinearity parameter B/A in biological media,'' J. Acoust. Soc. Am. **76**, 949–950 (1984).
[21] L. T. Chang, ''A method for attenuation correction in radionuclide computed tomography,'' IEEE Trans. Nucl. Sci. **Ns-25**, 28–32 (1978).
[22] P. J. Westervelt, ''Parametric acoustic array,'' J. Acoust. Soc. Am. **35**, 535–537 (1963).
[23] D. Zhang and X. F. Gong, ''Experimental investigation of the acoustic nonlinearity parameter tomography for excised pathological biological tissues,'' Ultrasound Med. Biol. **25**, 593–599 (1999).

Estimated transmission beam pattern of clicks recorded from free-ranging white-beaked dolphins (*Lagenorhynchus albirostris*)

Marianne H. Rasmussen[a)]
Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Magnus Wahlberg
Department of Zoophysiology, Aarhus University, CF Møllers Alle Building 131, DK-8000 Aarhus C, Denmark

Lee A. Miller
Institute of Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 29 February 2004; revised 17 May 2004; accepted 2 June 2004)

Recordings were made from white-beaked dolphins in Icelandic waters using a four-hydrophone array in a star configuration. The acoustic signals were amplified and sampled to a hard disk at a rate of 800 kHz per channel. The 3 and 10 dB beamwidths were calculated to be 8° and 10°, respectively, indicating a narrower transmission beam for white-beaked dolphins than that reported for bottlenose dolphins (*Tursiops truncatus*). The beamwidth was more similar to that found for belugas (*Delphinapterus lucas*). The measured beam pattern included large side lobes, perhaps due to the inclusion of off-axis clicks, even after applying several criteria to select only on-axis clicks. The directivity index was calculated to be 18 dB when using all data for angles from 0°–50°. The calculated sound radiation from a circular piston with a radius of 6 cm driven by a white-beaked dolphin click had a beam pattern very similar to the measured beam pattern for the main transmission lobe of the white-beaked dolphin. The directivity index was 29 dB. This is the first attempt to estimate the directionality index of dolphins in the field. [Work supported by the Oticon Foundation and the Danish National Research Council.] © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1775274]

PACS numbers: 43.80.Ka, 43.30.Vh [WWA] Pages: 1826–1831

I. INTRODUCTION

Odontocetes use echolocation for prey detection and orientation. Echolocation is performed by emitting ultrasonic clicks in a narrow beam and listening for the returning echoes (Au, 1993). It is important to know the width of the transmitted beam to evaluate several aspects of the sonar system, such as the angular resolution and the transmitted power. The transmission beam can give information about the space ensonified by the dolphin for hunting or navigating. Clicks have been recorded in at least 21 odontocete species, but the beam pattern has only been estimated for eight species (Au, 1993; Au *et al.*, 1999; Møhl *et al.*, 2003). Most of these recordings have been performed from animals in captivity; there are very few field measurements. Field studies are important as the echolocation behavior of the animals may differ significantly between wild and captive situations (Au, 1993). For example the frequency content of the signal can change and, therefore, also the transmission beam pattern. In addition, not all species of odontocetes can be kept in captivity, emphasizing the need to develop recording techniques for evaluating echolocation abilities of free-ranging animals (see Madsen, Kerr and Payne, 2004).

The beam pattern is described in detail from captive bottlenose dolphins, *Tursiops truncatus* (Evans *et al.*, 1973; Au, 1980; Au *et al.*, 1978, 1986), the beluga whale, *Delphinapterus leucas* (Au *et al.*, 1987), the false killer whale, *Pseudorca crassidens* (Au *et al.*, 1995) and the harbor porpoise, *Phocoena phocoena* (Au *et al.*, 1999). Only one field study is known to us where the directionality of odontocete signals was estimated (Møhl *et al.*, 2003, sperm whales, *Physeter macrocephalus*).

To determine the beam pattern, recordings on the acoustic axis of the whale must be compared with off-axis recordings. Therefore, both the distance to the whale and the orientation of the animal must be known. In field studies, clicks recorded on an array of hydrophones can be used for calculating the distance to the whale. The orientation must be known to within a few degrees, since the transmission beam pattern of odontocetes is from 6.5° to 16.5° wide (Au, 1993; Au *et al.*, 1999). Various approaches can be used to obtain the orientation. The whale can be tracked from successive acoustic locations obtained from a click sequence. This assumes that the acoustic axis of the animal is identical to its swimming direction. This is the case for the odontocetes investigated so far in the horizontal plane, whereas the vertical beam pattern is tilted by about 5° in relation to the swimming direction for the bottlenose dolphin and the beluga (Au, 1993). Other approaches, such as comparing the sound level at the different hydrophones in the array, or comparing the frequency content of the received signals have been used in previous studies (Au and Benoit-Bird, 2003; Au and Herzing

a)Electronic mail: mhr@biology.sdu.dk

 © 2004 Acoustical Society of America

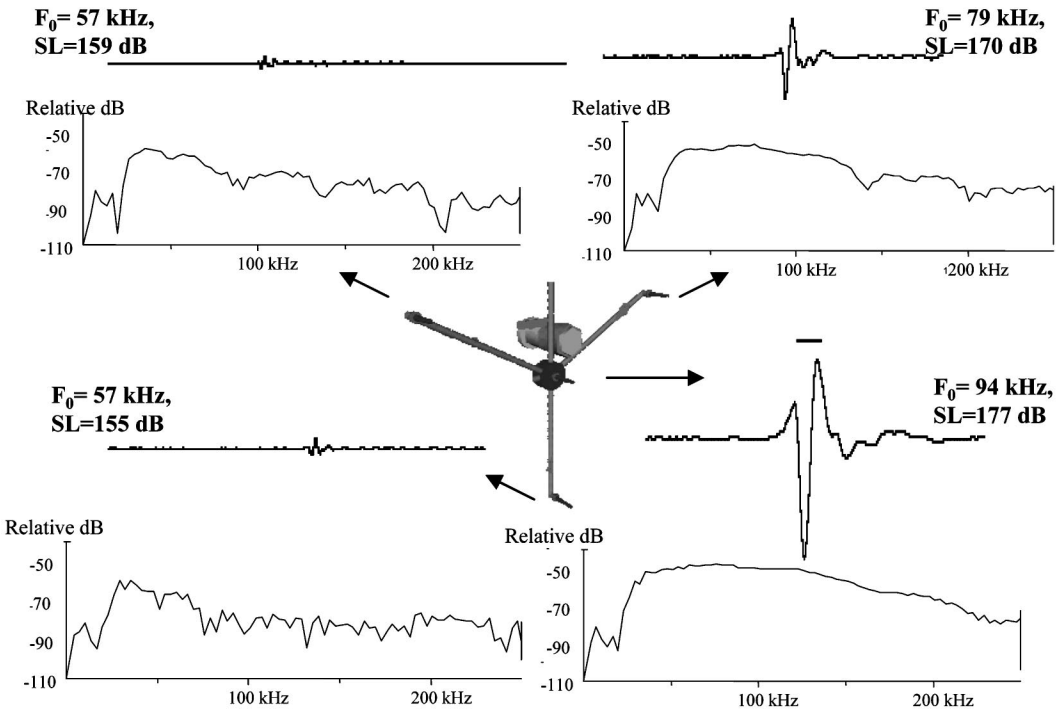


FIG. 1. An example of one on-axis click recorded on the center hydrophone. The waveform and amplitude spectra are shown for the center and peripheral hydrophones. In addition, the centroid frequency and source levels *re* 1 μPa (p–p) are shown for each channel. Note the high-frequency component (about 200 kHz) at the center hydrophone, a horizontal time line above the click indicates 10 μs.

2003; Møhl *et al.*, 2003; Rasmussen *et al.*, 2002). All these criteria (acoustic tracking, sound level, and spectral content) have their limitations and all can include off-axis clicks in the analysis.

The aim of this study was to investigate whether it is possible to use field recordings to describe the beam pattern of clicks from white-beaked dolphins, *Lagenorhynchus albirostris*. White-beaked dolphin clicks resemble those of bottlenose dolphins. The source levels of the clicks are between 189–219 dB *re* 1 μPa (peak-to-peak, p–p) (Rasmussen *et al.*, 2002). This is similar to what is found from other free-ranging dolphins (Au and Herzing, 2003; Au and Benoit-Bird, 2003). The broadband white-beaked dolphin clicks have maximum frequency components around 120 kHz (Rasmussen and Miller, 2002), similar to the clicks of bottlenose dolphins. However, the white-beaked dolphin clicks have secondary frequency components between 200–250 kHz, which can be 6–10 dB less than the maximum spectral components for on-axis clicks (Rasmussen and Miller, 2002, and Fig. 1 here). It is not known whether these secondary spectral components serve any function for the animal or if they are artifacts of the sound production system. Here we report that the main lobe of the estimated transmission beam of white-beaked dolphin clicks is narrower than that reported for bottlenose dolphins, possibly reflecting the broad bandwidth of white-beaked dolphin clicks.

II. MATERIAL AND METHODS

Clicks were recorded from white-beaked dolphins off the coast from Keflavik, Iceland (64°00.49′N, 22°33.37′W). Here groups of white-beaked dolphins gather in summer to forage and mate (Rasmussen, 1999; Rasmussen and Jakobsen, 2003). Recordings were made during the summers of 2002 and 2003 using a star-shaped hydrophone array similar to the one used by Rasmussen *et al.* (2002); Au and Herzing (2003); and Schotten *et al.* (2004). The array consisted of four Reson TC4034 hydrophones (frequency range 1 Hz–350 kHz±3 dB) and connected to a multichannel amplifier (Etec, Copenhagen, Denmark) on-board a small motor boat. The acoustic centers of the three outer hydrophones of the array were either 0.5 or 1 m from that of the center hydrophone. An underwater video camera was mounted approximately 10 cm above the center hydrophone (Fig. 1).

A recording session started when the dolphins were close to the boat. After idling the engine, the hydrophone array was lowed about two meters below the surface. Video/audio recordings were started and digitized on-line to a separate hard disk on a lunch-box computer using a video/audio card (Dazzle Inc.). A custom made click detector was connected to the center hydrophone for audio monitoring and recording on one of the audio channels of the video card. When dolphins were seen on a monitor or heard through the click detector, we started recording dolphin clicks from all four hydrophones simultaneously on line to two mirrored hard disks of the computer at a rate of 800 ksamples/s per channel (AD-Link, NuDAQ 20 MHz 4-channels, 12 bit) using a custom-written software program. For synchronizing video/audio signals and click recordings, a short beep sound was recorded on the second audio channel of the video card.

The data files were later separated into four files, one for each hydrophone, using a custom-made software program (''Sigpro,'' Simon Boel Pedersen, Copenhagen, Denmark) to check if clicks were recorded on all four hydrophones. Cool Edit Pro (version 2, Syntrillium Software) was used for ini-

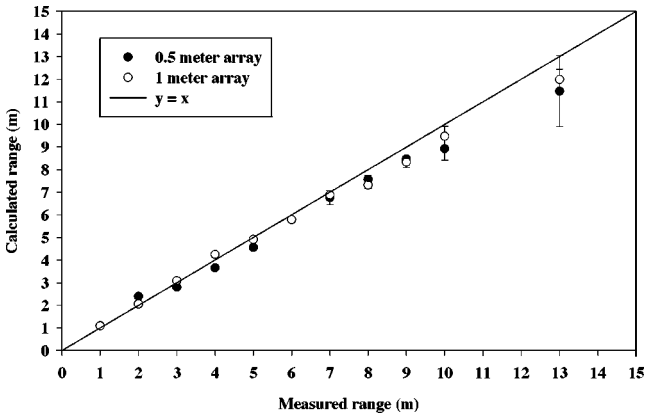


FIG. 2. A calibration of the hydrophone array for measuring distance. The black dots show the calculated distance using the 0.5 m array and the white dots show the calculated distance using the 1 m array (the line $y=x$ is shown for comparison). At 10 m there is an error of about −1 m in range for the 0.5 m array and about a −0.5 m in range for the 1 m array.

tial screening of the files. Files consisting of multiple animals clicking at the same time were excluded from further analyses. A program written in Matlab (6.1) was used for the calculations described below. The positions of the dolphins were calculated from the time of arrival differences between the center hydrophone and the peripheral hydrophones using same equations as those given in Au and Herzing (2003). Also, the angle between the line from the dolphin to each peripheral hydrophone and the line from the dolphin to the center hydrophone was calculated. The peak-to-peak (p–p) received sound pressure level of the click was compared with the values from a 250 Hz calibration tone at +162 dB *re* 1 μPa rms (Brüel & Kjær 4223 fitted with a custom-built adaptor for the Reson hydrophones). Source levels were then calculated, compensating for spherical spreading and sound absorption. The sound levels of clicks were estimated at various angles relative the dolphin's acoustic axis using the notation of the apparent source level (Møhl *et al.*, 2000), and the directivity index of the clicks was estimated from an interpolated beam pattern. In addition, the centroid frequency (Au, 1993) of each click was calculated. Figure 1 shows an example of one on-axis click with the corresponding waveform and amplitude spectra on each of the four hydrophones. The centroid frequencies and source levels *re* 1 μPa (p–p) @ 1 m for each click are also shown in Fig. 1.

The localization system was calibrated in open water using an underwater loudspeaker (Reson TC2130) connected to a Yamaha amplifier. Artificial dolphin clicks were generated using a Toshiba computer with a DSP board and a custom written program (Simon Boel Pedersen, Copenhagen, Denmark) and connected to the amplifier. The results of the calibration are shown in Fig. 2 for both the 0.5 and 1.0 m array. The error between the calculated range and the measured range increases with distance (Fig. 2). Linear regressions between the measured range and the acoustically calculated range for both the 0.5 m array ($r^2=0.9952$) and the 1 m array ($r^2=0.9975$) were high. However, the slopes of the measured range and the acoustically calculated range were significantly different from the slope of the line $y=x$ in Fig. 2 [two-sample t test; $P<0.05$, (Zar, 1996)].

TABLE I. Criteria used for determining on-axis clicks. Criteria 1–6 were used for set A in Fig. 4 and criteria 1, 2, and 6 (italics) were used for set B in Fig. 4. Criterion 4 was motivated from the analyses of white-beaked dolphin signals presented in Rasmussen and Miller (2002).

Criteria	Description
1	*A maximum source level (p–p) on the center hydrophone (at least 1 dB above the apparent source level at the peripheral hydrophones)*
2	*The dolphin is heading towards the array as observed on the underwater video camera.*
3	The acoustic track of the dolphins, calculated from an acoustic location on a series of clicks, shows the dolphin swimming toward the array.
4	Centroid frequency on the center hydrophone is above 85 kHz.
5	The vertical and horizontal angle between the dolphin and center hydrophone is within 35° (ideally it should be 0°)
6	*The dolphin must be at least 1 m from the array to avoid near field effects* (Au, 1993).

It is very difficult to determine if a click is an on-axis click from field recordings. In captive studies the animal's position is usually fixed with a bite plate; thus the orientation relative the receiver can be carefully monitored (Au, 1993). In field recordings from free-ranging animals the situation is much more problematic and an on-axis click may only be assumed from rather subjective criteria. In most studies only one criterion is used, and there is a great risk of including off-axis clicks in the database. In this study we used a series of criteria to establish a very conservative sorting routine (Table I). We tried to compare a conservative set of criteria (set A) with a less conservative one (set B). All six criteria in Table I must be fulfilled before we consider a click recorded at the center hydrophone to be on-axis for criteria set A. Only criteria 1, 2, and 6 had to be fulfilled for an on-axis click using criteria set B.

III. RESULTS

The dolphins approached the array in groups consisting of females, males, and young animals. However, only adult animals (probably both females and males) approached the array close enough to be observed on the video camera (about 5 m). Females and their young calves never approached the array.

Only 64 on-axis clicks giving 192 angles from eight different sequences (four from 2002 and four from 2003) fulfilled all our criteria for being on-axis clicks (Table I, set A). These clicks were chosen from recordings containing over 1700 different sequences each with from 20 to 500 clicks giving a total of more than 100 000 clicks. For the on-axis clicks, the centroid frequency was 94±5 (SD) kHz on the center hydrophone and 73±22 kHz on the peripheral hydrophones. A track of a dolphin swimming towards the array, derived from consecutive acoustic locations in a series of clicks, is shown in Fig. 3. The differences in the apparent source levels (p–p) between the center and outer hydrophones for each click were calculated along with the corresponding angles. These data were distributed into bins of five degrees: 0°–5°, 6°–10°, 11°–15°, 16°–20°, 21°–25°, and 26°–30°, due to a low sample size at larger angles, bins of

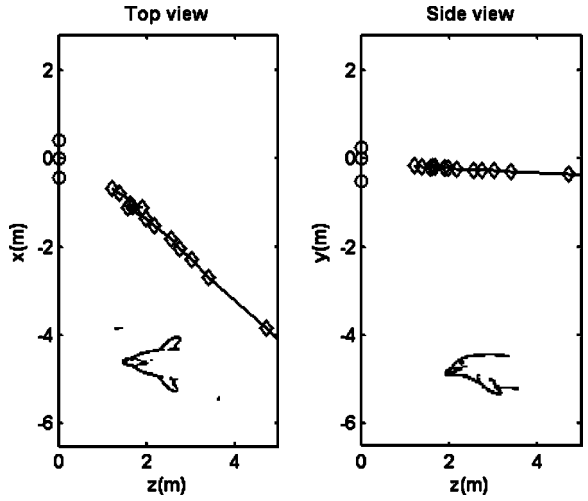


FIG. 3. An example of an acoustic track of a dolphin swimming towards the hydrophone array. The receivers of the array are indicated as circles and the diamonds mark the location of each click made by the dolphin in the whole sequence. The track starts at the location farthest away from the array. The sequence is plotted as seen from the side and from above. The six clicks closest to the array in this sequence fulfilled all our criteria for on-axis clicks (Table I).

10° were used for 31°–40° and for 41°–50°. These bins were chosen to facilitate a comparison with beam patterns from the studies of captive animals. For example, Au (1987) used the angles 0°, ±5°, ±10°, and ±20° in the horizontal plane and −10°, 0°, 5°, 10°, 15°, 20°, and 30° in the vertical plane when measuring the beam pattern of the beluga. In this study we used eight angular bins and the average angle in each bin was derived. These angles were 4°±3° (SD) ($n=24$), 10°±12° ($n=10$), 15°±4° ($n=8$), 20°±6° ($n=47$), 25°±6° ($n=69$), 31°±6° ($n=16$), 41°±7° ($n=8$), and 50°±8° ($n=10$), and the result is shown in Fig. 4, set A. The average difference in source levels were then plotted with standard error bars for each angular bin (Fig. 4). We found the 3 dB

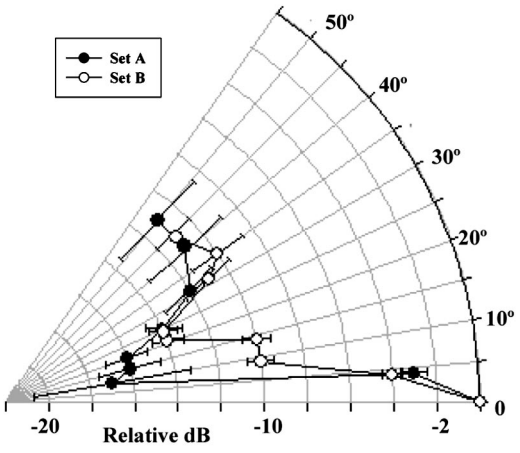


FIG. 4. A polar plot of the beam-pattern of white-beaked dolphin clicks. dB values in the polar plots are differences in source levels (p–p measurements) between the center hydrophone at 0° (on-axis) and a peripheral hydrophone for the respective angles out to 50°±S.E. (see Sec. III). Criteria set A (see Table I) gave 64 on-axis clicks and 192 angles (black dots). Criteria set B gave 267 clicks and 801 angles (white dots). The inclusion of some off-axis clicks for criterion set B may explain the difference between the two curves beyond 5°. The 3 and 10 dB beamwidths were found to be 8° and 10°, respectively (for set A).

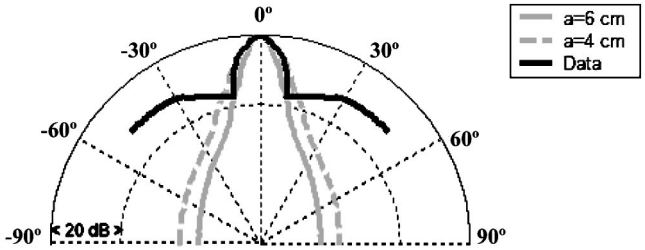


FIG. 5. A comparison between interpolated beam pattern of measured values (see Fig. 4, criterion set A) of white-beaked dolphin clicks (black solid line) and the piston model using a radius of 4 cm (gray dashed line) and 6 cm (gray solid line) excited with a presumed on-axis white-beaked dolphin click. The deviation of the interpolated beam pattern and the model beyond about 10° probably results from slightly off-axis clicks satisfying even our most conservative criteria set (A) in Table I. The directivity index of the whole interpolated beam pattern is 18 dB while that of the central part of the interpolated beam patterns (from 0° to 12°), which closely matches the pattern for the 6 cm piston model, is 29 dB. (See also Fig. 4.)

beamwidth to be 8° and the 10 dB beamwidth to be 10°.

To illustrate the importance of the choice of on-axis criteria we compared the derived results for sets A and B. Set A, used above, uses all the criteria in Table I. Set B is less conservative and uses criteria 1, 2, and 6 in Table I. Set B rendered 267 clicks, giving 801 angles and apparent source levels. These were grouped into angular bins like for data in set A, giving average angles of 4°±3° ($n=228$), 9°±7° ($n=96$), 14°±6° ($n=96$), 20°±8° ($n=96$), 24°±6° ($n=133$), 31°±6° ($n=25$), 35°±6° ($n=16$), and 44°±6° ($n=34$). Figure 4 shows a comparison between the beam patterns of clicks satisfying criteria sets A and B. The 3 dB beamwidth was 8° for criteria sets A and B, but the 10 dB beam width was 18° for clicks selected by criteria set B and broader than that calculated using the stricter set of criteria A.

We used underwater video recordings to judge the orientation of the dolphins during acoustical recordings. Interestingly, we found criterion 2 in Table I (the observations through the video camera) to be of little value. Often when the dolphin seemed to look into the video camera, the source level was not highest on the center hydrophone, and these sequences were therefore not used in the analyses.

For the 64 on-axis clicks satisfying all six criteria in Table I the source level at each peripheral hydrophone relative to the source level of the central hydrophone was plotted against the off-axis angles. We calculated the directivity index from the interpolated beam pattern (Fig. 5) using the formula in Møhl *et al.* (2003) after Urick (1983). The calculation assumes a symmetrical beam around 0° and used an angular step of 1° in the interpolation. The directivity index was 18 dB when using all the measured data for on-axis clicks (from 0° to 50° in Fig. 4).

IV. DISCUSSION

Our study is the first description of an estimated transmission beam pattern measured from free-ranging dolphins. The only other study known to us estimating a beam pattern in a free-ranging odontocete is that of Møhl *et al.* (2003) for the sperm whale. Just as in Møhl *et al.* (2003), our data are also prone to errors for several reasons. First, the source level

calculations assume that the range to the animal is known. With the acoustic location system used here, the range error at ten meters (which approximately corresponds to the maximum distance at which dolphins can be observed through the under water video camera) was 0.5 m for the 1 m array and 1 m for the 0.5 m array (Fig. 2). This gives a maximum error of only 1 dB in source levels corrected for transmission loss. Second, the axis of the acoustic beam may be oriented differently from the swimming direction of the animal (Fig. 3) as noted by Au (1993) and Au *et al.* (1999). Third, determining which clicks are recorded on-axis and which are not is a very difficult problem. Fourth, we assume the beam to be cylindrical, which is approximately the case for (captive) bottlenose dolphins (Au, 1993) and the harbor porpoise (Au *et al.*, 1999). Fifth, the beam pattern will vary with the size of the sound producing organ, and therefore with the size of the animal. As we believe that all recorded dolphins in this study were approximately of equivalent size (grown individuals, body length approximately 2.5–3 m; Rasmussen, 1999), the size problem is probably a less important one. Finally, we assume that all clicks have the same beamwidth. This may not be true, as the beamwidth is expected to vary with the frequency emphasis of the clicks, a parameter known to vary with the emitted source level in odontocetes (Au, 1993).

Despite these problems, some interesting observations can be made from the derived beam pattern of white-beaked dolphins. The 3 dB beamwidth of white-beaked dolphins is measured to be 8°, and the 10 dB beam width to be 10°. In bottlenose dolphins the 3 dB beam width is 8°–10° and the 10 dB beam width is 21°–23° in the vertical and horizontal planes, respectively (Au, 1993). This indicates that white-beaked dolphins have a narrower beam than that reported for bottlenose dolphins. The transmitted beam of white-beaked dolphins is more similar to that of the beluga, where a 3 dB beam width of 6.5° and a 10 dB beamwidth of 13.5° were found (Au *et al.*, 1987). The directivity index of the sound beam of white-beaked dolphins is 18 dB when using all our data (see Fig. 5), whereas that of bottlenose dolphins is 26 dB. The difference is caused by the large side lobes measured in the beam pattern of white-beaked dolphin clicks (Fig. 4). These side lobes are probably not a feature of the transmission beam pattern, but instead they may represent an artifact in the choice of on-axis clicks (see above). A beam of only 8° is very narrow and off-axis clicks may have entered our analysis, in spite of the six criteria (Table I) used for rejecting off-axis clicks.

The sound production in odontocetes can be modeled as a circular piston in a baffle excited by a dolphin click (Au, 1993). This piston model used a radius of 4 cm and a click of a bottlenose dolphin and gave a beam pattern similar to the one measured with a directivity index of 26 dB. A piston radius of 8.9 cm gave a beam pattern similar to that measured from a beluga and a directivity index of 32.1 dB (Au, 1993). We used the piston model excited with a white-beaked dolphin click. Pistons with radii of 4 and 6 cm were chosen because these gave the best fit with the main beam of our data (Fig. 5). For the piston model, the relative intensity at angles larger than 90° was assumed to be constant. The beam pattern of the 6 cm piston model has a 3 dB beam width of 8° and a 10 dB beamwidth of 12° (Fig. 5), which is comparable to the actual measurements of the main lobe of white-beaked dolphin clicks (Fig. 4). The directivity index of the 6 cm piston model was 29 dB, or 3 dB higher than that for the bottlenose dolphin.

Au *et al.* (1999) made a comparison between the measured head diameter of the animals and the directivity index. They found a good correlation between the directivity index and the diameter of the head divided by the peak frequency used by the animal. The directivity index is proportional to the size of the transducer; being greater for a large radiator and *vice versa*. Conversely, the beam width is inversely proportion to the size of a transducer (Urick, 1983 in Au *et al.*, 1999). The white-beaked dolphin is similar in size (also head size) to the bottlenose dolphin and smaller than the belugas. The white-beaked dolphin has similar peak frequency as the bottlenose dolphin and almost the same as the beluga (Au, 1993). According to the previous argument we would have expected the beam width of the white-beaked dolphin to be more like that of the bottlenose dolphin rather than the beluga. Of course the beam width is not only correlated to the head size, but more likely to the size of the sound producing organ, the phonic lips, also called monkey lips/dorsal bursas complex (Cranford *et al.*, 1996, 2000), and presumably also to the acoustic properties of the melon.

We measured large side lobes at angles above 20° when using all our data that satisfied all the criteria in Table I. According to the piston model we would expect the first side lobe at 25° and at −8 dB relative to 0° (Fig. 6.16 in Au, 1993). However, our data show −14 dB at 25° and −11 dB at 50° (Fig. 4). The large angles in our data are derived from dolphins close to the array. It is conceivable that even with our very conservative criteria these clicks are not exactly on-axis with the center hydrophone, and therefore we do not measure the main beam. If off-axis clicks were included in the data, this would broaden the beam as shown in Fig. 4 for the less conservative on-axis criteria (B). We made a simple calculation for illustrating the difficulties of getting on-axis recordings at the center hydrophone at close range. For instance, when the dolphin is 1 m from the array and the main beam of the click is 8°, the circular area will have a diameter of only 14 cm, assuming that the sound beam is cylindrical and symmetrical. With the receivers separated by 50 cm, the dolphin can easily direct its beam between two hydrophones. We may still get the highest source level on the center hydrophone without recording the main beam. When the receivers are 1 m apart, it is even more likely that the dolphin can project the sound beam between the receivers. When the dolphin is 5.7 m away from the 0.5 m array, the 8° beam will ensonify the whole array with highest source levels on the center hydrophone and −3 dB at the outer hydrophones for an on-axis click.

Our measurements indicate that the transmission beam width of white-beaked dolphins is narrower than that for bottlenose dolphins and harbor porpoises and is more similar to the one of belugas. The high centroid frequency of the clicks (Fig. 1) may be the main contributor to the narrow beam. A narrow transmission beam may also be accompa-

nied by a narrow receiving beam, and such an arrangement would improve the performance of the dolphin biosonar under noise-limiting conditions (Au, 1993). A narrow transmission beam may be very useful during prey capture. White-beaked dolphins feed mainly on sand eels (*Ammodytes* sp.) in Faxaflói Bay (Rasmussen and Miller, 2002). Underwater video footage of feeding white-beaked dolphins (Husavik, Whale Museum) shows that white-beaked dolphins herd schools of fish and individual dolphins swim into the school to capture one fish at a time. In order to catch just one sand eel out of a whole school, and to determine its swimming direction, it is an advantage to use a narrow beam to single out the target prey.

ACKNOWLEDGMENTS

This study was supported by the Oticon Foundation and the Danish National Research Foundation. Thanks to Gisli Vikingsson at the Marine Research Institute and Jörundur Svarvarsson at the Institute of Biology, Iceland University in Reykjavik for their cooperation. Thanks to Troels Jacobsen, Malene Simon, Nina Eriksen, Katja Vinding Pedersen, Mille Rasch, Kimie Salo, Jakob Højer Kristensen, and Thorvaldur Benediktsson for assisting in the field. Thanks to the captain and boat owner David Thor Olafsson. We thank Kristian Beedholm for debugging the piston radiation pattern algorithm and Ulrik Nørum for helping with the statistics. We appreciated the constructive comments from two reviewers.

Au, W. W. L. (**1980**). ''Echolocation signals of the Atlantic bottlenose dolphin (*Tursiops truncatus*) in open waters,'' in *Animal Sonar Systems*, edited by R. G. Busnel and J. F. Fish (Plenum, New York), pp. 251–282.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., Floyd, R. W., and Haun, J. E. (**1978**). ''Propagation of Atlantic bottlenose dolphin echolocation signals,'' J. Acoust. Soc. Am. **64**, 411–422.

Au, W. W. L., Moore, P. W. B., and Pawloski, D. (**1986**). ''Echolocation transmitting beam of the Atlantic bottlenose dolphin,'' J. Acoust. Soc. Am. **80**, 688–691.

Au, W. W. L., Penner, R. H., and Turl, C. W. (**1987**). ''Propagation of Beluga echolocation signals,'' J. Acoust. Soc. Am. **82**, 807–813.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Blonz, M., and Gisiner, R. C. (**1995**). ''Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **98**, 51–59.

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). ''Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*),'' J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., and Benoit-Bird, K. J. (**2003**). ''Automatic gain control in the echolocation system of dolphins,'' Nature (London) **423**, 861–863.

Au, W. W. L., and Herzing, D. L. (**2003**). ''Echolocation of wild Atlantic spotted dolphin (*Stenella frontalis*),'' J. Acoust. Soc. Am. **113**, 598–604.

Cranford, T., Amundin, M., and Norris, K. S. (**1996**). ''Functional morphology and homology in the Odontocete Nasal complex: Implications for sound generation,'' J. Morphol. **228**, 223–285.

Cranford, T. W., Elsberry, W. R., Blackwood, D. J., Carr, J. A., Kamolnick, T., Todd, M., Van Bonn, W. G., Carder, D. A., Ridgway, S. H., Bozlinski, D. M., and Decker, E. C. (**2000**). ''Two Independent sonar signal generators in the bottlenose dolphin: Physiologic evidence and implications,'' J. Acoust. Soc. Am. **108**, 2613.

Evans, W. E. (**1973**). ''Echolocation by marine delphinids and one species of fresh-water dolphin,'' J. Acoust. Soc. Am. **54**, 193–199.

Madsen, P. T., Kerr, I., and Payne R. (**2004**). ''Echolocation clicks of two free-ranging, oceanic delphinids with different food preferences: False killer whales (*Pseudorca crassidens*) and Risso's dolphins (*Grampus griseus*),'' J. Exp. Biol. **207**, 1811–1823.

Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (**2000**). ''Sperm whale clicks: Directionality and source level revisited,'' J. Acoust. Soc. Am. **107**, 638–648.

Møhl, B., Wahlberg, M., Madsen, P. T., Heerfordt, A., and Lund, A. (**2003**). ''The monopulsed nature of sperm whale clicks,'' J. Acoust. Soc. Am. **114**, 1143–1152.

Rasmussen, M. H. (**1999**). ''Hvidnæsens lydproduktion, adfærd samt udbredelse,'' Masters thesis, Institute of Biology, University of Southern Denmark, Odense (in Danish).

Rasmussen, M. H., and Miller, L. A. (**2002**). ''Whistles and clicks from white-beaked dolphins, *Lagenorhynchus albirostris* recorded in Faxaflói Bay,'' Aquat. Mamm. **28**, 78–89.

Rasmussen, M. H., Miller, L. A., and Au, W. W. L. (**2002**). ''Sounds and calculated source levels from White beaked dolphins (*Lagenorhynchus albirostris* Gray 1846) recorded in Icelandic waters,'' J. Acoust. Soc. Am. **111**, 1122–1125.

Rasmussen, M. H., and Jakobsen, T. (**2003**). ''Photo-identification of White-beaked dolphins in Icelandic waters,'' in *Proceedings of the 17th Annual Conference of the European Cetacean Society, 9–13 March 2003*, Las Palmas, Gran Canaria.

Schotten, M., Au, W. W. L., Lammers, M. O., and Aubauer R. (**2004**). ''Echolocation recordings and localizations of wild spinner dolphins (*Stenella longirostris*) and pantropical spotted dolphins (*Stenella attenuata*) using a four hydrophone array,'' in *Echolocation in Bats and Dolphins*, edited by J. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 393–400.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).

Zar, J. H. (**1996**). *Biostatistical Analysis* (Prentice–Hall, Englewood Cliffs, NJ).

Correlation of ultrasonic scatterer size estimates for the statistical analysis and optimization of angular compounding

Anthony Gerig,[a] Quan Chen, James Zagzebski, and Tomy Varghese
Department of Medical Physics, University of Wisconsin—Madison, 1300 University Ave. Rm. 1530, Madison, Wisconsin 53706

(Received 15 October 2003; revised 8 March 2004; accepted 9 April 2004)

Ultrasonic scatterer size estimates generally have large variances due to the inherent noise of spectral estimates used to calculate size. Compounding partially correlated size estimates associated with the same tissue, but produced with data acquired from different angles of incidence, is an effective way to reduce the variance without making dramatic sacrifices in spatial resolution. This work derives theoretical approximations for the correlation between these size estimates, and the coherence between their associated spectral estimates, as functions of ultrasonic system parameters. A Gaussian spatial autocorrelation function is assumed to adequately model scatterer shape. Both approximations compare favorably with simulation results, which consider validation near the focus. Utilization of the correlation/coherence expressions for statistical analysis and optimization is discussed. Approximations, such as the invariance of phase and amplitude terms with angle, are made to obtain closed-form solutions to the derived spectral coherence near the focus and permit analytical optimization analysis. Results indicate that recommended parameter adjustments for performance improvement generally depend upon whether, for the system under consideration, the primary source of change in total coherence with rotation is phase term variation due to the change in the relative position of scattering sites, or field amplitude term variation due to beam movement.
© *2004 Acoustical Society of America.* [DOI: 10.1121/1.1756615]

PACS numbers: 43.80.Vj, 43.80.Qf [FD] Pages: 1832–1841

I. INTRODUCTION

Ultrasonic scatterer size estimation and imaging has proven to be both feasible and useful for the monitoring, diagnosis, and study of disease.[1–9] However, it is hampered by low signal-to-noise ratios, which ultimately result from the random placement of scatterers in tissue.[10–12] This difficulty is intrinsic to the majority of techniques used in ultrasonic imaging, including standard B-mode processing in the form of speckle.[13–15] One method that can be used to reduce the severity of this problem is to spatially compound partially correlated results which are associated with the same tissue, but which are produced with data that are taken from different angles of incidence. The application of spatial compounding to B-mode imaging has been investigated extensively, and has yielded excellent results.[16–19] As a result, many modern clinical scanners include a compounding option to reduce speckle and provide the additional benefit of improved specular reflector imaging. Recent work done by our group has been devoted to adapting the spatial compounding technique for use in elastography and parametric imaging that involves spectral analysis, such as scatterer size and attenuation imaging.[20] For the case of parametric imaging, either the power spectral estimates necessary for parameter estimation can be compounded beforehand, or parametric estimates can be generated for each angle of incidence and then averaged.

Much theoretical work, with implications for the statistical analysis and optimization of B-mode spatial compounding, has also been done.[16,18] Wagner *et al.* and O'Donnell *et al.* investigated the dependence of the rf signal and amplitude correlation upon experimental parameters. O'Donnell *et al.* used the results to explore optimizing transducer translation for compounding. However, because work concerned with the application of angular compounding to parametric imaging has been recent, the analogous theoretical research for signal power spectra and parameter estimates has not been done.

The purpose of the following is to develop a theoretical framework that, although pertinent for any type of tissue characterization involving power spectra, will permit statistical analysis and optimization in the application of angular compounding to scatterer size estimation and imaging in particular. Approximate expressions for both spectral coherence and size estimate correlation in terms of experimental parameters, including the angular separation between different incidence directions, are derived. Results of a comparison between these expressions and simulations are presented. A general discussion of the statistical analysis and optimization of angular compounding techniques applied to size estimation, given knowledge of the correlation as derived, ensues. Finally, further assumptions are made which yield analytic solutions to the expression for the spectral coherence, and, based upon the results, parameter adjustment recommendations for performance improvement are given.

II. THEORY

A. Spectral coherence

The essential quantity for determining the correlation between size estimates is the coherence between the Fourier

[a]Corresponding author. Electronic mail: algerig@wisc.edu

transforms of the rf (radio frequency) data segments used to generate those estimates. In general form, the Fourier transform of an rf data segment gated with a rectangular window is given approximately by[21]

$$V(\omega)\approx G(\omega)B(\omega)\omega^2\int_{\Delta\Omega} dr\,\gamma(\mathbf{r})A_t(\mathbf{r},\omega)A_{\mathbf{r}}(\mathbf{r},\omega), \quad (1)$$

where ω is the angular frequency, $G(\omega)$ is the complex transfer function for the transducer, $B(\omega)$ is the complex superposition coefficient corresponding to the insonifying pulse, and A_t and A_r are the field integrals (see the appendix), which include attenuation terms, corresponding to spatial point $\mathbf{r}$ (the origin at the transducer face) for transmit and receive, respectively. It is assumed that a linear, shift-invariant system is scanning a stationary, medium in which only incoherent scatter is significant. $\gamma(\mathbf{r})=[\kappa(\mathbf{r})-\kappa_0]/\kappa_0 -[\rho(\mathbf{r})-\rho_0]/\rho(\mathbf{r})$ is the reflectivity of the scattering medium, where κ and ρ are compressibility and density, and κ_0 and ρ_0 are their corresponding mean values for the medium.[22] $\Delta\Omega$ is the region of the medium which contributes significantly to the gated rf signal. It includes positions in space for which $2|\mathbf{r}|/c$ falls within the time gate, where c is the speed of sound for the medium. Approximating the time gate as a spatial gate in this manner is viable given that the time duration of the system response for a single scatterer is short in comparison to the gate length. As a result, it is most accurate when broadband pulses are used.

Within this framework, reflectivity is a random variable with zero mean. The covariance between two Fourier transformed signal segments, obtained from the same medium but different transducer positions, is therefore given by

$$\begin{aligned}\langle V_1(\omega)V_2^*(\omega)\rangle\approx|G(\omega)B(\omega)|^2\omega^4\int\int_{\Delta\Omega_{1,2}} d\mathbf{r}_1 d\mathbf{r}_2 \\ \times A_t(T(\mathbf{r}_1),\omega)A_r(T(\mathbf{r}_1),\omega) \\ \times A_t^*(\mathbf{r}_2,\omega)A_r^*(\mathbf{r}_2,\omega) \\ \times\langle\gamma(\mathbf{r}_1)\gamma^*(\mathbf{r}_2)\rangle, \quad (2)\end{aligned}$$

where the spatial integration is done in the coordinate system defined by transducer position 2, and T transforms coordinates in this system to those of the system defined by transducer position 1. In this instance, it is used to place the acoustic field generated at position 1 into the coordinate system defined by position 2. This formulation assumes that the acoustic field produced by the transducer is identical for positions 1 and 2. However, it can be easily adapted by changing the field integral for one of the positions to include cases where they are different. Making the change of integration variable, $\mathbf{r}_2=\mathbf{r}_1+\boldsymbol{\Delta}\mathbf{r}$, and invoking the approximation[21] often used in the far-field, $k|\mathbf{r}_1+\boldsymbol{\Delta}\mathbf{r}-\mathbf{r}'|\approx k|\mathbf{r}_1-\mathbf{r}'|+k\cdot\boldsymbol{\Delta}\mathbf{r}$ and $|\mathbf{r_1}+\boldsymbol{\Delta}\mathbf{r}-\mathbf{r}'|\approx|\mathbf{r_1}-\mathbf{r}'|$, yields

$$\begin{aligned}\langle V_1(\omega)V_2^*(\omega)\rangle\approx|G(\omega)B(\omega)|^2\omega^4 \\ \times\int_{\Delta\Omega_1} d\mathbf{r}_1 A_t(T(\mathbf{r}_1),\omega) \\ \times A_r(T(\mathbf{r}_1),\omega)A_t^*(\mathbf{r}_1,\omega)A_r^*(\mathbf{r}_1,\omega) \\ \times\int_{\Delta\Omega_2} R_\gamma(\boldsymbol{\Delta}\mathbf{r})e^{-2i\mathbf{k}\cdot\boldsymbol{\Delta}\mathbf{r}}d\boldsymbol{\Delta}\mathbf{r}, \quad (3)\end{aligned}$$

where $R_\gamma(\boldsymbol{\Delta}\mathbf{r})$ is the autocorrelation function of the reflectivity, is a function of $\Delta\mathbf{r}$ alone for statistically stationary media, and falls to zero quickly for media containing small scatterers.[22] It is, in general, the sharp decrease in this value with distance that makes the approximation described above appropriate. $\mathbf{k}$ (the magnitude denoted by k) is the wavenumber corresponding to the angular frequency, ω, and has direction $(\mathbf{r}_1-\mathbf{r}')/|\mathbf{r}_1-\mathbf{r}'|$ where $\mathbf{r}'$ is the variable of integration in the field integrals. Assuming that the autocorrelation function possesses spherical symmetry, Eq. (3) can be simplified to

$$\begin{aligned}\langle V_1(\omega)V_2^*(\omega)\rangle\approx|G(\omega)B(\omega)|^2C(\omega)\omega^4 \\ \times\int_{\Delta\Omega_{1,2}} d\mathbf{r}_1 A_t(T(\mathbf{r}_1),\omega) \\ \times A_r(T(\mathbf{r}_1),\omega)A_t^*(\mathbf{r}_1,\omega)A_r^*(\mathbf{r}_1,\omega), \quad (4)\end{aligned}$$

where $C(\omega)$ is the Fourier transform of the autocorrelation function, and is directly related to the form factor for the medium.[22] As before, the integration is restricted to values of $\mathbf{r}_1$ for which signals from both fields fall into the time gate. The coherence between Fourier transformed segments is therefore given by

$$\rho_{1,2}(\omega)=\frac{\langle V_1(\omega)V_2^*(\omega)\rangle}{\sqrt{\langle|V_1(\omega)|^2\rangle}\sqrt{\langle|V_2(\omega)|^2\rangle}}\approx\frac{\int_{\Delta\Omega_{1,2}}d\mathbf{r}_1 A_t(T(\mathbf{r}_1),\omega)A_r(T(\mathbf{r}_1),\omega)A_t^*(\mathbf{r}_1,\omega)A_r^*(\mathbf{r}_1,\omega)}{\int_{\Delta\Omega}d\mathbf{r}|A_t(\mathbf{r},\omega)|^2|A_r(\mathbf{r},\omega)|^2}, \quad (5)$$

where it has been assumed that the time gates for the two positions are identical in duration and type. Assuming that the medium contains many independent scatterers per gated segment, the real and imaginary parts of the segment transforms are approximately jointly Gaussian distributed. As a result, the mathematical formalism pertaining to Gaussian distributed variables used by Wagner in determining speckle correlation also applies here, and the coherence between the periodogram power spectral estimates for the segments, given by $|V(\omega)|^2$, can be expressed by the simple formula[18]

$$\rho_S(\omega)=|\rho_{1,2}(\omega)|^2. \quad (6)$$

Although Eqs. (5) and (6) are valid for any two data segments taken from the same medium, the primary concern

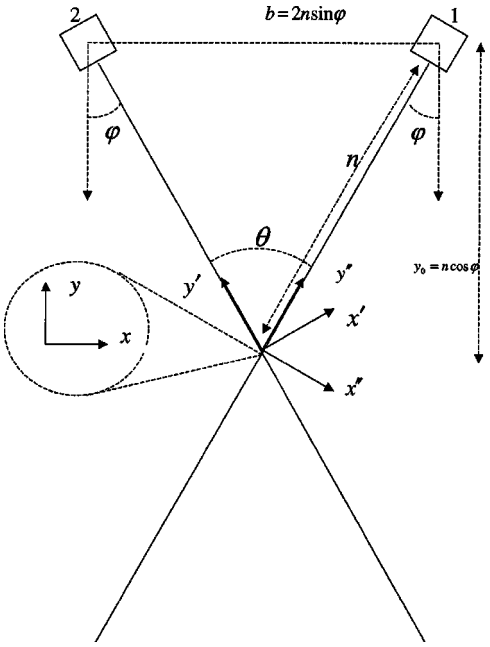


FIG. 1. The coordinate system used in deriving the spectral coherence between angled beamlines.

here is with the coherence between segments obtained from the same location but from different angles of incidence. For this case, if the spatial coordinate system for integration is redefined such that the origin is located at the transducer's axis of rotation, as illustrated in Fig. 1, then the transformation of Eq. (5) becomes a simple rotation about that axis where

$$x''=x'\cos\theta-y'\sin\theta,\quad y''=x'\sin\theta+y'\cos\theta, \tag{7}$$

and θ is the angular separation between positions 1 and 2. In general, Eq. (5) is not easy to calculate analytically due to the inherent difficulty of obtaining closed-form expressions for the acoustic fields involved, and to the complexity of those expressions as integrands when they are available in closed-form. However, it is possible to simplify the integral in certain cases. One of the simulations included in this paper is devoted to verifying Eq. (5) for one such case, namely, when the far-field approximation can be used. The field at the transmit focus of a transducer can be accurately represented by this approximation[23] and will receive particular attention here, although the approximation is appropriate in other instances as well. To further simplify matters for the simulation, attenuation will be neglected. Applying the far-field approximation to Eq. (5) yields

$$\rho_{1,2}(\omega)\approx\frac{\int_{\Delta\Omega_{1,2}}P(T(\mathbf{r}_1),\omega)P(\mathbf{r}_1,\omega)\frac{e^{2ik|T(\mathbf{r}_1)|}}{|T(\mathbf{r}_1)|^2}\frac{e^{-2ik|\mathbf{r}_1|}}{|\mathbf{r}_1|^2}d\mathbf{r}_1}{\int_{\Delta\Omega}P^2(\mathbf{r},\omega)\frac{1}{|\mathbf{r}|^4}d\mathbf{r}}, \tag{8}$$

where P is the point-spread function (amplitude only) of the transducer, which can include terms for dynamic receive aperture and apodization. Assuming approximate field symmetry in the elevational direction such that integration in that direction is unnecessary, and applying the geometry of Fig. 1, where the rotation axis is perpendicular to the diagram and passes through the transducer transmit focus, gives

$$\rho_{1,2}(\omega)\approx\frac{\int_{\Delta\Omega_{1,2}}dx\,dy\,p(x\cos\varphi-y\sin\varphi,x\sin\varphi+y\cos\varphi,\omega)p(x\cos\varphi+y\sin\varphi,-x\sin\varphi+y\cos\varphi,\omega)}{\int_{\Delta\Omega}dx\,dy\,p^2(x,y,\omega)}$$
$$\times\frac{e^{2ik(\sqrt{(x-b/2)^2+(y_0-y)^2}-\sqrt{(x+b/2)^2+(y_0-y)^2})}\dfrac{1}{[(x-b/2)^2+(y_0-y)^2][(x+b/2)^2+(y_0-y)^2]}}{\dfrac{1}{[x^2+(n-y)^2]^2}}, \tag{9}$$

where $p(x,y,\omega)$ is the point-spread function when the coordinate system has been translated from the transducer face to the transmit focus, and half-angle transformations ($\varphi=\theta/2$) have been applied to both transducer positions rather than a full-angle transformation to position 1 alone. b is the distance from position 1 to position 2 and n is the distance from the transducer face to the transmit focus.

B. Size estimate correlation

Size estimation generally involves dividing a spectral estimate obtained from a medium by a type of transfer function for the system used to gather the data, and an attenuation correction term.[1,4] What results is a backscatter estimate for the medium, which can be fit to theoretical curves, whose frequency dependencies are functions of scatterer size, to produce a size estimate. Several different theoretical models have been used to generate size estimates,[4] however, the emphasis here will be upon the Gaussian model for scatterer spatial autocorrelation functions (SAF). For this model, the size estimate (effective radius) corresponding to an individual backscatter estimate is given by[11]

$$\hat{a}^2 = \frac{-d_1^2 c^2 \Sigma_{\omega_{\min}}^{\omega_{\max}} (y(\omega)\omega^2 - \overline{y\omega^2})}{80 \Sigma_{\omega_{\min}}^{\omega_{\max}} (\omega^2 - \overline{\omega^2})^2}, \tag{10}$$

where $y(\omega) = 10 \ln(\hat{BSC}(\omega)/\omega^4)$, $d_1 \approx 3.1$ is a constant that relates the scatterer size to the characteristic length of the Gaussian SAF, and the summation is over discrete frequency values within the bandwidth of the transducer. Although backscatter estimates $(\hat{BSC}(\omega))$ at neighboring frequency points can, in general, be correlated (e.g., if zero-padding or nonrectangular gates are used in a spectral estimation), the assumption here will be that only uncorrelated terms have been included in the summation and that correlated terms have been discarded.

Assuming that a periodogram is used to generate spectral estimates for size estimation, the results from the preceding section can be used to determine the correlation between both backscatter and size estimates produced using correlated rf data segments. If knowledge of the associated system transfer function and attenuation is near perfect, then $\hat{S}(\omega) = f(\omega)\hat{BSC}(\omega)$, where $\hat{S}(\omega)$ is a power spectral estimate and $f(\omega)$ is dependent solely upon the transfer function and attenuation [includes all terms from Eq. (4) for a zero degree angle but $\omega^4 C(\omega)$]. The correlation between backscatter estimates is therefore given by

$$\begin{aligned}
&\rho_{\mathrm{BSC}}(\omega) \\
&= \frac{\frac{1}{f_1(\omega) f_2(\omega)} \langle (\hat{S}_1(\omega) - \langle \hat{S}_1(\omega) \rangle)(\hat{S}_2(\omega) - \langle \hat{S}_2(\omega) \rangle) \rangle}{\frac{1}{f_1(\omega)} \mathrm{std}(\hat{S}_1(\omega)) \frac{1}{f_2(\omega)} \mathrm{std}(\hat{S}_2(\omega))} \\
&= \rho_S(\omega),
\end{aligned} \tag{11}$$

where the subscripts, as previously, refer to different locations in the same medium. From here, the approximation[24]

$$\begin{aligned}
\mathrm{cov}(\hat{a}_1, \hat{a}_2) \approx \sum_{\omega_i} \sum_{\omega_j} &\left(\frac{\partial \hat{a}_1}{\partial \hat{BSC}_1(\omega_i)} \right) \left(\frac{\partial \hat{a}_2}{\partial \hat{BSC}_2(\omega_j)} \right) \\
&\times \mathrm{cov}(\hat{BSC}_1(\omega_i), \hat{BSC}_2(\omega_j))
\end{aligned} \tag{12}$$

can be used to determine the correlation between size estimates, where each partial derivative is evaluated at the expected value of the associated backscatter. Given that the backscatter estimates are uncorrelated from frequency to frequency for both medium locations, the covariance between backscatter estimates for the two locations must be zero unless the frequencies are identical, i.e. if $\hat{BSC}_1(\omega_i)$ and $\hat{BSC}_1(\omega_j)$ are uncorrelated, it follows that $\hat{BSC}_1(\omega_i)$ and $\hat{BSC}_2(\omega_j)$ are also uncorrelated. As a result, Eq. (12) can be simplified to

$$\mathrm{cov}(\hat{a}_1, \hat{a}_2) \approx \sum_{\omega_i} \left(\frac{\partial \hat{a}}{\partial \hat{BSC}(\omega_i)} \right)^2 \langle \hat{BSC}(\omega_i) \rangle^2 \rho_{\mathrm{BSC}}(\omega_i), \tag{13}$$

where the expected value of the backscatter is equivalent to, and has been substituted for, the standard deviation of the backscatter estimator[10] for positions 1 and 2 (legitimate given the earlier assumption that joint-Gaussian statistics apply to the data segment transforms), and the subscripts have been dropped since the quantities involved, other than the backscatter correlation, are independent of spatial position. Finally, using Eq. (10) to calculate the partial derivatives, and dividing the result by the standard deviations of the size estimates, which are also approximated using Eq. (13) where the subscripts are identical, yields (see the Appendix)

$$\rho_a \approx \frac{\Sigma_{\omega_i} (\omega_i^2 - \overline{\omega^2})^2 \rho_{\mathrm{BSC}}(\omega_i)}{\Sigma_{\omega_i} (\omega_i^2 - \overline{\omega^2})^2}, \tag{14}$$

where ρ_a is the correlation between size estimates.

III. METHOD

A. Spectral coherence

To test Eq. (9), rf data were simulated for a linear array transducer interrogating a uniform scattering medium.[25] Each of 11 data sets contained 181 rf echo data lines simulated for transducer rotation increments of 0.5 degrees about the transmit focus of the transducer (3 cm) from a minimum angle of 0 degrees to a maximum of 90 degrees. The background medium speed of sound was set to 1540 m/s, and the attenuation to zero for convenience. Scatterers ($4000/\mathrm{cm}^3$) of diameter 50 micrometers were randomly dispersed within the background medium, and their locations reordered for each data set. The transducer aperture was set to 1.5 cm, the center frequency to 5 MHz with 100 percent bandwidth, and neither apodization, dynamic receive focus, nor dynamic aperture were used. However, to simplify the calculation of Eq. (9) while approximating the effects of dynamic receive focus and dynamic aperture, the field amplitude was artificially restricted to be invariant with depth near the transmit focus, i.e., the region from which rf data segments were selected.

For a specified segment length, spectral estimates were produced for each rf line by gating the signal with a rectangular window of that length centered about the rotation center/transmit focus of the transducer, and generating a periodogram with the result. [Although other windows could have been employed, the accuracy of the gating approximation of Eq. (1) may be reduced for some of the more typical types given that it does not take window shape into account.] To estimate spectral coherence as a function of rotation angle and frequency for comparison with Eq. (9), for every spectral frequency point of each data set, covariance was estimated as a function of angle, theta, by pairing the periodogram value for each rf line with the value for the line corresponding to a rotation increment of theta degrees. The covariance estimates were then averaged across data sets to reduce noise, and normalized to produce coherence curves for each frequency point.

Theoretical values were obtained by numerically integrating Eq. (9). Because the simulation incorporated a linear array without apodization, the point-spread function used was sinc $(\sin(\pi x)/(\pi x))$ squared. Since the field amplitude was artificially made invariant in the focal region, only the position and phase of the fields changed over the range of integration. The integration limits were approximated roughly by setting them according to the gate length: $y_{\min} = (-l/2)\cos\varphi$ and $y_{\max} = (l/2)\cos\varphi$.

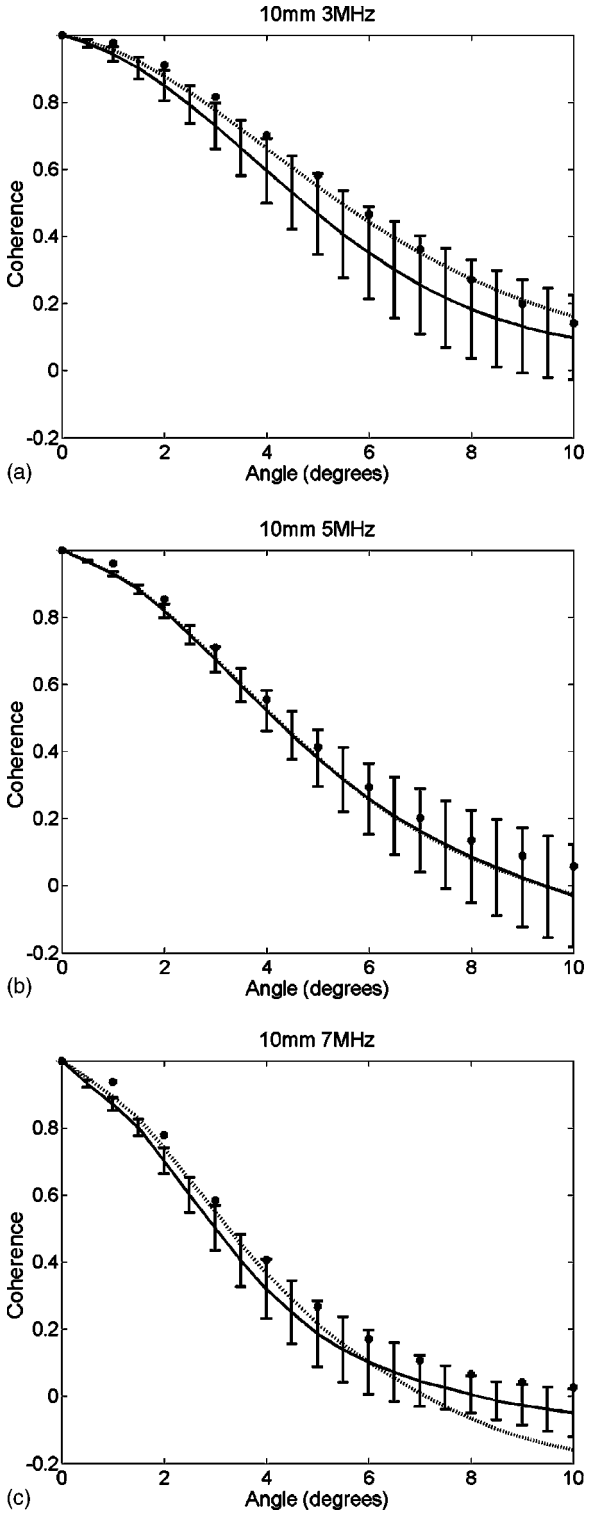


FIG. 2. Spectral coherence versus angle for a gate length of 10 mm and frequencies of (a) 3 MHz, (b) 5 MHz, and (c) 7 MHz. Dots are theoretical values, and solid and dashed lines are the results of simulation experiments.

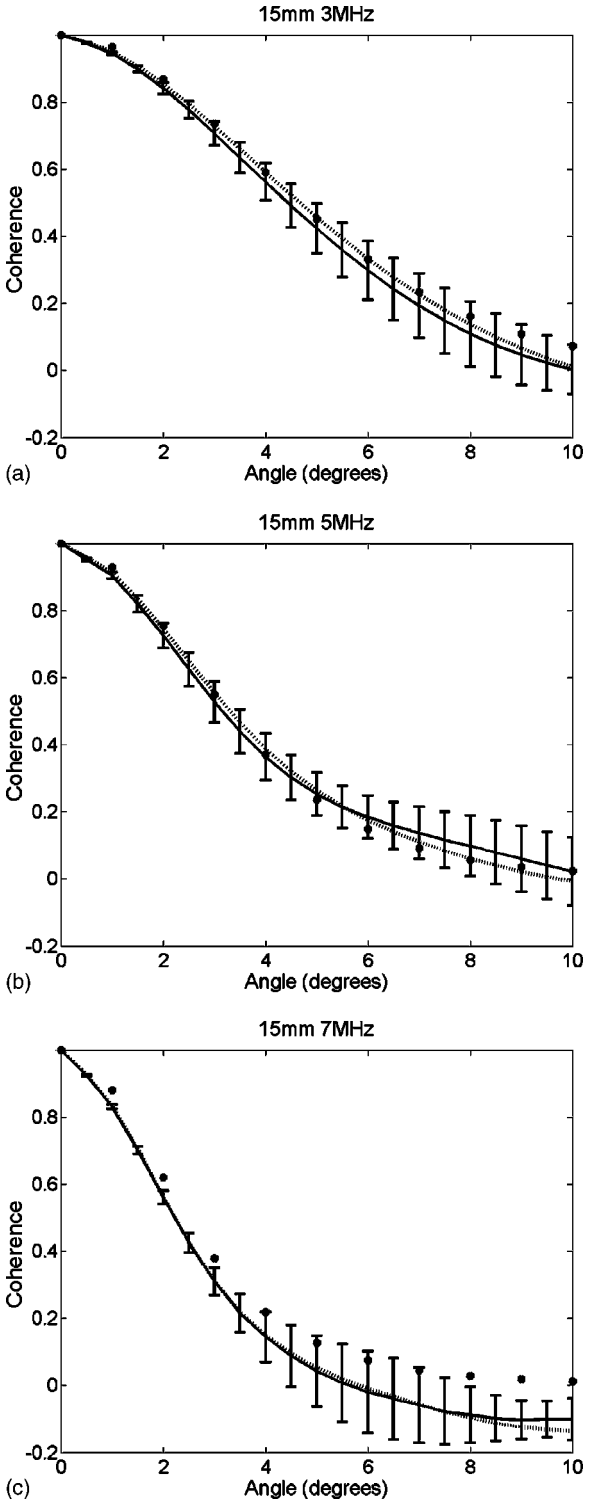


FIG. 3. Spectral coherence versus angle for a gate length of 15 mm and frequencies of (a) 3 MHz, (b) 5 MHz, and (c) 7 MHz. Dots are theoretical values, and solid and dashed lines are the results of simulation experiments.

B. Size estimate correlation

The accuracy of Eq. (14) was also explored using simulated data. However, rather than use the code of the previous section, for which a backscatter correlation must be calculated, new code was written which would generate pairs of rf Fourier transforms according to a specified function of coherence over frequency. Assuming Rayleigh scattering statistics, this could be done by drawing randomly the real and imaginary parts of the transform for individual frequency points in pairs, corresponding to the two different transducer positions, from a joint-Gaussian distribution with zero mean,[15] $\mathrm{var}=\sqrt{\mathrm{BSC}(\omega)/2}$, and $\rho(\omega)=\sqrt{\rho_{\mathrm{BSC}}(\omega)}$, where, again, $\rho_{\mathrm{BSC}}(\omega)$ was selected. $\mathrm{BSC}(\omega)$ had a functional form

which corresponds to a Gaussian spatial autocorrelation function, where the effective size was also a selected parameter.

Ten data sets of 300 Fourier transform pairs each were generated for several different functions of backscatter correlation. The effective scatterer diameter was set to 75 um, and the bandwidth to 100 percent at a center frequency of 5 MHz. For each backscatter correlation function, a periodogram was produced and the scatterer size estimated for every transform. Scatterer size correlation was then estimated for each data set by pairing the size estimates for each Fourier transform pair. The results were averaged across data sets to reduce estimate noise. Theoretical values, for comparison, were generated numerically according to Eq. (14).

IV. RESULTS

A. Spectral coherence

Figures 2 and 3 contain spectral coherence results for two gate lengths and typical frequencies used in size estimation and imaging. For each graph, theoretical and simulation values of coherence are plotted as a function of rotation angle. Dots represent theoretical values, while dashed lines connect simulation results for which averaging, as described in the previous section, was performed across the eleven covariance estimates before normalization. Coherence estimates produced using this method are inherently biased low (i.e., the estimator itself is biased), and become worse as fewer covariance estimates are averaged before normalization. As a result, proper error bars for these values could not be generated by calculating the standard deviation for individual normalized covariance estimates. The results of a separate calculation were therefore included which, although less accurate, allow for error bar generation. Nine of the eleven covariance estimates were divided into three groups of three. Each group of three was averaged and subsequently normalized. Means and standard deviations were then calculated across groups. The results appear as solid lines in the figures. The dashed lines are less biased than the solid lines, but the error bars of the solid lines indicate the general precision of the estimates.

In all cases, the agreement between simulation and theoretical values appears to be good. The coherence estimation bias is believed to be a contributing factor to consistently lower simulation values. However, the disagreement appears to become worse with increasing angle and decreasing frequency for gate lengths shorter than those of Figs. 2 and 3, suggesting that the integration limit approximation described in Sec. III may fail to include significant portions of nonoverlapping acoustic fields in these cases. A more complex treatment of the integration limits and the inclusion of additional, independent data sets may therefore alleviate some of the disagreement.

B. Size estimate correlation

Figure 4 displays the results for size estimate correlation as a function of the associated coherence between rf Fourier transforms. Two simple cases were explored. The first assumes that backscatter correlation for a given rotation angle is constant with frequency, which is generally expected when the rotated point spread function overlaps significantly with the original over the gated region for all frequencies within the transducer bandwidth (see Sec. V). The second assumes a monotonic decrease of backscatter correlation with frequency, which is commonly expected when the rotated point spread function does not overlap significantly with the original. For the sake of simplicity, a linear decrease in correlation with frequency was assumed for this case. Although some agreement between theory and simulation is evident, the theoretical approximation appears to consistently overestimate the true size estimate correlation. In both cases, the error becomes progressively worse as correlation increases. The source of the error appears to be in the approximation itself [Eq. (14)], and caution is therefore suggested in its use, especially for higher values of correlation.

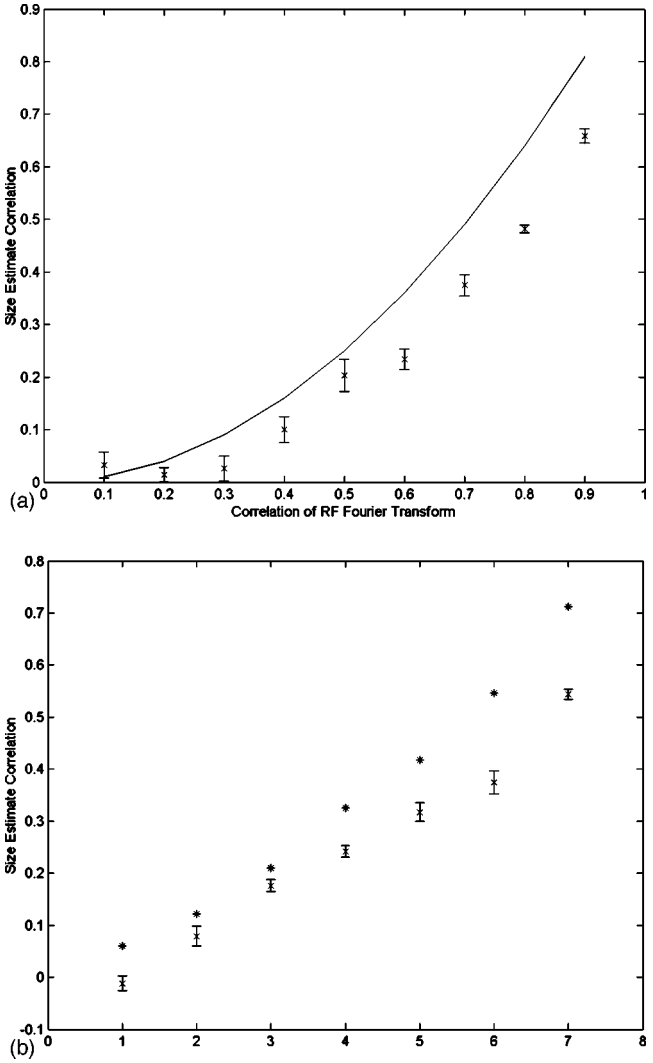


FIG. 4. Size estimate correlation versus rf Fourier transform coherence assuming (a) a constant transform coherence and (b) a linearly decreasing transform coherence with frequency. Minimum and maximum transform coherence values for (b) are as follows: 1: 0.2–0.3, 2: 0.2–0.5, 3: 0.2–0.7, 4: 0.2–0.9, 5: 0.4–0.9, 6: 0.6–0.9, and 7: 0.8–0.9.

V. DISCUSSION: IMPLICATIONS FOR THE STATISTICAL ANALYSIS AND OPTIMIZATION OF ANGULAR COMPOUNDING

Given sufficient accuracy of the theoretical spectral coherence and size estimate correlation formulas as shown in Figs. 2–4, the statistical analysis and optimization of angular compounding techniques applied to scatterer size estimation and imaging becomes possible. For example, consider the simple case discussed previously, where an aperture is rotated about the location of its transmit focus as in Fig. 1. Assume that rf signals are collected at regular angular intervals, $\Delta\theta$, and that compounding of the data is done after size estimation rather than before, i.e., scatterer size estimates, as opposed to spectral or backscatter estimates, are averaged. Using Eq. (9) in conjunction with Eq. (14), an expression for the correlation between size estimates as a function of angular separation, $\rho_{\hat{a}}(n\,\Delta\theta)$, for a specific set of ultrasonic system parameters, can be derived. Assuming that a total of N signals are recorded, corresponding to a maximum compounding angle of $N\,\Delta\theta$, the theoretical improvement in the size estimate SNR after compounding is given by[16]

$$\frac{\mathrm{SNR}_{\mathrm{comp}}}{\mathrm{SNR}}=\left[\sqrt{\frac{1}{N}+2\sum_{n=1}^{N}\frac{(N-n)}{N^2}\rho_{\hat{a}}(n\,\Delta\theta)}\right]^{-1}. \quad (15)$$

By analyzing this value as a function of system parameters and angular increment, it becomes possible not only to design compounding experiments with the desired level of effectiveness (as quantified by the SNR) without correlation measurements, but also to optimize these parameters in order to maximize performance.

Although statistical analysis and optimization can always be done numerically, it is difficult in general to determine, analytically, the dependencies of spectral coherence, size estimate correlation, and size estimate SNR upon system parameters. However, for the simple case described above, a few assumptions and simplifications can be introduced that make analytical progress possible. This, in turn, allows for some general conclusions regarding the use and optimization of angular compounding to be made, connects this work with that done for speckle, and provides insight into the physical mechanisms responsible for size estimate decorrelation. Minimizing size estimate correlation as a function of angular separation is generally necessary for optimizing compounded size estimate/image SNR, and is the case for the simple set-up under consideration here based upon the form of Eq. (15). The following analysis will therefore focus both upon the first steps toward obtaining an analytical solution for size estimate correlation as a function of angle and the adjustment of experimental parameters for its minimization.

Assume, for the sake of simplicity, that the point-spread function for the system is adequately represented by a sinc squared function, and that the amplitude of the field, as earlier, does not vary significantly in the axial direction near the transmit focus.

Approximating the distance in the phase of Eq. (9) as $-bx/z_0$ yields[18]

$$\rho_{1,2}(\omega)\approx\frac{\int_{\Omega_{1,2}}dx\,dy\,\mathrm{sinc}^2\left[\frac{D}{z_0\lambda}(x\cos\varphi-y\sin\varphi)\right]\mathrm{sinc}^2\left[\frac{D}{z_0\lambda}(x\cos\varphi+y\sin\varphi)\right]e^{-2ki(bx/z_0)}}{\int_{\Omega_{1,2}}dx\,dy\,\mathrm{sinc}^4\left(\frac{Dx}{z_0\lambda}\right)}, \quad (16)$$

for the coherence between Fourier transformed segments, where D is the transducer aperture width, λ is the wavelength which corresponds to wavenumber k, z_0 is the distance from the transducer to the transmit focus, and b is again the distance from position 1 to position 2 (see Fig. 1). From here, two simplifying assumptions can be made which will be considered separately. First, the dependence of the point-spread function position upon rotation angle will be neglected by using the approximation $\varphi\approx 0$. Assuming that the integration limits can be adequately handled by limiting integration in the y-direction from $-l/2$ to $l/2$ gives

$$\rho_{1,2}(\omega)\approx\frac{\int dx\,\mathrm{sinc}^4\left(\frac{Dx}{z_0\lambda}\right)e^{-2ki(bx/z_0)}}{\int dx\,\mathrm{sinc}^4\left(\frac{Dx}{z_0\lambda}\right)}. \quad (17)$$

This is simply a normalized, one-dimensional Fourier transform, and is equivalent to the result, studied in detail by both Wagner and O'Donnell,[16,18] obtained for speckle correlation. To distinguish this result from the next, hereafter it will be referred to as phase decoherence (i.e., loss of coherence with angle), since any deviation from complete coherence is the result of a nonzero rotation angle contained in the phase term. The physical mechanism responsible for this type of decoherence is the change in position, relative to the transducer aperture, of scatterers contained within the gated, insonifying beam as the aperture rotates. The phase decoherence mechanism tends to be the dominant source of decoherence when the beam overlaps its original location significantly with rotation.

Without going into the details of the solution to Eq. (17), the result, like most correlation curves, starts at one for a zero rotation angle, and falls to zero as the angle increases. As both Wagner and O'Donnell indicate, it is a function of

$$\frac{b}{D}\approx\frac{\theta}{D/z_0}. \quad (18)$$

The characteristic length of the phase decoherence curve is,

therefore, completely independent of frequency and gate length, but proportional to the aperture size and inversely proportional to the depth of the transmit focus. In essence, this result indicates that wider fields lead to quicker decoherence, and thus a more rapid decorrelation of size estimates with angle, when the phase term is the source of that decoherence. With wider fields, more scatterers away from the rotation center, where the relative change in position with aperture rotation is greater, contribute to the rf signal. However, the invariance of Eq. (17) with frequency indicates that changes in field width due to frequency shifts are negated by offsetting changes to the frequency dependent phase term.

The second case neglects decoherence due to phase changes completely and considers decoherence due to beam movement [i.e., the change in position of the field amplitude term of Eq. (16) with rotation] only by utilizing the approximation $b \approx 0$, $\cos\varphi \approx 1$:

$$\rho_{1,2}(\omega) \approx \frac{\int_{\Omega_{1,2}} dx\, dy\, \mathrm{sinc}^2\left[\frac{D}{z_0\lambda}(x - y\sin\varphi)\right]\mathrm{sinc}^2\left[\frac{D}{z_0\lambda}(x + y\sin\varphi)\right]}{\int_{\Omega_{1,2}} dx\, dy\, \mathrm{sinc}^4\left(\frac{Dx}{z_0\lambda}\right)}, \tag{19}$$

where the integration limits will be handled as before. The numerator is simply a convolution of the sinc squared function in the x-direction, and an integration of the result in the y-direction. The decoherence associated with this approximation, hereafter, will be referred to as field amplitude decoherence. The primary physical mechanism responsible for this type of decoherence is the incorporation of new scatterers within, and exclusion of old scatterers from, the beam as it rotates about the center of the gated region. This mechanism tends to be the dominant source of decoherence when the field amplitude rotates off its original position quickly.

Without generating an analytical solution, it can be demonstrated that Eq. (19) is a function of

$$\frac{lD2\sin\varphi}{z_0\lambda} \approx \frac{lD\theta}{z_0\lambda}. \tag{20}$$

Making the change of variable $x = (\lambda z_0/D)x' - y\sin\varphi$, the numerator of Eq. (19) becomes

$$\int_{-l/2}^{l/2}\int_{-\infty}^{\infty}\frac{\lambda z_0}{D}\mathrm{sinc}^2\left(\frac{2D}{\lambda z_0}y\sin\varphi - x'\right)\mathrm{sinc}^2(x')\,dx'\,dy. \tag{21}$$

Integrating yields

$$\int_{-l/2}^{l/2}\frac{\lambda z_0}{D}f\left(\frac{2D}{\lambda z_0}y\sin\varphi\right)dy, \tag{22}$$

where f is the convolution of sinc^2. Integrating again produces

$$\frac{\lambda z_0 l}{DC}[F(C/2) - F(-C/2)], \tag{23}$$

where C is the quantity of Eq. (20) and F is the integral of f. Upon normalization, Eq. (23) becomes a function of C alone.

In contrast to the previous results for phase decoherence, Eq. (20) indicates that narrower fields yield faster field amplitude decoherence. Essentially, narrower fields tend to separate from their original positions to incorporate new scatterers and exclude old ones much more quickly with rotation than broader fields. The characteristic length of the field amplitude decoherence is also inversely proportional to gate length. The larger the gate, the more nonoverlapping field is generally included in the numerator integral of Eq. (19).

The results for field amplitude and phase decoherence suggest some general principles for the minimization of size estimate correlation as a function of angle. Practically, external constraints on experimental parameters, due to resolution requirements for example, restrict any given system to operation in a specific region of parameter space. If the mechanism responsible for phase decoherence is negligible within that region, then the minimization conditions discussed for field amplitude decoherence (i.e., maximally narrow beams and long gates) minimize both the total decoherence of Eq. (16) and, with adjustments to account for the frequency dependent weighting of Eq. (14), size estimate correlation. Conversely, if the mechanism responsible for field amplitude decoherence is negligible, then the phase decoherence minimization conditions apply for minimizing the total decoherence and size estimate correlation [the weighting of Eq. (14) is irrelevant in this case since phase decoherence is frequency independent].

In reality, however, both mechanisms will generally contribute to the total decoherence with aperture rotation. If this is indeed the case, then minimizing conditions may be more complex. One physical mechanism will often dominate changes to the total decoherence as parameters are adjusted within a particular parameter space, in which case the minimizing conditions of that mechanism will apply. However, a local maximum or minimum in total decoherence may exist within the open parameter space where, regardless of parameter changes, total decoherence either exclusively increases or decreases. In such cases, as with any similar optimization problem, either the location of the local minimum or the boundaries of the open parameter space will determine the total decoherence minimization conditions.

It should be noted that if the field amplitude decoherence mechanism is a source of decoherence for a given set of parameters, an inherent loss of resolution accompanies the

implementation of angular compounding. This is due to the fact that, upon rotation, new scatterers located within neighboring resolution cells contribute significantly to the echo signal. Different spatial compounding techniques (e.g., compounding estimates from parallel beams), which degrade resolution to a corresponding degree, may therefore be just as effective as, and perhaps easier to implement than, angular compounding in instances where the phase decoherence mechanism contributes little to the total decoherence. In contrast, little to no resolution is lost when phase decoherence is predominant. Regardless of the rotation angle, here it is the same scatterers, located within a single resolution cell, which produce the echo signal. As a result, if angular compounding is being considered for a system and set of parameters for which the field amplitude decoherence mechanism is significant yet losses in resolution are unacceptable, then angular compounding should not be implemented. However, if the phase decoherence mechanism is the sole source of decoherence, then angular compounding will only be beneficial.

VI. CONCLUSION

General expressions for the correlation between backscatter coefficient and scatterer size estimates, generated for the same tissue location but from different angles of incidence, were derived and compared against simulation results. These formulas can be used to determine the relationships between experimental system parameters and measures of performance such as SNR, making technique optimization possible at the theoretical level.

In general, two different mechanisms appear to be the sources of rotation decoherence between spectral estimates. The first is the change of relative scatterer position with aperture rotation, which is responsible for changes in the phase term of Eq. (16). The second is the change in position of the acoustic field itself, such that successive rotated fields separate from prior ones. These two mechanisms generally affect the total decoherence in an opposing manner with changing experimental parameters, making optimization conditions dependent upon the dominant source of the change in combined decoherence. Although both phase and field amplitude spectral decoherence ultimately result in improved SNR for compounded size estimates, it should be noted that there is an inherent loss of resolution associated with field amplitude decoherence. As a result, angular compounding may not significantly outperform other spatial compounding techniques that trade image resolution for SNR in cases where phase decoherence is insignificant.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grants No. R21EB002722 and No. T32CA09206.

APPENDIX: FIELD INTEGRALS AND DERIVATION OF EQ. (14)

1.

The field integrals, which first appear in Eq. (1), have the form

$$A_t(\mathbf{r},\omega)=\int_S d\mathbf{r}'\, K_t(\mathbf{r}')\,\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,e^{-\alpha(\omega)|\mathbf{r}-\mathbf{r}'|},$$

$$A_r(\mathbf{r},\omega)=\int_S d\mathbf{r}'\, K_r(\mathbf{r}',|\mathbf{r}-\mathbf{r}'|)\,\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}\,e^{-\alpha(\omega)|\mathbf{r}-\mathbf{r}'|},$$

where the integration is over the face of the transducer (S), K_t contains transmit phase and amplitude modifying terms, such as those for apodization and transmit focusing, and K_r contains analogous receive terms, including those for dynamic receive focusing, dynamic aperture, and apodization. $\alpha(\omega)$ is the medium attenuation, and k is the wavenumber corresponding to ω.

2.

Using Eq. (10) to calculate the partial derivative for Eq. (13) yields

$$\frac{\partial\hat{a}}{\partial\hat{\mathrm{BSC}}(\omega_i)}=\frac{-d_1^2c^2}{80}\,\frac{(\omega_i^2-\overline{\omega^2})}{\sum_{\omega_{\min}}^{\omega_{\max}}(\omega^2-\overline{\omega^2})^2}\,\frac{10}{\langle\hat{\mathrm{BSC}}(\omega_i)\rangle}\,\frac{1}{2\hat{a}},$$

where, as described in the text, the derivative has been evaluated at the expected value of the backscatter estimator. Inserting this expression into Eq. (13) gives

$$\mathrm{cov}(\hat{a}_1,\hat{a}_2)\approx\left(\frac{-d_1^2c^2}{80}\right)^2\left(\frac{10}{2\hat{a}}\right)^2\frac{\sum_{\omega_i}(\omega_i^2-\overline{\omega^2})^2\rho_{\mathrm{BSC}}(\omega_i)}{[\sum_{\omega_{\min}}^{\omega_{\max}}(\omega^2-\overline{\omega^2})^2]^2}.$$

Thus,

$$\mathrm{var}(\hat{a})\approx\left(\frac{-d_1^2c^2}{80}\right)^2\left(\frac{10}{2\hat{a}}\right)^2\frac{1}{[\sum_{\omega_{\min}}^{\omega_{\max}}(\omega^2-\overline{\omega^2})^2]}$$

and dividing the covariance by the variance yields the correlation of Eq. (14).

[1] T. Hall, M. Insana, L. Harrison, and G. Cox, "Ultrasonic measurement of glomerular diameters in normal adult humans," Ultrasound Med. Biol. **22**, 987–997 (1996).

[2] M. Insana, "Modeling acoustic backscatter from kidney microstructure using an anisotropic correlation function," J. Acoust. Soc. Am. **97**, 649–655 (1995).

[3] M. Insana, "Ultrasonic imaging of microscopic structures in living organs," Int. Rev. Exp. Pathol. **36**, 73–92 (1996).

[4] M. Insana, R. Wagner, D. Brown, and T. Hall, "Describing small-scale structure in random media using pulse-echo ultrasound," J. Acoust. Soc. Am. **87**, 179–192 (1990).

[5] M. Insana, T. Hall, J. Wood, and Z.-Y. Yan, "Renal ultrasound using parametric imaging techniques to detect changes in microstructure and function," Invest. Radiol. **28**, 720–725 (1993).

[6] M. F. Insana, T. J. Hall, and J. L. Fishback, "Identifying acoustic scattering sources in normal renal parenchyma from the anisotropy in acoustic properties," Ultrasound Med. Biol. **17**, 613–626 (1991).

[7] M. F. Insana, J. G. Wood, and T. J. Hall, "Identifying acoustic scattering sources in normal renal parenchyma in vivo by varying arterial and ureteral pressures," Ultrasound Med. Biol. **18**, 587–599 (1992).

[8] M. L. Oelze, J. F. Zachary, and J. William D. O'Brien, "Parametric imaging of rat mammary tumors *in vivo* for the purposes of tissue characterization," J. Ultrasound Med. **21**, 1201–1210 (2002).

[9] R. L. Romijn, J. M. Thijssen, and G. W. J. V. Beuningen, "Estimation of scatterer size from backscattered ultrasound: A simulation study," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **36**, 593–606 (1989).

[10] P. Chaturvedi and M. Insana, "Error bounds on ultrasonic scatterer size estimates," J. Acoust. Soc. Am. **100**, 392–399 (1996).

[11] A. Gerig, J. Zagzebski, and T. Varghese, ''Statistics of ultrasonic scatterer size estimation with a reference phantom,'' J. Acoust. Soc. Am. **113**, 3430–3437 (2003).
[12] M. L. Oelze, J. F. Zachary, and W. D. O'Brien, ''Characterization of tissue microstructure using ultrasonic backscatter: Theory and technique for optimization using a Gaussian form factor,'' J. Acoust. Soc. Am. **112**, 1202–1211 (2002).
[13] F. Lizzi, E. Feleppa, M. Astor, and A. Kalisz, ''Statistics of ultrasonic spectral parameters for prostate and liver examinations,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 935–942 (1997).
[14] F. L. Lizzi, M. Astor, E. J. Feleppa, M. Shao, and A. Kalisz, ''Statistical framework for ultrasonic spectral parameter imaging,'' Ultrasound Med. Biol. **23**, 1371–1382 (1997).
[15] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, ''Statistics of speckle in ultrasound B-scans,'' IEEE Trans. Sonics Ultrason. **30**, 156–163 (1983).
[16] M. O'Donnell and S. D. Silverstein, ''Optimum displacement for compound image generation in medical ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 470–476 (1988).
[17] G. E. Trahey, S. W. Smith, and O. T. V. Ramm, ''Speckle pattern correlation with lateral aperture translation: Experimental results and implications for spatial compounding,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 257–264 (1986).
[18] R. Wagner, M. Insana, and S. Smith, ''Fundamental correlation lengths of coherent speckle in medical ultrasonic images,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 34–44 (1988).
[19] R. R. Entrekin, B. A. Porter, H. H. Sillesen, A. D. Wong, P. L. Cooperberg, and C. H. Fix, ''Real-time spatial compound imaging: Application to breast, vascular, and musculoskeletal ultrasound,'' Semin Ultrasound CT MR **22**, 50–64 (2001).
[20] A. L. Gerig, T. Varghese, and J. A. Zagzebski, ''Improved parametric imaging of scatterer size estimates using angular compounding,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 708–715 (2004).
[21] J.-F. Chen, J. Zagzebski, F. Dong, and E. Madsen, ''Estimating the spatial autocorrelation function for ultrasound scatterers in isotropic media,'' Med. Phys. **25**, 648–655 (1998).
[22] M. Insana and D. Brown, ''Acoustic scattering theory applied to soft biological tissues,'' in *Ultrasonic Scattering in Biological Tissues*, edited by K. Shung (CRC Press, Boca Raton, 1993), pp. 75–124.
[23] A. Macovski, *Medical Imaging Systems* (Prentice–Hall, Upper Saddle River, NJ, 1983).
[24] P. Bevington and K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed. (McGraw-Hill, New York, 1992).
[25] Y. Li and J. A. Zagzebski, ''A frequency domain model for generating B-mode images with array transducers,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 690–699 (1999).

Erratum: 5aMU1. Movement amplitude and tempo change in piano performance [J. Acoust. Soc. Am. 115, 2590 (2004)]

Caroline Palmer
Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC H3A 1B1, Canada

Simone Dalla Bella
Kazimierz Wielki University, Bydgoszcz, 85-867 Poland

(Received 7 June 2004; accepted for publication 25 June 2004)

[DOI: 10.1121/1.1782934]

PACS numbers: 43.75.Cd, 43.10.Vx

The correct authors and affiliations for this manuscript are Caroline Palmer, Dept. of Psychology, McGill Univ., 1205 Dr. Penfield Ave., Montreal, QC H3A 1B1, Canada, Caroline.palmer@mcgill.ca and Simone Dalla Bella, Kazimierz Wielki Univ., Bydgoszcz, 85-867 Poland.

 © 2004 Acoustical Society of America